Genetic basis of O-antigen variation in *Salmonella* spp. : *rfb* gene clusters of *S. typhi* and *S. paratyphi A*

Naresh K. Verma, M.V.Sc. (Pantnagar)
Department of Microbiology
The University of Sydney

A thesis submitted for the degree of
Doctor of Philosophy
June 1988

TABLE OF CONTENTS

Chapter 2: Materials and methods 41

Chapter 4: Localisation of DNA responsible for specificity in the biosynthesis of paratose and tyvelose, determination and analysis of its nucleotide sequence 61

ABSTRACT

The O-antigens of *Salmonella* serogroups A, B and D differ structurally in their side-chain sugar residue. The genes encoding O-antigen biosynthesis are clustered in the *rfb* operon. This thesis describes the molecular cloning and analysis of the *rfb* genes from *S. paratyphi A* IMVS1316 (serogroup A) and *S. typhi* Ty21a (serogroup D). The *rfb* gene cluster from *S. typhimurium* LT2 (serogroup B) has also been studied in our laboratory (Brahmbhatt *et al* 1986, 1988; P. Wyk, Ph.D. thesis, Univ. of Adelaide, 1988). The genetic basis of O-antigen variation in serogroups A, B and D of *Salmonella* is discussed.

The *rfb* gene cluster of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 were cloned into cosmid vectors. The regions of DNA homology and nonhomology between the *rfb* regions of the serogroup A, B and D representatives were identified, and the junction between the homologous and nonhomologous regions defined by heteroduplex analysis. The only gross difference found between the *rfb* regions of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 was a 2.8 kb DNA segment that was repeated three times in *S. paratyphi A* IMVS1316, and gave rise to 5.6 kb additional DNA.

The *rfbS* gene of *S. typhi* Ty21a (equivalent to the *rfbJ* gene of *S. typhimurium*) has been localised on a 1.2 kb *rfb* DNA fragment. This fragment, when transformed into a *S. typhimurium* strain, caused it to produce a mixed type of LPS (both 0-2 and 0-4 specific). We sequenced the *rfbS* gene from *S. typhi* Ty21a. The products of genes *rfbS* and *rfbJ* both have dehydrogenase function and use the

same substrate for their reactions. The amino acid sequence of protein encoded by the *rfbS* of *S. typhi* Ty21a showed substantial homology with the amino acid sequence of the protein encoded by *rfbJ* of *S. typhimurium* ; however, there was a high degree of variation between these genes at the DNA level.

Matsuhashi (1966) showed that CDP-paratose is reversibly converted to CDP-tyvelose by an epimerisation at C-2. The gene for this epimerase function has been localised on a 1.4 kb *rfb* DNA fragment of *S. typhi* Ty21a and the product of this gene can make a group A strain produce tyvelose. The nucleotide sequence of the *rfbE* gene in *S. typhi* Ty21a has been determined. The amino acid sequence of proteins encoded by *rfbE* and *galE* of *E. coli* show homology at the N-terminus region. The proteins of the *rfbS* (*S. typhi*), *rfbE* (*S. typhi*), and *galE* (*E. coli*) genes were found to have significant similarities with the NAD binding domain, identified by Rossman *et al* (1975), in structurally known dehydrogenases. The *rfb* regions sequenced were found to have low G+C content (only 35% of all bases). The boundaries of homologous and nonhomologous DNA between *S. typhi* Ty21a and *S. typhimurium* in the vicinity of position 9.57 (*S. typhimurium*) and 5.5 (*S. typhi* Ty21a) have been defined at the base sequence level. We sequenced the *rfbE* gene in *S. paratyphi A* IMVS1316 and found a frame shift mutation in the coding region. Lack of *rfbE* gene product was confirmed by minicell analysis of that region. No evidence of a transpositional event or structural repeat in the sequence at either end of the triplicated DNA in *S. paratyphi A* IMVS1316 was found.

The *rfb* genetic structure in several strains of groups A, B and D, independently isolated from natural populations, was studied. A 2.8 kb *rfb* DNA segment is repeated two, three or four times in the *rfb* region of *S. paratyphi A* strains. All group A strains studied were found to contain the *rfbE* gene; but in inactive form, indicating derivation from group D progenitors. We have shown that, in at least one instance, a naturally occurring group A strain was derived by loss of its epimerase function.

STATEMENT

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

N.K. Verma

Naresh K. Verma

June 1988

ACKNOWLEDGMENTS

I would like to thank Professor Peter Reeves for his helpful discussions, valuable comments and guidance throughout the course of this work.

I would also like to thank Donna for her invaluable help in typing, and reading of this thesis. Thanks are due to all colleagues in the laboratory for their helpful suggestions and to John Angles for the use of his computer.

Finally, I am greatly indebted to my parents for their continuous encouragement and patience during the course of my Ph.D. studies.

CHAPTER 1

Introduction

The external layer of the cell envelope of *Salmonella* and related gram-negative bacteria is a membranous structure which contains the lipopolysaccharide (LPS) of the envelope in addition to phospholipid and protein. LPS carries the O-specific polysaccharide (O-antigen) which is the major cell surface antigen and it shows high structural variation in the genus *Salmonella*. The genetic determination of the LPS has been studied in *Salmonella typhimurium*. The genes specifying the synthesis of the O-polysaccharide are encoded mainly in the *rfb* gene cluster.

The peripheral "O-chain" portion of cell wall LPS of *Salmonella* serogroups A, B and D is known to differ in short branches of single 3, 6-dideoxyhexose residues that are attached to a main chain (Lüderitz *et al* 1971). In this thesis the cloning and analysis of the *rfb* genes from *S. paratyphi A* (serogroup A) and *S. typhi* (serogroup D), and comparison with the known corresponding cluster of *S. typhimurium* (serogroup B) at the molecular level is presented. This chapter briefly reviews the literature on the structure, biosynthesis and genetics of individual LPS components and related cell surface antigens in *Salmonella* and other enterobacteria.

1.1 LPS structure and function

LPS are integral components of the outer membrane of gram-negative bacteria, where they form complexes with proteins and phospholipids (Lugtenberg and Van Alphen 1983). The chemical structure of LPS indicates that the molecule consists of three separate parts: lipid A, an oligosaccharide core and a chain of polymerized units carrying the O-antigenic determinants (Lüderitz *et al* 1968). The core oligosaccharide and lipid A are linked by 3-deoxy-D-*manno* -2-octulosonic acid (KDO), which was originally thought to be typical of enterobacterial LPS. It is however, a component of many capsular polysaccharides of invasive bacteria (Jann and Jann 1983). The ketosidic KDO linkages are hydrolysed under very mild conditions and allow the degradation of LPS into lipid A and a carbohydrate moiety (Jann *et al* 1975).

Most of the LPS molecules appear to be located in the outer leaflet of the bilayer (Mühlradt and Golecki, 1975; Funahara and Nikaido, 1980) while the inner leaflet of the bilayer is almost entirely composed of phospholipids (Kamio and Nikaido, 1976; Schindler and Teuber, 1978). Takeuchi and Nikaido (1981) have shown, by using spin-labelled phospholipids reconstituted with LPS, that LPS-LPS interaction is very strong in comparison with LPS-phospholipid interactions, and that domains containing pure LPS are almost completely stable in bilayers containing LPS and phospholipids in 'physiological ratios'. This inherent stability of LPS-only domains obviously contributes to the segregation of LPS and phospholipids (Vaara and Nikaido 1984). Takeuchi and Nikaido (1981) measured the strength of LPS-LPS interactions 'relative' to the LPS-phospholipid interaction, but there is no indication of the absolute magnitude of these forces.

Available information on the structure and conformational properties of LPS and free lipid A (Labischinski *et al* 1985) shows that the LPS exhibits a remarkably highly ordered state, when compared to natural phospholipids, which is mediated by its well-ordered lipid A arrangement and results in a relatively rigid structure. The O-specific chains assume a heavily coiled conformation but it is not known whether these sugar chains are intermingled in the native state. It has been suggested that the hydrophilic region of lipid A (phosphorylated glucosamine disaccharide) is oriented 45^{o} to the membrane surface, forming a 'sawtooth roof'-like surface stucture. LPS molecules with their lipid A units tend to form domains and the more or less ordered arrangement of these anisotrophic molecules can be preserved for long periods of time.

The majority of LPS types derived from different bacterial genera show acidic reactions following deionization by electrophoresis, indicating a predominance of negatively charged groups in the molecule (Galanos and Lüderitz 1975). Galanos and Lüderitz (1975) have also shown that the LPS in free acid form undergoes hydrolysis when heated in distilled water at 100^{o} C; the ketosidic linkage between lipid A and polysaccharide is cleaved and the two parts separated. Metal cations (Na^{+}, K^{+}, Ca^{+}, Mg^{+}, Fe^{2+} and Fe^{3+}) and basic anions (ethanolamine, putrescine, spermidine and spermine) both play a crucial role in the organization of the LPS monolayer in the outer membrane. LPS molecular interaction with outer membrane proteins was evidenced by Ames *et al* (1974) and Koplow and Goldfine (1974). They observed a drastic decrease in outer membrane proteins in deep rough mutants of *S. typhimurium* and *E. coli* ; these mutants were extremely sensitive to various hydrophobic agents. Concomitant with the decrease in proteins, the phospholipid content of the outer membrane showed significant increase, whereas the LPS content stayed constant (Smit *et al* 1975). Lugtenberg *et al* (1976) reported a 30 to 45% decrease in major outer membrane protein/total cellular protein ratio, and showed that the decrease was mainly due to the almost total disappearance of the OmpF porin, with less pronounced changes in the amount of the OmpA protein. It has been suggested from membrane reconstitution experiments that the *E. coli* OmpC porin interacts with the core region of LPS while the OmpA protein has been clearly shown to interact with whole LPS (Beher *et al* 1980).

The Rd and Re mutants of *S. typhimurium* and heptose-less mutants of *E. coli* K12 are sensitive to hydrophobic agents and dyes (Schlecht and Schmidt 1969; Schmidt *et al* 1969; Schlecht and Westphal 1970; Sanderson *et al* 1974; Roantree *et al* 1977), fatty acids (Sheu and Freese, 1973), phenol (Schmidt *et al* 1969), polycyclic hydrocarbons (Ames *et al* 1973), and anionic (Schmidt *et al* 1969; Stocker and Mäkelä 1971; Sanderson *et al* 1974) as well as cationic detergents (Vaara 1981; Nikaido and Vaara 1985). Deep rough mutants show only a slightly increased sensitivity to the polycationic antibiotic polymyxin (Roantree *et al* 1977), which is not particularly hydrophobic, but has an affinity to LPS. In contrast to the increased sensitivity to hydrophobic antibiotics, the deep rough mutants do not show hypersensitivity to hydrophilic agents.

In some cases a decrease of sensitivity is found presumably due to decrease in the number of porins in the outer membrane.

Major alterations in the outer membrane composition may be the main reason behind the molecular basis for the increased permeability of outer membrane of deep rough mutants. It has been shown that, in deep rough mutants, phospholipid molecules are found in the outer leaflet of the membrane (Kamio and Nikaido 1976) and this strongly suggests that a portion of the outer membrane must be occupied by phospholipid bilayer regions. Generation of these phospholipid bilayer regions, which are highly permeable to hydrophobic molecules, could adequately explain the increased permeability to all hydrophobic agents (Vaara and Nikaido 1984). Deep rough mutants also release periplasmic enzymes into the medium (Lindsay *et al* 1972; Chatterjee *et al* 1976). Transient ruptures of the outer membrane are the most likely explanation of this phenomenon (Nikaido and Nakae 1979). EDTA in presence of Tris buffer is known to release about one-half of LPS (Leive 1965) also with a concomitant increase in outer membrane permeability to hydrophobic agents (Leive 1974). This phenomenon is thought to be due to a similar mechanism of weakening LPS-LPS interactions, disrupting the outer membrane and forming a phospholipid bilayer in the outer leaflet of the outer membrane. More recently, it has been shown that short and long O-specific chains are equally susceptible to the effects of EDTA and polycations (Hukari *et al* 1986). The polycationic decapeptide antibiotic, polymyxin, exerts its lethal activity by binding to the phospholipids in the cytoplasmic membrane (Teuber and Bader 1976), thereby destroying its barrier property (Storm *et al* 1977).

It gains access to the cytoplasmic membrane by disrupting the outer membrane (Teuber and Bader 1976), making the membrane permeable to hydrophobic agents (Rosenthal and Storm 1977), detergents (Vaara and Vaara 1981), and lysozyme (Teuber 1970). There is growing evidence that certain cationic host defence factors increase the outer membrane permeability. One such agent is the 'bacterial permeability increasing protein' or BPI (Elsbach and Weiss 1983), isolated from the granule contents of polymorphonuclear leukocytes, complexing with isolated LPS, binding to the outer membrane, and increasing its permeability to hydrophobic antibiotics.

1.1.1 Lipid A

Lipid A has a very highly conserved structure in the enteric organisms so far examined. It represents the active centre of LPS, being responsible for the induction of typical endotoxin effects observed in mammals, such as fever, hemodynamic changes, disseminated intravascular coagulation, and shock states (Galanos *et al* 1977). It represents the endotoxic principle of biologically active LPS and, as an LPS component, it may be regarded as a virulence factor. It plays a significant role in organization, stability and barrier function of the membrane and appears to be essential for the integrity, growth and survival of bacterial cells (Osborn 1979; Nikaido 1979; Nikaido and Nakae 1979).

The chemical structure of lipid A's from distinct bacterial organisms has been elucidated in various studies (Lüderitz *et al* 1978, 1982; Rietschel *et al* 1982, 1983, 1984a, 1984b; Takayama *et* *al* 1984). More recently, physical methods such as NMR, fast atom bombardment mass spectroscopy (FABMS) and laser desorption mass spectroscopy (LDMS) have been used for analysis of lipid A architecture. The lipid A of *E. coli* has recently been investigated very intensively by chemical and physical methods and is known to consists of β -1'.6-interlinked 2-deoxy-2-amino-D-glucopyranisyl (GlcN) residues (Hase and Rietschel 1976; Rosner *et al* 1979; Blache *et al* 1980). This disaccharide carries phosphate groups in position 4' of the nonreducing glucosaminyl residue (GlcNII) and in position 1 of the reducing glucosaminyl group (GlcNI), the latter being α -linked (Rosner *et al* 1979; Strain *et al* 1983a, 1983b). The hydrophilic backbone is substituted by 3-deoxy-D-*manno*- 2- octulosonic acid (KDO), a further phosphoryl group, and a characterstic spectrum of long-chain fatty acids.

The structure of the lipid A backbone in *Salmonella* was first established by analysis of the LPS in a *S. minnesota* Re mutant R595 (Gmeiner *et al* 1969). Recent studies (Batley *et al* 1985a, 1985b, 1985c) using NMR spectroscopy on *S. minnesota* R595 LPS, have shown that there are two KDO residues and only four fatty acid substituents per molecule of LPS. Comparison of the lipid A's shows that they do not differ significantly in their architecture from those of *E. coli, P. vulgaris, S. minnesota,* or *S. typhimurium* (Rietschel *et al* 1984b).

1.1.2 Core

Structural studies on LPS cores of *Enterobacteriaceae* have been carried out mainly on *Salmonella* (Hämmerling *et al* 1970), *E.*

coli (Feige and Stirm 1976; Feige *et al* 1977; Jansson *et al* 1981) and *Shigella* (Jansson *et al* 1979; Katzenellenbogen and Romanowska 1980; Gamian and Romanowska 1982). The core structure is quite common to all *Salmonella* species examined and only minor variations of this structure are found in other members of the *Enterobacteriaceae*. The *Salmonella* core contains a lipid A-distal hexose oligosaccharide consisting of D-glucosamine, D-glucose, and D-galactose, and an inner lipid A - proximal region consisting of an oligosaccharide of the core specific sugars, L-glycero-D-*manno* -heptose (L-D-Hep) and 2-keto-3-deoxy-D-*manno* -octonate (KDO), each forming a branched trisaccharide (Lüderitz *et al* 1966a; Osborn 1966; Hellerqvist and Lindberg 1971). The exact linkages in the KDO trisaccharide region have not been definitely established (for review of chemistry and biology of KDO, see Unger 1981) and it is known now that KDO units are present as dimers (Strain *et al* 1983a, 1983b). Compositional analysis of the core from various bacterial species have revealed the presence of such unusual constituents as D-glycero-D-*manno*-heptose, uronic acid (Kotelko *et al* 1974), amino acids (Drewry *et al* 1975) and, in the case of *Vibrio cholerae* strains, fructose (Redmond *et al* 1973; Jann *et al* 1973; Raziuddin 1980) or seduheptulose may be present. These are thought to replace KDO in linking the core to lipid A (summarized by Galanos *et al* 1977; Wilkinson 1977). The O-chains are attached to the subterminal glucose II unit of the core. The polar groups such as phosphate and ethanolamine, and the carboxyl groups of KDO contribute to the acidic character of LPS. Several bacteriophages use the core as their site of primary adsorption (Hellerqvist and Lindberg 1971; Wilkinson *et al* 1972; for review of phage receptors see Lindberg

1977). Substituents of the main chain of the inner core, the nature of which are not clear, are not necessarily present in molar amounts.

1.1.3 O-specific chain

O-specific chains are extremely polymorphic in their constituent sugars and linkages in *Salmonella* and other genera of the *Enterobacteriaceae*. They form the serologically dominant part of the LPS molecule (Lüderitz *et al* 1966b, 1971). Each bacterial serotype synthesizes a unique LPS, characterized by a specific composition and structure of the O-chain, and by an individual O-antigenicity. Kauffmann (1966) classified *Salmonella* strains into various groups based on shared O-antigen factors. The members of each group could show differences in other antigen factors, and some factors could be shared by some or all strains of different groups (Lüderitz *et al* 1968).

The O-specific polysaccharide chain length of a given bacterial strain is always heterogeneous (ranging between 10-40 units or more) as demonstrated by electrophoresing LPS preparations on SDS-PAGE gels or by degradation and gel permeation chromatography of the polysaccharide moieties (Goldman and Leive 1980; Jann *et al* 1975; Palva and Mäkelä 1980). This phenomenon is interpreted as the presence of individual LPS molecules with O-specific polysaccharide chains differing in size by multiple increments of oligosaccharide repeating units. *Salmonella* O-specific chains contain neutral polysaccharides while some strains of *E. coli* and *Shigella* have been shown to have acidic components such as hexuronic acids, N-acetylneuraminic acid (Neu NAc) or phosphate. LPS with neutral and with acidic O-polysaccharides can easily be differentiated by immunoelectrophoresis (Ørskov *et al* 1971, 1977; Ørskov and Ørskov 1972). The O-specific chain of *S. typhimurium* has a backbone of mannose-rhamnose-galactose with an abequose residue branching from the mannose. Detailed chemical characterization of the determinant groups and immunodominant sugar residues (Table1.1) has been the subject of various reviews (Lüderitz *et al* 1971; Jann and Westphal 1975). It has been shown that some polysaccharides act as receptors for specific bacteriophages (Lindberg 1973). Several bacteriophages encoding specific endoglycosidases (Iwashita and Kanegasaki 1973; Lindberg 1973; Svenson and Lindberg 1983) are known to cleave the bond between repeat units resulting in serotype conversion.

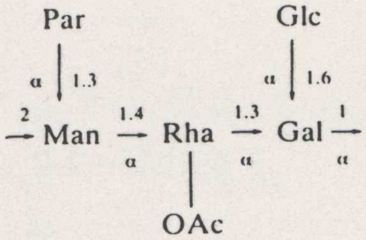
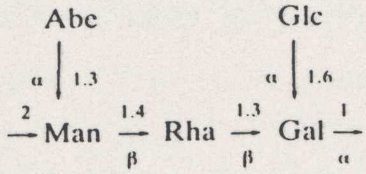
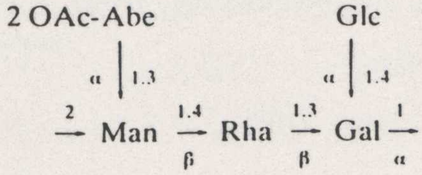
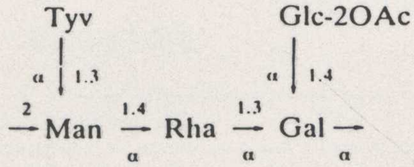
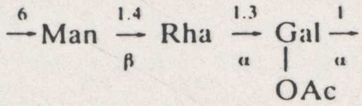
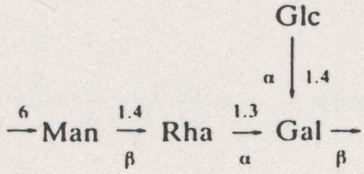
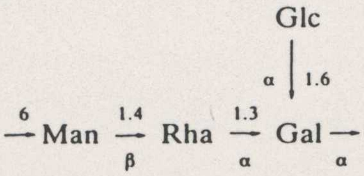
1.1.3.1 O-specific chain similarities between various enterobacterial strains

The structure of O-specific chains from strains of various enterobacterial genera has been studied (for reviews, see Lüderitz *et al* 1971; Kenne and Lindberg 1983; Shibaev 1986). *Salmonella* O groups A, B, D and E are shown to have an identical backbone (mannose-rhamnose-galactose) which differs from that of other *Salmonella* such as O groups C_1, C_2 and C_3 (Table1.2a,b). The O-specific chains of serogroups A, B and D can be distinguished, since each has a different 3,6-dideoxyhexose attached to the mannosyl residues as a side branch. The side branch is tyvelose in serogroup D, paratose in serogroup A, and abequose in serogroup B (Jann and

TABLE 1.1

Structure of O-repeating units from *Salmonella* with a Man-Rha-Gal backbone of the O-specific polysaccharide from O groups A, B, D and E. The table is adapted from Jann and Jann (1984).

Table 1.1

Salmonella (O group)	Structure of O-repeating unit	Reference
S. paratyphi A (A)	 $\xrightarrow{2}\overset{\text{Par}}{\overset{\alpha\downarrow 1.3}{\text{Man}}}\xrightarrow[\alpha]{1.4}\underset{\text{OAc}}{\underset{\vert}{\text{Rha}}}\xrightarrow[\alpha]{1.3}\overset{\text{Glc}}{\overset{\alpha\downarrow 1.6}{\text{Gal}}}\xrightarrow[\alpha]{1}$	Hellerqvist et al 1971c
S. bredeney (B)	 $\xrightarrow{2}\overset{\text{Abe}}{\overset{\alpha\downarrow 1.3}{\text{Man}}}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\beta]{1.3}\overset{\text{Glc}}{\overset{\alpha\downarrow 1.6}{\text{Gal}}}\xrightarrow[\alpha]{1}$	Hellerqvist et al 1969c
S. typhimurium (B)	 $\xrightarrow{2}\overset{\text{2 OAc-Abe}}{\overset{\alpha\downarrow 1.3}{\text{Man}}}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\beta]{1.3}\overset{\text{Glc}}{\overset{\alpha\downarrow 1.4}{\text{Gal}}}\xrightarrow[\alpha]{1}$	Hellerqvist et al 1969a
S. typhi (D)	 $\xrightarrow{2}\overset{\text{Tyv}}{\overset{\alpha\downarrow 1.3}{\text{Man}}}\xrightarrow[\alpha]{1.4}\text{Rha}\xrightarrow[\alpha]{1.3}\overset{\text{Glc-2OAc}}{\overset{\alpha\downarrow 1.4}{\text{Gal}}}\xrightarrow[\alpha]{}$	Hellerqvist et al 1969b
S. anatum (E1)	 $\xrightarrow{6}\text{Man}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\alpha]{1.3}\underset{\text{OAc}}{\underset{\vert}{\text{Gal}}}\xrightarrow[\alpha]{1}$	Robbins and Uchida 1962 Hellerqvist et al 1971a
S. newington (E2)	$\xrightarrow{6}\text{Man}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\alpha]{1.3}\text{Gal}\xrightarrow[\beta]{1}$	Robbins and Uchida 1962 Hellerqvist et al 1971a
S. illinois (E3)	 $\xrightarrow{6}\text{Man}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\alpha]{1.3}\overset{\text{Glc}}{\overset{\alpha\downarrow 1.4}{\text{Gal}}}\xrightarrow[\beta]{}$	Robbins and Uchida 1962
S. senftenberg	 $\xrightarrow{6}\text{Man}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\alpha]{1.3}\overset{\text{Glc}}{\overset{\alpha\downarrow 1.6}{\text{Gal}}}\xrightarrow[\alpha]{}$	Hellerqvist et al 1971d

Par = paratose (3,6-dideoxy-D-glucose); Abe = abequose (3,6-dideoxy-D-galactose); Tyv = tyvelose (3,6-dideoxy-D-mannose)

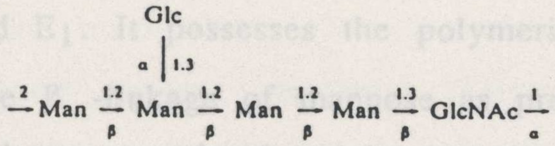
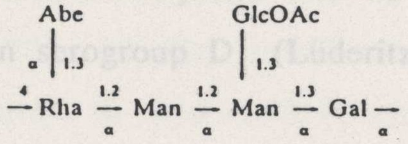
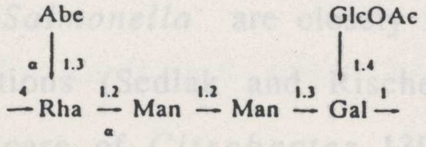
TABLE 1.2

A) Structure of O-repeating units from *Salmonella* C1, C2 and C3.

B) Comparison of O structures from *Citrobacter* 396 and *Salmonella* strains

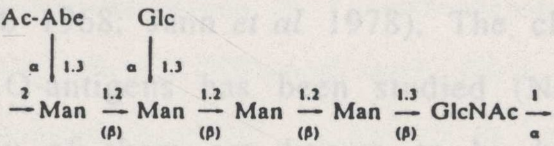
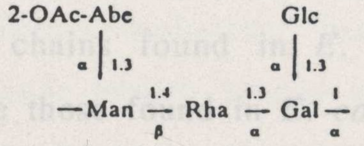
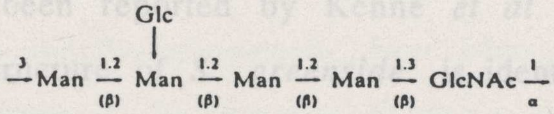
(from Jann and Jann 1984).

Table 1.2a

Salmonella	O-group	Structure of the O-repeating unit	Reference
S. thompson *S. montevideo* *S. choleraesuis*	C1	 $\xrightarrow{2}\text{Man}\xrightarrow[\beta]{1.2}\overset{\substack{\text{Glc}\\ \alpha\downarrow 1.3}}{\text{Man}}\xrightarrow[\beta]{1.2}\text{Man}\xrightarrow[\beta]{1.2}\text{Man}\xrightarrow[\beta]{1.3}\text{GlcNAc}\xrightarrow[\alpha]{1}$	Fuller and Staub 1968 Fuller et al 1968
S. newport	C2	 $\xrightarrow{4}\overset{\substack{\text{Abe}\\ \alpha\downarrow 1.3}}{\text{Rha}}\xrightarrow[\alpha]{1.2}\text{Man}\xrightarrow[\alpha]{1.2}\overset{\substack{\text{GlcOAc}\\ \downarrow 1.3}}{\text{Man}}\xrightarrow[\alpha]{1.3}\text{Gal}\xrightarrow[\alpha]{}$	Hellerqvist et al 1971b
S. kentucky	C3	 $\xrightarrow{4}\overset{\substack{\text{Abe}\\ \alpha\downarrow 1.3}}{\text{Rha}}\xrightarrow[\alpha]{1.2}\text{Man}\xrightarrow{1.2}\text{Man}\xrightarrow{1.3}\overset{\substack{\text{GlcOAc}\\ \downarrow 1.4}}{\text{Gal}}\xrightarrow{1}$	Hellerqvist et al 1970

Abe = abequose (3,6-dideoxy-D-galactose)

Table 1.2b

O antigen from	Structure of the O-repeating unit	Reference
Citrobacter 396	 $\xrightarrow{2}\overset{\substack{\text{2-OAc-Abe}\\ \alpha\downarrow 1.3}}{\text{Man}}\xrightarrow[(\beta)]{1.2}\overset{\substack{\text{Glc}\\ \alpha\downarrow 1.3}}{\text{Man}}\xrightarrow[(\beta)]{1.2}\text{Man}\xrightarrow[(\beta)]{1.2}\text{Man}\xrightarrow[(\beta)]{1.3}\text{GlcNAc}\xrightarrow[\alpha]{1}$	Jann et al 1978
S. typhimurium	 $\longrightarrow\overset{\substack{\text{2-OAc-Abe}\\ \alpha\downarrow 1.3}}{\text{Man}}\xrightarrow[\beta]{1.4}\text{Rha}\xrightarrow[\alpha]{1.3}\overset{\substack{\text{Glc}\\ \alpha\downarrow 1.3}}{\text{Gal}}\xrightarrow[\alpha]{1}$	Hellerqvist et al 1969a
S. montevideo	 $\xrightarrow{3}\text{Man}\xrightarrow[(\beta)]{1.2}\overset{\substack{\text{Glc}\\ \downarrow}}{\text{Man}}\xrightarrow[(\beta)]{1.2}\text{Man}\xrightarrow[(\beta)]{1.2}\text{Man}\xrightarrow[(\beta)]{1.3}\text{GlcNAc}\xrightarrow[\alpha]{1}$	Fuller and Staub 1968 and J. Lönngren, personal communication

Abe = abequose (3,6-dideoxy-D-galactose)

Jann 1984; Mäkelä and Stocker 1984). The structure of serogroup D_2 is interesting since it appears to be a chemical intermediate between group D and E_1. It possesses the polymerization linkage (gal-1,6-Man) and the β -linkage of mannose as present in group E_1, but its galactose is not acetylated and its mannose carries a side chain of tyvelose as in serogroup D_1 (Lüderitz *et al* 1971).

Citrobacter and *Salmonella* are closely related and show many serological cross-reactions (Sedlak and Rische 1961; Edwards and Ewing 1966); in the case of *Citrobacter* 139 and *Salmonella* of group B the O-specific polysaccharides have been shown to have the same structure (Yuan and Horecker 1968). Structural analysis of the polysaccharide has shown that O-antigen of *Citrobacter* 396 is a molecular hybrid between those of *S. montevideo* (serogroup C_1) (Fuller and Staub 1968) and *S. typhimurium* (serogroup B) (Fuller and Staub 1968; Jann *et al* 1978). The chemical composition of *Klebsiella* O-antigens has been studied (Nimmich and Korten 1970), and many of them are known to be homopolysaccharides. The mannan chains found in *E. coli* O9 and *Klebsiella* O3 are identical while those found in *E. coli* O8 and *Klebsiella* O5 are very similar. It has been reported by Kenne *et al* (1983) that the O-specific chain structure of *S. greenside* is identical with that of *E. coli* O55 (Lindberg *et al* 1981) while the O-specific chain polysaccharide structure is very similar to that of *E. coli* O111 (Eklund *et al* 1978).

With respect to immunochemistry and biochemistry, there is a close relationship between *E. coli* and *Shigella*. The acidic heteropolysaccharides of *E. coli* O58 and O124 are identical with the O-specific polysaccharides of *Shigella dysenteriae* 5 and 3, respectively (Dmitriev *et al* 1975, 1977) and these bacteria are known causative agents of dysentery. The O-antigens of *Proteus* species contain several unusual structural features (Sutherland 1977). Some contain ribitol phosphate, ethanolamine phosphate or galacturonic acid residues. Lysine, linked to the carboxyl group of the latter is also common in these LPS (Gromska and Mayer 1976).

1.1.3.2 O-specific chains in other bacteria

Yersinia pseudotuberculosis and *Yersinia enterocolitica* have a variety of O-antigens which differ from those of *Salmonella* (Winblad 1968; Samuelsson *et al* 1974; Thomshich *et al* 1976; Hoffman *et al* 1980; Kenne and Lindberg 1983), although the LPS of *Y. enterocolitica* has not been studied as completely as that of *Salmonella*. Several serotypes of *Y. enterocolitica* contain the rare sugar, 6-deoxy-L-altrose (Ellwood and Kirk 1971; Acker *et al* 1980). *Y. pseudotuberculosis* is known to have all forms of 3,6-dideoxyhexoses, including ascarylose, which comprise the immunodominant branch groups in the oligosaccaride O units. The structure and chemical composition of O-antigens in other bacteria such as *Vibrio* and *Pseudomonas* have been studied and found to be different to those of *Salmonella* (Kenne *et al* 1978; Dmitriev *et al* 1980).

1.2 Lipopolysaccharide biosynthesis

The different structural regions of LPS are synthesized by separate biosynthetic pathways. Lipid A is made from two monosaccharide precursors, which are substituted with fatty acids, KDO, phosphates, ethanolamine and 4-amino-D-arabinose. The synthesis of KDO involves three sequential reactions. These three reactions are catalyzed by the enzymes d-ribulose-5-phosphate isomerase, KDO-8-phosphate synthase, and KDO-8-phosphate phosphatase, respectively. Finally, free KDO is converted to CMP-KDO, the donor of KDO residues for inner core biosynthesis (Ghalambor and Heath 1966). The biosynthesis of the outer core region of *S. typhimurium* LPS has been studied extensively (for reviews, see Osborn and Rothfield 1971; Rothfield and Romeo 1971; Nikaido 1973). Synthesis of this region involves a series of membrane-bound glycosyltransferases that catalyse the sequential transfer of sugars from nucleotide-sugar donors to the acyl carrier lipid (ACL); the sugars are then polymerized on the same carrier. The ACL-linked polysaccharide is subsequetly transferred to core-lipid A completing the formation of the O-antigen. The complete LPS molecule is then translocated across the membrane system.

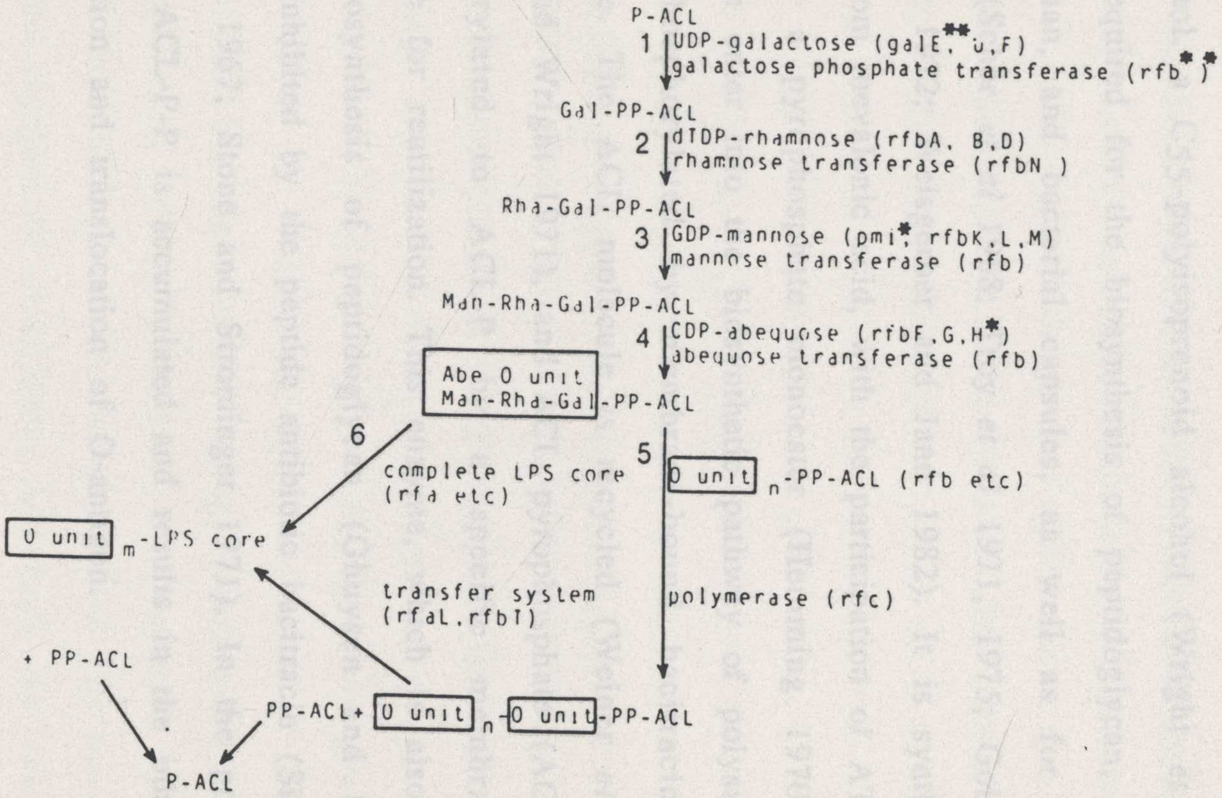
1.2.1 Biosythesis of O-specific polysaccharide

The general pathways of O-antigen biosynthesis are shown in Fig1.1. The first stage is assembly of the oligosaccharide repeating unit on the acyl carrier lipid (ACL). The ACL can only function as an acceptor in the form of its monophsphate ester (ACL-P). The lipid bound oligosaccharides are then polymerized, still in association with the ACL. The final step in the sequence is translocation of the O- polysaccharide from the ACL to core-lipid A.

FIGURE 1.1

O-unit assembly, polymerization and attachment to LPS core

The steps (1-4) in the assembly of O unit of *Salmonella typhimurium*, its polymerization (step 5) and attachment (step 6) of either the single O unit or the polymer to the LPS core. The mono- or oligosaccharide donors and transferases (with the genes determining them in parentheses) are shown to the right at each step. Notice the utilization and release of P-ACL (antigen carrier lipid), monophosphate undecaprenol. Mutation of some genes (marked with one asterisk) has been shown to lead to detergent-sensitivity of the mutant bacteria, whereas this sensitivity is relieved by mutation of other genes (those found to do so are marked with a double asterisk) (from Mäkelä and Stocker 1984).


P-ACL
1
UDP-galactose (galE**, u, F)
galactose phosphate transferase (rfb**)
Gal-PP-ACL
2
dTDP-rhamnose (rfbA, B, D)
rhamnose transferase (rfbN)
Rha-Gal-PP-ACL
3
GDP-mannose (pmi*, rfbK, L, M)
mannose transferase (rfb)
Man-Rha-Gal-PP-ACL
4
CDP-abequose (rfbF, G, H*)
abequose transferase (rfb)
Abe 0 unit
Man-Rha-Gal-PP-ACL
5
0 unit n-PP-ACL (rfb etc)
polymerase (rfc)
6
complete LPS core
(rfa etc)
0 unit m-LPS core
transfer system
(rfaL, rfbT)
+ PP-ACL
PP-ACL+ 0 unit n-0 unit-PP-ACL
P-ACL

1.2.1.1 Acyl carrier lipid (ACL)

ACL has been isolated from *S. anatum*, and its structure determined by mass spectroscopy. It is a monophosphoester of undecaprenol, a C55-polyisoprenoid alcohol (Wright *et al* 1967). ACL is required for the biosynthesis of peptidoglycan, yeast cell wall mannan, and bacterial capsules, as well as for O-antigen synthesis (Scher *et al* 1968; Troy *et al* 1971, 1975; Goldman and Strominger 1972; Weisgerber and Jann 1982). It is synthesized in the cell from mevalonic acid, with the participation of ATP, and is formed as a pyrophosphate monoester (Hemming 1970). Before ACL-P can enter into the biosynthetic pathway of polysaccharides, it is dephosphorylated by membrane-bound bacitracin-sensitive phosphatase. The ACL molecule is recycled (Weiner *et al* 1965; Robbins and Wright 1971), and ACL pyrophosphate (ACL-P-P) is dephosphorylated to ACL-P by a specific membrane-bound phosphatase for reutilization. This enzyme, which is also involved in the biosynthesis of peptidoglycan (Ghuysen and Shockman 1973), is inhibited by the peptide antibiotic bacitracin (Siewert and Strominger 1967; Stone and Strominger 1971). In the presence of bacitracin, ACL-P-P is accumulated and results in the inhibition of polymerization and translocation of O-antigen.

1.2.1.2 Transferases

In *S. typhimurium* the galactose-1-phosphate is transferred to the ACL with the formation of a pyrophosphate bond. Subsequently other sugars are added without incorporation of one of their phosphates into the product (Osborn and Tze-Yuen 1968; Robbins and Wright 1971). The transfer of galactose phosphate in the assembly is reversible (Nikaido and Nikaido 1971; Nikaido *et al* 1971), whereas subsequent incorporations of sugar constituents with the formation of glycosidic bonds, are irreversible. The specificity of the rhamnosyl transferase has been studied using chemically synthesized monosaccharide-P-P-ACL (Shibaev 1978; Danilov *et al* 1980, 1981; Shibaev *et al* 1982), and it was found that the rhamnose residue can be added to fucose and talose, as well as to galactose attached to the lipid carrier.

1.2.1.3 Polymerization

The *rfc* gene is required for the polymerization of the O-units in *Salmonella* strains of serogroups A, B, D and E, whereas *Salmonella* C_1 and L and *E. coli* O8 and O9 strains need a functional *rfe* gene (Mäkelä *et al* 1970; Stocker and Mäkelä 1978). Strains which depend on a functional *rfe* gene for O-antigen expression do not need the *rfc* - directed oligosaccharide polymerase. The *rfc* - directed polymerization of lipid-linked oligosaccharides has been studied using exogenously generated oligosaccharide-ACL (Weiner *et al* 1965; Robbins *et al* 1966; Osborn and Weiner 1968; Kanegasaki and Wright 1970). The reaction takes place in the cytoplasmic membrane, and both the intermediates and the enzymes participating in it are membrane-bound. The chain grows at its proximal 'reducing' end, so that the already partly polymerized polysaccharide is transferred to the end of a single O unit-diphospho-undecaprenol serving as its acceptor (Robbins *et al* 1967). The undecaprenol diphosphate is concomitantly released from the polymer (Fig 1.1), and probably then dephosphorylated to undecaprenol phosphate which can be used in subsequent synthesis of O units or other polysaccharides (Mäkelä and Stocker 1984). The linkage formed by the polymerase contributes to the structure and hence the antigenic specificity of the O polysaccharide (Staub and Bagdian 1966). In *S. typhimurium* it was shown that abequose is transferred to the oligosaccharide before polymerization (Weiner *et al* 1965; Osborn and Weiner 1968), and that its presence is necessary for polymerization to occur (Yuasa *et al* 1969). The polymerase in this system is not absolutely specific, lipid linked oligosaccharides generated in the membranes of *S. senftenberg, S. anatum* and *S. typhimurium* which contain analogs of galactose and mannose, can act as substrates of the polymerase (Shibaev 1978; Shibaev 1982). The regulation of polymerase activity is very poorly understood. The chain length of the products was thought to be in the region of 10-30 repeating units (Nikaido 1973). However, SDS-PAGE patterns of complete O-antigens have revealed a heterogeneity of the chain lengths which extends up to more than 40 repeating units (Jann *et al* 1975; Goldman and Leive 1980; Palva and Mäkelä 1980)

1.2.1.4 Postpolymerization modifications

The O-antigen of a *Salmonella* strain may be heterogeneous with respect to certain subsituents. Thus, glucosyl and/or O-acetyl substituents may or may not be present on the O-specific polysaccharide of a particular strain. The substitution of the O-antigen with a-glucose takes place at the level of membrane-associated, ACL-linked polysaccharide (Takeshita and Mäkelä 1971; Nikaido *et al* 1971; Mäkelä 1973). A glucosylated lipid intermediate has been characterised as β-glucosyl-P-ACL (NIkaido and Nikaido 1971). Here, glucose is transferred to ACL-P, as opposed to the transfer of Gal-1-P by the galactosyl trasferase of the first sugar of the O-unit. Whereas the transfer of Gal-1-P from UDP Gal occurs with retention of the anomeric galactosyl configuration, that of glucose from UDP Glc occurs with inversion of this configuration. The same type of glucosylation, as described for *Salmonella*, seems to operate in *Shigella*, whereby the main chain is glucosylated at different sites (Simmons 1971). It has also been shown that bacteriophage conversion seems to involve changes either in polymerization or in postpolymerization modifications (Bray and Robbins 1967b; Losick and Robbins 1967).

1.2.1.5 Translocation

The translocation of the completed O-specific chain to the core is catalysed by translocase. This enzyme shows dual specificity; for the O-specific polysaccharide and for the specific core structure. Translocation has been achieved *in vitro,* and has been studied using *S. typhimurium* sugar activation defective mutants (Nikaido 1965), *galE* mutants (Osborn and Weiner 1968) and in wild type *S. anatum* and its phage converted derivatives (Bray and Robbins 1967a; Losick and Robbins 1967; Robbins and Wright 1971). The acceptor site for the O-specific polysaccharide in *S. typhimurium* was found to be the C-4 position of the subterminal glucose of the core (Nikaido 1970; Yuasa *et al* 1970). In *S. anatum*, during the translocation, the last molecule of ACL participating in the biosynthesis of the O-antigen molecule is liberated in the form of its pyrophosphate derivative. The energy required in the translocase reaction is provided by the cleavage of the glycosyl-pyrophosphate bond, with the formation of a lower energy glycosyl bond.

An unexpected finding resulted when the genes determining the *Shigella sonnei* form I antigen were transferred into *S. typhi* Ty21a. Here the form I antigen is not covalently bound to core lipid A but instead occurs as an independent, highly polymerized, immunogenic entity on the cell surface. This suggests that the polymerized O-antigen may not require covalent attachment to core lipid A (Seid *et al* 1984).

1.2.1.6 Export across to the outer membrane

The mechanism for LPS transport through the inner and outer membrane and its subsequent insertion into the outer membrane remains largely unknown. Mühlardt *et al* (1973) demonstrated that newly formed O-antigen appeared after approximately 2 minutes on about 200-300 distinct patches over the cell surface. These patches corresponded to regions of the bacterial cell surface where the cytoplasmic and the outer membrane come in contact, known as Bayer's junctions, now believed to be the sites for LPS export (Bayer 1975, 1979; Bayer *et al* 1982). Mühlradt and Golecki (1975) showed that in native state the LPS is asymmetrically located only on the outer face of the outer membrane; later confirmed by Funahara and Nikaido (1980). Recently, Mulford and Osborn (1983), also using the immunoelectron microscopy technique, found an intermediate stage in the biosynthesis where LPS localised on the periplasmic face of the inner membrane possibly indicating that LPS export occurs from the periplasmic and not the cytoplasmic surface of the inner membrane as was previously believed. The mechanism by which the LPS molecules cross the hydrophobic barrier of the cytoplasmic membrane remains unclear.

1.3 Lipopolysaccharide genetics

Two large blocks of genes, the *rfa* (map position 79 on the 100 unit map of *S. typhimurium* LT2) and *rfb* (map position 42) clusters are involved in core and O-specific chain biosynthesis respectively. Apart from these, there are common housekeeping genes in various parts of the chromosome which also have a role in the synthesis of many constituents of both oligosaccharides. The basic structure can be further modified by various additions determined by genes outside the *rfa* and *rfb* gene clusters, which may often be phage encoded. By contrast, the genetics of lipid A is not so well understood. A major block in the elucidation of lipid A genetics has been the difficulty in isolating mutants altered in lipid metabolism, (Raetz 1982). However, recent development of techniques for immobilization of bacterial colonies (Raetz 1975) on filter paper, allowing rapid autoradiographic detection of specific lipid enzymes *in situ* (Raetz 1982) has circumvented this problem. Two genes have been identified specifying enzymes UDP-GlcNAc acyltransferase (*lpxA*) and disaccharide synthase (*lpxB*), which are involved in the biosynthesis of lipid A. The molecular cloning and nucleotide sequence determination of these genes (Crowell *et al* 1986, 1987; Coleman and Raetz 1988) indicate that they are situated in an operon near min 4 on the *E. coli* chromosome and transcribed in clockwise direction.

1.3.1 Genetics of the core

Mutations of *S. typhimurium* defective in several of the core oligosaccharides, as well as other enzymes involved in core synthesis, are members of a group of rough mutants designated rough A or *rfa*. These mutants have been classified as such on the basis of a variety of phenotypic criteria involving in each case an alteration of core structure (Ra to Re) (for reviews, see Stocker and Mäkelä 1971; Mäkelä and Stocker 1984). The genes whose mutations cause these defects, fall into three classes. The first consists of structural genes for enzymes involved in the synthesis of the sugar nucleotides: some of these genes also have functions in the cell economy unrelated to LPS core biosynthesis. The second class comprises the genes specifying enzymes which catalyse the transfer of a sugar unit from a donor compound to the incomplete LPS core; all these genes map in the *rfa* cluster. The third class regulates the activity of other *rfa* genes: gene *rfaH* of *S. typhimurium*, and its equivalent, *sfrB*, of *E. coli* K12, are the only genes identified so far.

Genetic studies of a series of outer core mutants of *S. typhimurium* have led to a fairly detailed elucidation of the genes involved in biosynthesis and identification of the five transferases (*rfaK, J, I, G, B*) required in assembly. All these genes are located in the main *rfa* gene cluster. The order of *rfa* genes has been defined using R-prime plasmids carrying the *pyrE-rfa-cysE* region of the chromosome of *S. typhimurium* for complementation analysis (Kadam *et al* 1985). The heptose component of the core is genetically less well characterized though several mutants (*rfaC* to *rfaF*) are known that prevent its completion. The KDO part of the molecule is genetically very poorly understood and mutants of only two genes participating in its biosynthesis are known. Some defects have incomplete effects and the result may be a 'part-rough' phenotype, with a fraction of the LPS molecules having the expected R structure, while others are complete with the O-polysaccharide. Thus, even an amber mutation, presumably causing complete or virtually complete loss of function in gene *rfaK* for addition of the N-acetylglucosamine unit on glucose II, produces only a 'part-rough' phenotype (Gemski and Stocker 1967), apparently because the O-chain translocase system is able to add O-chains to a small fraction of core molecules of type Rb1.

1.3.2 Genetics of O-specific chains of *Salmonella*

The genetics of the O-specific polysaccharide has been almost exclusively studied in *Salmonella*, and most of that work has used *S. typhimurium* LT2 as the experimental organism. Mutants blocked in the synthesis of the O-chain can be selected by looking for an R phenotype with complete core. The simplest method is to select for resistance to 'S-specific' phages such as the phage P22 in *S. typhimurium*. Most of the enzymes used in the biosynthesis of the O unit are encoded in one gene cluster, the *rfb*, at position 42 on the *S. typhimurium* chromosome (Fig1.2).

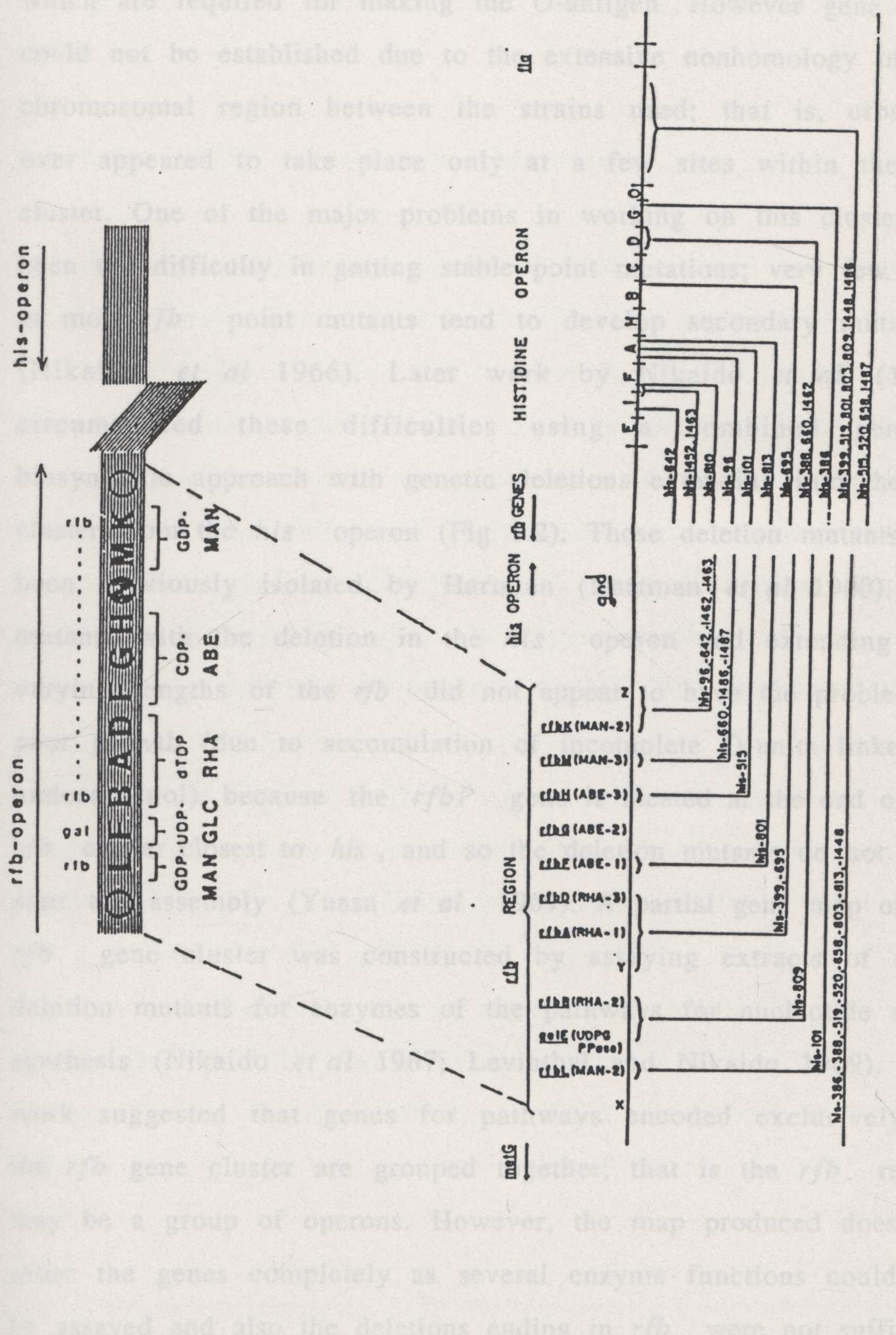
Transconjugant of crosses exchanging the *rfb* regions between *Salmonella* of different O groups usually results in a change of the O-antigen expressed (Mäkelä 1965, 1966; Nikaido *et al* 1966;

FIGURE 1.2

The *rfb* operon of *S. typhimurium*

The *rfb* operon of *S. typhimurium* and the nearby *his* operon are shown here. The arrows show the direction of transcription. The *rfb* cistrons are shown on the 'chromosome', the biosynthetic pathways in which their products participate are indicated underneath. Encircled cistrons determine monosaccharide transfers, of which *rfbN* has been definitively identified. MAN = mannose; GLC = glucose; RHA = rhamnose; ABE = abequose (from Mä kelä and Stocker 1984). The bottom segment shows the *his* - *rfb* deletions with respective *his* numbers (from Nikaido *et al* 1967). The *galE* (bottom segment) gene was later found to be *galF* (top segment). The locus 'Y' was mapped between *rfbA* and *rfbD*, while locus 'X' mapped to the left of *rfbL*.

Kochibe 1970; Kishi and Iseki 1973a, 1973b; Jousimies and Mäkelä 1974). Thus the *rfb* cluster must contain all the specific genes which are required for making the O-antigen. However gene order could not be established due to the extensive nonhomology of this chromosomal region between the strains used; that is, crossing-over appeared to take place only at a few sites within the *rfb* cluster. One of the major problems in working on this cluster has been the difficulty in getting stable point mutations; very few exist as most *rfb* point mutants tend to develop secondary mutations (Nikaido *et al* 1966). Later work by Nikaido *et al* (1967) circumvented these difficulties using a combined genetic-biochemical approach with genetic deletions extending into the *rfb* cluster from the *his* operon (Fig 1.2). These deletion mutants had been previously isolated by Hartman (Hartman *et al* 1960). The mutants with the deletion in the *his* operon and extending into various lengths of the *rfb* did not appear to have the problem of poor growth (due to accumulation of incomplete O-units linked to undecaprenol) because the *rfbP* gene is located at the end of the *rfb* cluster closest to *his*, and so the deletion mutants do not even start O-unit assembly (Yuasa *et al* 1969). A partial gene map of the *rfb* gene cluster was constructed by assaying extracts of these deletion mutants for enzymes of the pathways for nucleotide sugar synthesis (Nikaido *et al* 1967; Levinthal and Nikaido 1969). This work suggested that genes for pathways encoded exclusively by the *rfb* gene cluster are grouped together; that is the *rfb* region may be a group of operons. However, the map produced does not order the genes completely as several enzyme functions could not be assayed and also the deletions ending in *rfb* were not sufficient to order all the genes with respect to each other. The expression of



Kochibe 1970; Kishi and Iseki 1973a, 1973b; Jousimies and Mäkelä 1974). Thus the *rfb* cluster must contain all the specific genes which are required for making the O-antigen. However gene order could not be established due to the extensive nonhomology of this chromosomal region between the strains used; that is, crossing-over appeared to take place only at a few sites within the *rfb* cluster. One of the major problems in working on this cluster has been the difficulty in getting stable point mutations; very few exist as most *rfb* point mutants tend to develop secondary mutations (Nikaido *et al* 1966). Later work by Nikaido *et al* (1967) circumvented these difficulties using a combined genetic-biosynthetic approach with genetic deletions extending into the *rfb* cluster from the *his* operon (Fig 1.2). These deletion mutants had been previously isolated by Hartman (Hartman *et al* 1960). The mutants with the deletion in the *his* operon and extending into varying lengths of the *rfb* did not appear to have the problem of poor growth (due to accumulation of incomplete O-units linked to undecaprenol) because the *rfbP* gene is located at the end of the *rfb* cluster closest to *his* , and so the deletion mutants do not even start the assembly (Yuasa *et al* 1969). A partial gene map of the *rfb* gene cluster was constructed by assaying extracts of these deletion mutants for enzymes of the pathways for nucleotide sugar synthesis (Nikaido *et al* 1967; Levinthal and Nikaido 1969). This work suggested that genes for pathways encoded exclusively by the *rfb* gene cluster are grouped together; that is the *rfb* region may be a group of operons. However, the map produced does not order the genes completely as several enzyme functions could not be assayed and also the deletions ending in *rfb* were not sufficient to order all the genes with respect to each other. The expression of *rfb* genes in deletion mutants, where the *his* operon is fused to the *rfb* operon, suggests that the direction of transcription in the *rfb* gene cluster is toward the *his* operon (Levinthal and Nikaido 1969).

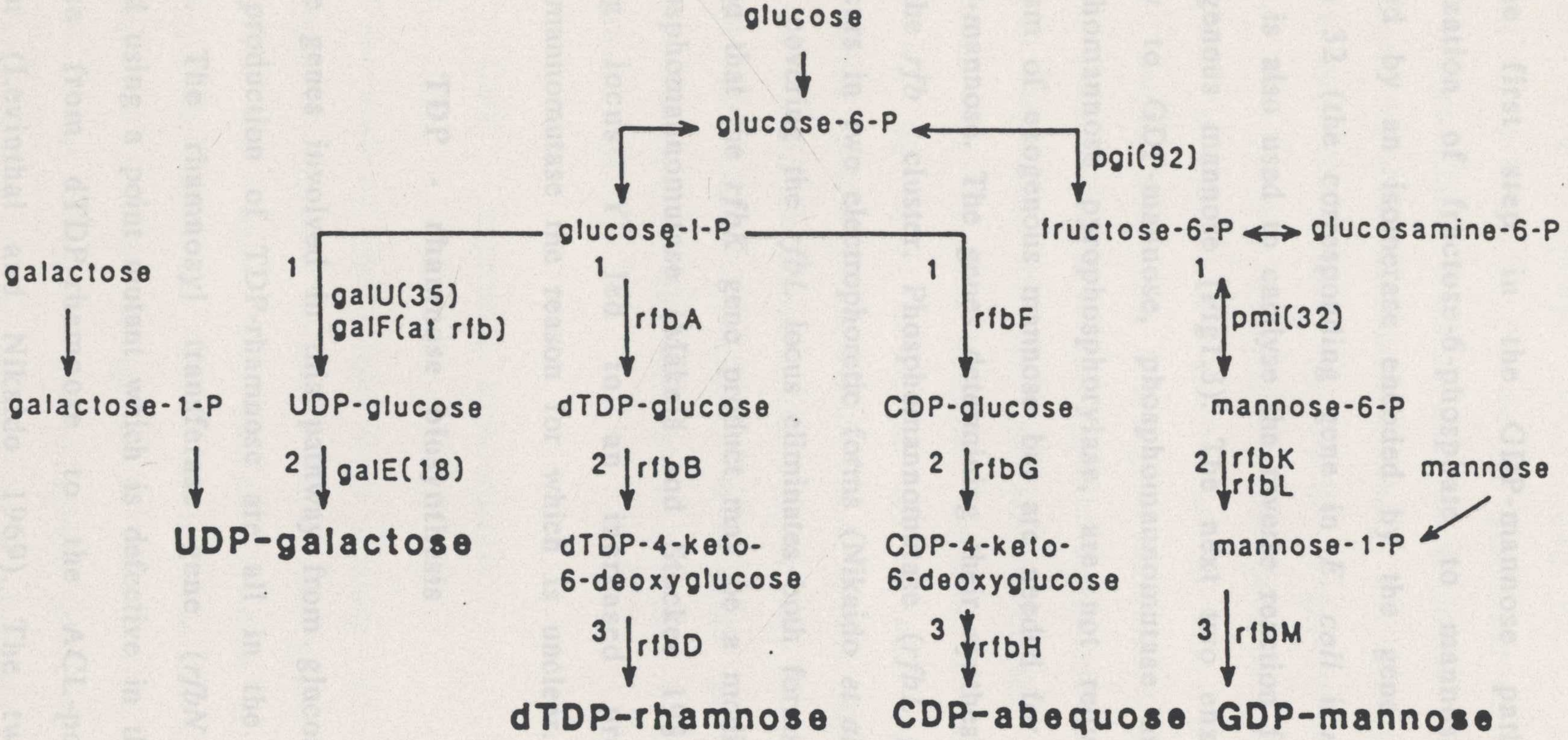
1.3.2.1 UDP - galactose biosynthesis

The genes required to synthesize UDP-galactose are located in the *gal* operon at position 18 outside the *rfb* region (Fig1.3). Two *rfb* genes, *galF* and *rfbP*, are involved in the biosynthesis of the galactose portion of the O-unit in *S. typhimurium*. The gene *rfbP* encodes the transferase that transfers the sugar from UDP-galactose to ACL, while *galF* participates in the synthesis of UDP-galactose in an indirect way. The *galF* gene product modifies the *galU* gene product (UDP-glucosepyrophosphorylase, a general housekeeping enzyme of the cell which may be involved in assisting in O-antigen synthesis) (Nakae 1971). This modification was detected by an altered electrophoretic mobility when compared to that form of the enzyme in a *his-rfb* deletion mutant lacking the *galF* gene (Nakae and Nikaido 1971a, 1971b; Nakae 1971). Mutants in this pathway are defective in synthesis of both the core and O-specific unit as galactose is present in both. It is hypothesized that this modification assists the use of the enzyme for purposes of O-unit synthesis, in some unexpected way (Nakae 1971).

FIGURE 1.3

Genetic determination of the monosaccharide constituents in the O unit of *Salmonella typhimurium*

The map position of genes outside the *rfb* cluster (map position 42) is indicated in parentheses. The biosynthetic steps in each pathway are numbered (from Mäkelä and Stocker 1984).


glucose
glucose-6-P
pgi(92)
glucose-1-P
fructose-6-P
glucosamine-6-P
galactose
1
galU(35)
galF(at rfb)
1
rfbA
1
rfbF
1
pmi(32)
galactose-1-P
UDP-glucose
dTDP-glucose
CDP-glucose
mannose-6-P
2
galE(18)
2
rfbB
2
rfbG
2
rfbK
rfbL
mannose
UDP-galactose
dTDP-4-keto-
6-deoxyglucose
CDP-4-keto-
6-deoxyglucose
mannose-1-P
3
rfbD
3
rfbH
3
rfbM
dTDP-rhamnose
CDP-abequose
GDP-mannose

1.3.2.2 GDP - mannose biosynthesis

The first step in the GDP-mannose pathway is the isomerization of fructose-6-phosphate to mannose-6-phosphate catalyzed by an isomerase encoded by the gene *pmi* at map position 32 (the corresponding gene in *E. coli* is *manA*). This enzyme is also used to catalyse the reverse reaction for metabolism of exogenous mannose (Fig1.3). The next two enzymes on the pathway to GDP-mannose, phosphomannomutase and guanosine diphosphomannose pyrophosphorylase, are not required for the catabolism of exogenous mannose but are needed for the synthesis of GDP-mannose. The genes determining their synthesis are located within the *rfb* cluster. Phosphomannomutase (*rfbL* gene product) also occurs in two electrophoretic forms (Nikaido *et al* 1967), and a deletion covering the *rfbL* locus eliminates both forms. It has been suggested that the *rfbK* gene product may be a modifying enzyme for phosphomannomutase (Mäkelä and Stocker 1984). Deletions covering locus 'Y' led to an increased production of phosphomannomutase the reason for which is unclear.

1.3.2.3 TDP - rhamnose biosynthesis

The genes involved in this pathway from glucose-1-phosphate to the production of TDP-rhamnose are all in the *rfb* cluster (Fig1.3). The rhamnosyl transferase gene (*rfbN*) has been identified using a point mutant which is defective in the transfer of rhamnose from dTDP-rhamnose to the ACL-polysaccharide precursor (Levinthal and Nikaido 1969). The two identified intermediates, dTDP-glucose and dTDP-4-keto-6-deoxyglucose, are also intermediates on the pathway to dTDP-aminodeoxygalactose (Matsuhashi and Strominger 1964), and were recently found (Lugowski *et al* 1983) as part of the glycolipid of enterobacterial common antigen (ECA) (Mäkelä and Mayer 1976). Point mutations in the genes for dTDP-rhamnose synthesis are not expected to be harmful to the cell because the transfer of galactose phosphate is reversible. The enzyme dTDP-glucose oxidoreductase (*rfbB* gene product) also occurs in two electrophoretic forms (Nikaido *et al* 1967) and deletions covering this region eliminate both forms. The reason for these different electrophoretic forms is not known and the gene(s) responsible for the modification of the dTDP-glucose oxidoreductase has not been identified. Deletions covering locus 'X' were shown to have increased levels of dTDP-rhamnose oxidoreductase activity (Nikaido *et al* 1967). Since in these deletions gene *rfbB* was also deleted, the activity in the 'X' mutants was thought to be based on a different structural gene determining a second enzyme with at least partial dTDP-rhamnose oxidoreductase activity.

1.3.2.4 CDP - abequose biosynthesis

CDP-abequose is synthesized from glucose-1-phosphate via a series of specific reactions (Fig1.3), and all the genes involved are located in the *rfb* cluster. The enzyme cytidine diphosphate-D-glucose pyrophosphorylase (*rfbF* gene product) adds cytidine diphosphate onto glucose-1-phosphate to form cytidine diphospho-D-glucose with the release of pyrophosphate (Nikaido *et al* 1967), which is the first step of the biosynthesis of abequose. This enzyme has been purified and characterized most fully in *Y* .

pseudotuberculosis type IV as a monomer (Rubenstein and Strominger 1974). The second step, the conversion of CDP-glucose to CDP-4-keto-6-deoxy-D-glucose, is catalyzed by the product of gene *rfbG* in *S. typhimurium*. This protein has been purified and characterized in *Y. pseudotuberculosis* type V, and found to be functional as a dimer (Matsuhashi *et al* 1966; Gonzalez-Porque and Strominger 1972). The final step is catalyzed by the *rfbH* gene product in *S. typhimurium* (Nikaido *et al* 1967), but was found to involve three proteins in *Y. pseudotuberculosis* (Matsuhashi and Strominger 1967), and later in *S. typhimurium* (Pape and Strominger 1969). CDP-4-keto-6-deoxy-D-glucose is converted to CDP-4-keto-3,6-dideoxy-D-glucose by two proteins E_1 and E_3 in *Y. pseudotuberculosis*. An *rfbH* mutant of *S. typhimurium* (Yuasa *et al* 1969) defective in protein E_1 was shown to produce Ra LPS indicating that abequose is essential for O-side chain polymerization. The E_2 protein reduces 4-keto-6-deoxy-D-glucose in various serotypes of *Y. pseudotuberculosis* using NADPH to provide energy required for its activity (Matsuhashi and Strominger 1967). In *S. typhimurium* the names *rfbI* and *rfbJ* were allocated to enzymes E_3 and E_2 respectively, because these were two names not yet allocated to *rfb* genes. The names *rfbH*, *rfbI* and *rfbJ* would in this instance reflect the order in which genes acted in the biosynthesis of CDP-abequose (P. Wyk, Ph.D. thesis, 1988).

1.3.2.5 CDP - paratose and CDP - tyvelose biosynthesis

The *rfb* genes of *S. typhimurium* (serogroup B) define the basic O-specific unit, in which the immunodominant part is abequose. The serogroups D and A differ from B in having the immunodominant sugars tyvelose or paratose in place of abequose (Lüderitz *et al* 1968). All three dideoxyhexoses share a common biosynthetic pathway: steps 1 and 2 as well as part of step 3 of the abequose pathway (Fig1.3). An enzyme E_2 would then be sufficient (although this has not been demonstrated) to convert the CDP-4-keto-3,6-dideoxyhexose to either CDP-abequose or CDP-paratose. Finally, paratose is converted by a single enzyme called CDP-paratose-2-epimerase to CDP-tyvelose (Matsuhashi 1966) (Fig 1.4). Mutation in a group D organism within the gene determining CDP-paratose-2-epimerase results in the synthesis of CDP-paratose instead of CDP-tyvelose (Uchida *et al* 1974; Sasaki and Uchida 1974). These mutants synthesize a smooth, group A-like O-antigen, indicating that tyvelose-transferase can transfer paratose in the O-unit assembly, and that this unnatural O-unit can be polymerized and transferred to the LPS core in a manner that is apparently fully efficient since the mutants are not sensitive to R-specific bacteriophages (P. H. Mäkelä, unpublished data). It was hypothesized that such mutants, altering the O group from D to A, have probably also occurred in nature; the simultaneous isolation of otherwise identical *Salmonella* with O-antigens 1, 9, 12 and 1, 2, 12 from a patient has been reported (Böhlck 1965).

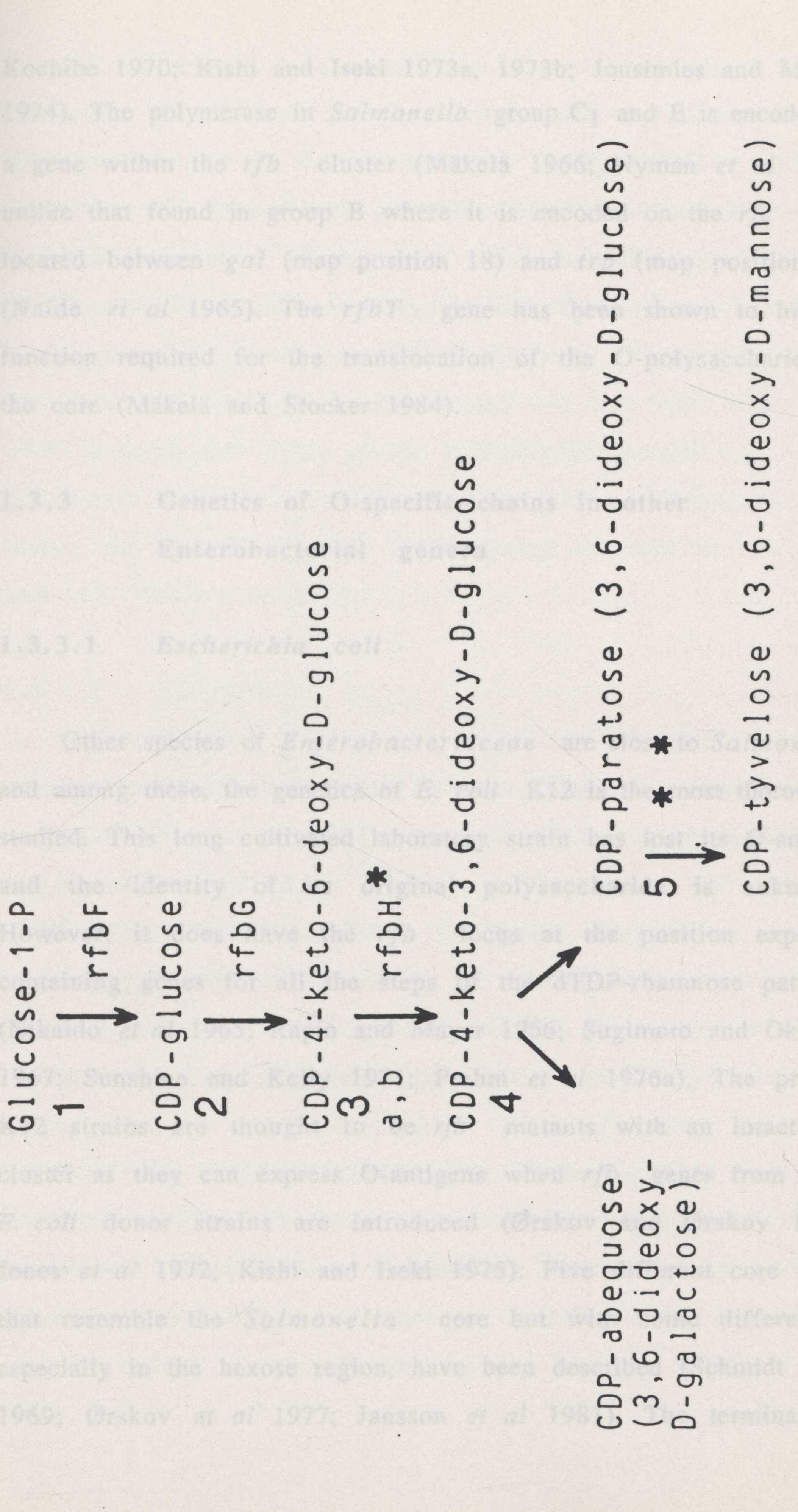
1.3.2.6 Other genes involved in O-specific chain biosynthesis

Recombination studies have shown that genes for all the O-unit transferases must be located in the *rfb* cluster in serogroup B, C_1, D, E, G, L, R and U (Mäkelä 1965; Mäkelä 1966; Nikaido *et al* 1966;

FIGURE 1.4

The biosynthesis of dideoxyhexoses present in O units of *Salmonella* of groups A, B and D

The biosynthesis of the dideoxyhexoses present in O units of *Salmonella*, and the genes of the synthesis as far as known. * = mutation shown to impair viability (Yuasa *et al* 1969); ** = mutation found to result in the incorporation of paratose instead of tyvelose into O polysaccharide and LPS (Uchida *et al* 1974) (from Mäkelä and Stocker 1984).


Glucose-1-P
1
rfbF
CDP-glucose
2
rfbG
CDP-4-keto-6-deoxy-D-glucose
3
a,b
rfbH*
CDP-4-keto-3,6-dideoxy-D-glucose
4
CDP-abequose (3,6-dideoxy-D-galactose)
CDP-paratose (3,6-dideoxy-D-glucose)
5
* *
CDP-tyvelose (3,6-dideoxy-D-mannose)

Kochibe 1970; Kishi and Iseki 1973a, 1973b; Jousimies and Mäkelä 1974). The polymerase in *Salmonella* group C_1 and E is encoded by a gene within the *rfb* cluster (Mäkelä 1966; Nyman *et al* 1979) unlike that found in group B where it is encoded on the *rfc* gene located between *gal* (map position 18) and *trp* (map position 34) (Naide *et al* 1965). The *rfbT* gene has been shown to have a function required for the translocation of the O-polysaccharide to the core (Mäkelä and Stocker 1984).

1.3.3 Genetics of O-specific chains in other Enterobacterial genera

1.3.3.1 *Escherichia coli*

Other species of *Enterobacteriaceae* are close to *Salmonella*, and among these, the genetics of *E. coli* K12 is the most thoroughly studied. This long cultivated laboratory strain has lost its O-antigen and the identity of its original polysaccharide is unknown. However, it does have the *rfb* locus at the position expected containing genes for all the steps of the dTDP-rhamnose pathway (Nikaido *et al* 1965; Rapin and Mayer 1966; Sugimoto and Okazaki 1967; Sunshine and Kelly 1971; Prehm *et al* 1976a). The present K12 strains are thought to be *rfb* mutants with an intact *rfa* cluster as they can express O-antigens when *rfb* genes from other *E. coli* donor strains are introduced (Ørskov and Ørskov 1962; Jones *et al* 1972; Kishi and Iseki 1975). Five different core types that resemble the *Salmonella* core but with some differences, especially in the hexose region, have been described (Schmidt *et al* 1969; Ørskov *et al* 1977; Jansson *et al* 1981). The terminal N-acetylglucosamine of the *Salmonella* core is missing from the *E. coli* core types R1, R3 and R4. In recombination experiments it was seen that *Salmonella* O-4,12 or O-9,12 O-polysaccharides could be attached to LPS in an *E. coli* O8:K27 strain of core type R1, without N-acetyl glucosamine (Kiefer *et al* 1976), as well as in a derivative of this O8:K27 strain which synthesized the K12-type core after receipt of the *rfa* genes from K12 (Prehm *et al* 1976b). Some of the *Salmonella* - *E. coli* O8 hybrids reacted with both anti-O8 and anti-O4,12 antisera and it was shown that the two specificities resided in separate molecules: the O8 specificity as long O-polysaccharide chains, the O-4,12 as single O-units, each attached to a separate core LPS (Kiefer *et al* 1976).

1.3.3.2 *Shigella*

In the intestinal pathogen *Shigella flexneri*, a number of O-antigenic serotypes and the structures of the corresponding O-polysaccharides are known (Simmons 1971; Kenne and Lindberg 1983). The genes determining the basic structure of the O-polysaccharide are located at the *rfb* locus analogous to the O-antigen genes of *E. coli* and *Salmonella* (Formal *et al* 1970). Whether or not the polymerization of the O-unit is also determined at *rfb* was not specifically tested. In *S. flexneri*, the genes responsible for enzymatic modifications of the O-antigen have been identified close to the proline locus (map position 6) on the chromosome. Some of these modifications are encoded by lysogenic viruses whose chromosomal sites are situated next to the *pro* genes (Formal *et al* 1970; Petrovskaya and Lincheva 1982).

All strains of *Shigella sonnei* exhibit 'phase I' and 'phase II' in culture. Only the phase I organisms are virulent and found to be smooth and stable (Kopecko *et al* 1980; Sansonetti *et al* 1980, 1981; Gamian and Romanowska 1982), whereas phase II organisms are unstable and rough with a complete core (like that of *E. coli* core type R1) but substituted with glucosamine at the branch heptose. The same core is present in phase I, the smooth form (S) in which the O-polysaccharide is a polymer of a disaccharide O-unit containing two unusual sugars, N-acetylaltrosaminuronic acid and 2-acetamido-4-amino-2,4,6-trideoxygalactose (Romanowska and Reinhold 1973; Kenne *et al* 1980). The phase I organisms carry a 120 Mdal plasmid which encodes the genes for O-antigen synthesis (Kopecko *et al* 1980; Sansonetti *et al* 1980, 1981). This plasmid is lost at a high rate in laboratory grown cultures resulting in phase II avirulent bacteria. *S. dysenteriae* is also known to have a 6 Mdal plasmid encoding one or more functions necessary for O-antigen expression (Watanabe and Timmis 1984).

1.3.3.3 Other enteric bacteria

The LPS of *Proteus mirabilis* strains contains a variety of constituents in the O-polysaccharide including amino acids and ribitol phosphate (Jann and Westphal 1975). R mutants, such as a deep rough form without heptose, have been studied biochemically (Sidorczyk *et al* 1983), but no genetic data are available. In both *Y. pseudotuberculosis* and *Y. enterocolitica* the regulation of LPS biosynthesis has been reported to depend on growth temperature. Plasmids are associated with the pathogenic potential of the *Yersiniae* and expression of a number of plasmid-encoded properties is also controlled by growth conditions (Gemski *et al* 1980a, 1980b; Ferber and Brabaker 1981).

1.4 O-specific chain modifications

The most common modifications are O-acetylation or glucosylation of specific residues in the O units, or changes in the link formed by the polymerase, although several other substituents are also found (Lüderitz *et al* 1968; Kenne and Lindberg 1983). The modifications may be a constant part of the LPS structure in one species or a clone of bacteria; in this form they are responsible for variation between serotypes within one O group. They may also cause variation within a single chain, depending on rapid on-off variation in the state of a gene; this is termed form variation. Similar modifications can be caused by genes on bacteriophages; in these cases they are called antigenic variation. Furthermore, the modifications are usually not stoichiometric and so cause heterogeneity within the LPS of even one bacterium (Hellerqvist *et al* 1969a; Lindberg *et al* 1978). The modifications are not essential for the completion of LPS or O-polysaccharide synthesis and the stage at which they occur is not defined.

1.4.1 O-acetylation

In *S. typhimurium* O-acetylation of the abequose residue (recognised antigen factor 5) is determined by the gene *oafA* at position 43, close to *rfb* but clearly separable from it (Mäkelä 1966; Johnson *et al* 1966; Kishi and Iseki 1973b). Even single O units are modified resulting in the 4,5,12 specificity of SR forms (Naide *et al*

1965). Tyvelose containing O units are not acetylated in natural strains of serogroup D or in hybrids combining the group D *rfb* with *oafA* from group B (Mäkelä 1965; Kishi and Iseki 1973a). Factor 5-negative variants (known as var. copenhagen) are known among *Salmonellae* of group B isolated from natural sources (Kauffmann 1935, 1966) and in laboratory isolated mutants (Iseki and Sasaki 1952; Smith and Stocker 1962). O-acetylation at C-2 of the rhamnosyl residue was found in hybrids between group B and group C_2 (O-antigens 6,8) strains which synthesize the group B-type O polysaccharide (Hellerqvist *et al* 1972; Ruden and Mäkelä 1974), thought to be mediated by modifying enzymes of group C_2; the rhamnose in the O unit of group C_2 is partially acetylated (Hellerqvist *et al* 1970). The specificity of the enzyme is low since the rhamnosyl residues in group B and C_2 occur in different linkages. A corresponding O-acetylation of the galactosyl residue of the O unit is present in group E; lysogenization by bacteriophage ε_{15} or g_{341} prevents this acetylation, and a phage mutant lost this blocking capacity (Robbins and Uchida 1965; Hagiwara *et al* 1966).

1.4.2 Glucosylation

In *S. typhimurium* glucosylation of the galactosyl residue in the O unit can occur through O-4 or O-6; resulting in two completely different antigen factors called 12_2 and 1 respectively (Stocker *et al* 1960; Tinelli and Staub 1960). The glucosyl donor in both cases is glucosyl-phospho-undecaprenol (Nikaido *et al* 1971; Nikaido and Nikaido 1971; Takeshita and Mäkelä 1971), and the polymerized O-polysaccharide-undecaprenol serves as an acceptor. This transfer is not affected by the dideoxyhexose substituent of the mannose as both group B and D are glucosylated to an equal extent (Nikaido *et al* 1971). Glucosylation through O-4 of the O-unit galactose is determined by chromosomal genes in the *oafR* - *oafE* cluster close to *purE* at position 11, with *oafE* involved in the synthesis of glucosyl-phospho-undecaprenol (Mäkelä and Mäkelä 1966; Mäkelä 1973). Although glucosylation is thus not phage-determined, it makes the LPS resistant to adsorption of the phages 9NA and KB1 (Wilkinson *et al* 1972; P.H.Mäkelä, unpublished data). The O-4 glucosylation in group E *Salmonella* is coded for by the phage ε_{34} gene and is recognized as antigenic factor 34 (Uetake *et al* 1955; Iseki and Hamano 1958). The O-6 glucosylation however is mediated by phage P22 genes (Iseki and Kashiwagi 1955, 1957; Zinder 1957). In group E_4 (O-antigens 1,3,19) the factor 1 (glucosyl-1-6-galactose) is thought to be determined by a chromosomal gene called *oafC*, close to *purE* (Stocker *et al* 1960; Plosila and Mäkelä 1972; Kishi and Iseki 1974).

1.4.3 Form variation

Variable expression of two antigen factors (1 and 12_2) has been described by Kauffmann in numerous *Salmonella* strains (Kauffmann 1941; Kauffmann and Rhode 1961). This variability is reversible (from one form to the other) and is detected by altered reactivity with antisera against these factors. Several other instances of variability have been described including those of the 6_1 factor in group C_1(Kauffmann 1937; Levine and Frisch 1936; Edwards 1945) and the 27 factor (change in O unit polymerization) in groups B and D (for reviews, see Mäkelä and Stocker 1984; Mäkelä and Rhen 1985). The bacteria can switch from one form (+

or - in respect to the particular antigen) to the other, and back again. This usually occurs at relatively high frequency (once per 10^2-10^4 cell divisions). Only one particular factor is affected; for example, only factor 1 and not 4 or 12 in *S. typhimurium* lysogenic for P22 (Almon and Stovall 1939; Hayes 1947; Mäkelä and Mäkelä 1966). Chemical studies have given quantitative estimations of LPS variation in a *Salmonella* strain lysogenic for P27. In such LPS, one third to one half of the linkages between the O units were found to be 27-type and, at the same time, 71-85% of the bacteria related with antifactor-27 serum; that is, there was little or no repression of bacterial polymerase (Lindberg *et al* 1978). However, in *S. bredeney*, the expression of factor 27 has been described as stable (Staub and Bagdian 1966). In all cases form variation was found to be the result of LPS modification and was determined by either phage or chromosomal genes outside *rfb* (such as *oafR*, the factor 12_2 in *S. typhimurium*) (Mäkelä and Mäkelä 1966; Mäkelä 1973). The similarity to flagellar phase variation (Kutsukake and Iino 1980; Silverman and Simon 1980) on a phenotypic and genetic level led to suggestions that the basic molecular mechanism for both may be the same.

1.4.4 Antigenic conversion

The O-polysaccharides act as receptors for many bacteriophages on attachment to the cell. A change in the O-antigenic structure of a strain, such as that caused by lysogenic conversion, is expected to alter its phage sensitivity (Lindberg 1977). This change in phage sensitivity is due to alterations of one or other of the monosaccharides between which the tail fibre enzyme hydrolyzes the polysaccharide. Presumably all these phages bind to the long O-polysaccharide chains and gain access to the surface of the bacterial outer membrane by hydrolyzing the polysaccharide that covers it (Lindberg 1977).

1.5 Polysaccharide components other than LPS

1.5.1 Vi antigen

Salmonella typhi is one of the few *Salmonella* that have a capsular polysaccharide, the Vi antigen; this is a polymer of O- and N-acetylated galactosaminuronic acid (Clark *et al* 1958). Two gene loci *viaB* (at about 92 map position) and *viaA* (46 map position) have been located by conjugational transfer to *S. typhimurium*, which does not naturally have a capsule (Johnson *et al* 1965, 1966). *viaA* maps close to the *rfb* region.

1.5.2 T antigen

A polysaccharide antigen called T_1 (a transient form between S and R forms) has been described in group B organisms *S. paratyphi B* and *S. typhimurium* (Kauffmann 1956); it may be present together with O-antigen but remains cryptic because of the stronger expression of the latter (Sarvas 1967). T_1 antigen consists of polymers of ribose and galactose in the unusual furanose form (Berst *et al* 1969, 1971).

The locus *rft*, at approximate position 15, determines T_1 antigen production. By genetic methods the T_1 determinant genes can be transferred to other strains, including smooth ones. In such recombinants, both O and T_1 chains can be linked to the core LPS, both requiring the complete core and a functional *rfaL* gene for attachment, (Sarvas 1967). Strains with the T_1 antigen become T_1-negative at a high frequency by mutation at or near *rft* but the reverse change does not occur with detectable frequency. By recombination studies it has been shown that *rfb* and *rfc* genes are not needed for T_1 synthesis but a functional *rfe* gene(s) is required for reasons that are not yet understood (Sarvas 1967; P. H. Mäkelä, unpublished data). Another T antigen, T_2 also attaches to the LPS core (Bruneteau *et al* 1974). Its chemical determinant appears to be an incompletely characterized polymer of N-acetylglucosamine which depends for expression on the *rfu* genes (Valtonen *et al* 1976; Sarvas *et al* 1976).

1.5.3 Colanic acid

Colanic acid is not a type-specific K antigen but, since it surrounds the bacterial cell as a thick slime layer, it functions as a capsular antigen. All *E. coli, Salmonella* and *Enterobacter cloacae* strains can synthesize this surface polysaccharide (which is antigenically very similar in all these strains). It consists of hexasaccharide repeating units and may vary in its O-acetylation and in the pyruvyl substituents peculiar to this molecule (Sutherland 1977). In wild type strains, the synthesis of large amounts of M antigen can be turned on by environmental conditions such as low temperature, high concentration of salts and excess of fermentable sugars (Markovitz 1977). The biosynthetic pathway for colanic acid synthesis in *E. coli* K12 has been postulated by Markovitz (1967b), and is thought to involve 11 biosynthetic enzymes, only four of which have been definitively identified. *E. coli* K12 is not highly mucoid and mutations at a number of loci eg. *lon* (*capR*), *capS*, *capT* (Markovitz 1977) dramatically increase the production of colanic acid. It has also been shown that some of the enzymes eg. *manA*, *galU* gene products and the enzymes encoded by the *gal* operon genes involved in colanic acid biosynthesis are present in increased amounts in *lon* strains (Markovitz *et al* 1967a). Synthesis of colanic acid is regulated by the *lon* gene product, which seems to function by limiting, by proteolysis, the *rcsA* gene product which acts as a positive regulator (Torres-Cabassa and Gottesman 1987). A series of *non* mutants (non-mucoid in *lon* strains) have been isolated (*non* -1 to 4, Lieberman *et al* 1970; *non* -9, Radke and Siegel 1971); the *non* -1 to 4 mutants have not been mapped but the *non* -9 was shown to be *his* -linked. The *non* -9 mutation was found due to its ability to mutate phage T7 resistance (Radke and Siegel 1971). This was explained as being due to a block in capsule formation in *non*-9 mutants. Recently, a new locus for exopolysaccharide overproduction (*ops*) has been mapped at 62 min on the *E. coli* linkage map (Zinkewich-Peotti and Fraser 1988). The polysaccharide produced was shown to be similar to colanic acid.

1.5.4 Enterobacterial common antigen (ECA)

The enterobacterial common antigen (ECA) is present in essentially all bacteria belonging to the family *Enterobacteriaceae* (Mäkelä and Mayer 1974; Mayer and Schmidt 1979; Ramia *et al*

1982). The serological specificity of ECA is determined by an amino sugar-containing heteropolysaccharide. It was shown previously that this heteropolysaccharide consisted of a linear chain of 1,4-linked *N*-acetyl-D-glucosamine (Glc NAc) and *N*-acetyl-D-mannosaminuronic acid (Man NAcUA) residues (Männel and Mayer 1978). More recently, it has been demonstrated that 4-acetamido-4,6-dideoxy-D-galactose (Fuc4 NAc) is also a component of ECA and that component sugars are arranged in a linear chain of trisaccharide repeat units (Lugowski *et al* 1983; Barr and Rick 1987). However, the identity of the reducing terminal sugar of the polymer has not been established.

ECA generally occurs in two forms; the haptenic form and the immunogenic form (Mayer and Schmidt 1979; Ramia *et al* 1982). The immunogenic form is covalently linked to the LPS core region of certain rough mutants with complete R1 or R4 cores, but lacking O-antigenic side chains. In contrast, the haptenic form is not linked to LPS and occurs in both rough and smooth strains (Kuhn *et al* 1983, 1984). Recently a lipid-free form of ECA was obtained from *Shigella sonnei* phase I and characterized as a mixture of cyclic heteropolysaccharides (Dell *et al* 1984). It has been proposed that the amino sugar chain of ECA might be assembled on a lipid carrier molecule analogous to C_{55}-isoprenoid alcohol involved in O-antigen synthesis (Mäkelä *et al* 1970; Mäkelä and Mayer 1974). Previous studies have shown that Glc NAc-pyrophosphoryl-undecaprenol functions as an early intermediate in ECA synthesis and that the antibiotic tunicamycin abolishes ECA synthesis *in vivo* (Rick *et al* 1985). More recently, an *in vitro* system has been developed in order to define the individual reactions involved in ECA biosynthesis (Barr and Rick 1987).

The gene clusters *rfe* and *rff* are both involved in the biosynthesis of ECA (Mäkelä and Mayer 1974; Mäkelä *et al* 1976). Both these gene clusters are linked to *ilv* at min 83 and 85 on *Salmonella* and *E. coli* chromosomes, respectively (Schmidt *et al* 1976; Mayer and Schmidt 1979; Meier and Mayer 1985). The *rff* genes are involved in the synthesis of UDP-Man NAc UA, the donor of Man NAc UA residues in ECA synthesis (Lew *et al* 1978). In some O groups of *Salmonella* spp. and *E. coli* the *rfc* genes are required for the synthesis of both the ECA and O-antigenic polysaccharide (Mäkelä *et al* 1970; Mäkelä and Mayer 1976). Genes *rfbA* and *rfbB* located in the *rfb* region of *S. typhimurium* are involved in the synthesis of dTDP glucose, an intermediate in the synthesis of rhamnose (O-antigen) and the ECA componenet 4-Fuc NAc.

1.6 Diversion into taxonomy of *Salmonella*

In the genus *Salmonella*, the species concept has evolved in to four overlapping phases, each of these phases relying on available methodology and knowledge, stressing (i) chemical evidence (ii) antigenic specificities (iii) biochemical properties and (iv) DNA relatedness (Le Minor and Popoff 1987). It has been demonstrated by DNA relatedness studies that all *Salmonella* serovars form a single DNA hybridization group with subgroups, based on thermal stability of hybrid DNA (Crosa *et al* 1973; Stoleru *et al* 1976; Le Minor *et al* 1986). They are considered to be members of a single species *Salmonella enterica* (Ewing 1986; Le Minor and Popoff

1987). However, the *Salmonella* typing community have used species epithets for each serotype (over 1,000 species epithets have been so used). Since this study relates to the genetic basis of variation in *Salmonella*, it is important to note that the strains of *S. typhimurium, S. paratyphi A*, and *S. typhi* used, are in reality not only members of the same species but also of the same subspecies, *enterica,* and have been proposed as of serovars Typhimurium, Paratyphi A and Typhi of *S. enterica.*

1.7 Aims of the project

The O-antigens of *Salmonella* vary greatly between strains. This variation or polymorphism has been used as a basis for serotyping and is well documented in the genus *Salmonella* . The O-antigens from strains of serogroups D (*S. typhi*) and A (*S. paratyphi A*) differ from that of B (*S. typhimurium*) essentially in the nature of the dideoxyhexose side chain sugar. Each of these O-antigens has a dideoxy hexose (tyvelose, paratose or abequose) and genes conferring this specificity are encoded in the *rfb* gene cluster. The aims of this thesis were to clone the *rfb* gene cluster from *S. typhi* Ty21a (group D) and *S. paratyphi A* IMVS1316 (group A) in order to compare them with the corresponding region of *S. typhimurium,* which has been cloned in our laboratory (Brahmbhatt *et al* 1988). The aims also involved localisation of the regions of homology and nonhomology, analysis of the variable region at the base sequence level, identification of the genes concerned with the biosynthesis of tyvelose and paratose, and possible determination of the evolutionary derivation of these serogroups.

CHAPTER 2

Materials and Methods

2.1 Growth media

Double strength Difco nutrient broth (Difco 0003, 16 g/L; NaCl, 5 g/L) was the general growth medium, referred to as "nutrient broth". It was supplemented with agar (Difco 0045, 15 g/L) to obtain nutrient agar. Luria broth (Difco tryptone, 10 g/L; Difco yeast extract, 5 g/L; NaCl, 5 g/L; pH adjusted to 7.2) was used as enriched liquid medium and to it was added Difco agar (15 g/L) to obtain Luria agar. 2X YT liquid medium (Difco yeast extract, 10 g/L; tryptone, 16 g/L; NaCl, 5 g/L) was used for growth of M13 bacteriophage. M13 minimal medium was prepared as described by Messing (1983) and was solidified with the addition of agar (Difco, 0140-10) to 16 g/L. Carbon source for minimal media was glucose, 5 mg/ml. Growth factors were added to a final concentration of 20 μg/ml, except vitamin B1 (thiamine) which was added to a final concentration of 1 μg/ml. Antibiotics were added to nutrient agar in the following concentrations: ampicillin, 25 μg/ml; chloramphenicol, 25 μg/ml; cycloserine, 200 μg/ml; kanamycin, 50 μg/ml; spectinomycin, 20 μg/ml; streptomycin, 100 μg/ml (for *rpsL* chromosome resistance) and 10 μg/ml (plasmid-borne resistance) and tetracycline, 16 μg/ml.

2.2 Reagents

Reagents were obtained from the following sources: acrylamide and bisacrylamide, BioRad; lysozyme, Sigma; sodium dodecyl sulphate (SDS), prod. no. 30175 and 44244, BDH; formamide, urea, Merck; Trizma base, Sigma; Methionine assay medium, Bacto, prod. no. 0423-15.2; cesium chloride, KBI; agarose, Seakem Inc.; nitrocellulose, Schleicher and Schuell; restriction endonucleases, Bal31, DNaseI, linkers, DNA polymerase I, deoxynucleotide triphosphates and dideoxynucleotide triphosphates, Boehringer-Mannheim, New England Biolabs and Pharmacia; *In vitro* λ packaging kit, Promega Biotech; large fragment of DNA polymerase I (Klenow), M13 primer, T_4 ligase and dideoxy sequencing kit, BRESA; ampicillin, Beecham; chloramphenicol and pronase, Calbiochem; kanamycin sulphate, Sigma; streptomycin sulphate, Glaxo; tetracycline, Gist-Brocade; L-[^{35}S] methionine (800 Ci/mmol), Amersham; [α-^{32}P] dCTP (2000 Ci/mmol), BRESA. All other chemicals were of analytical grade.

2.3 Bacterial strains, bacteriophages and plasmids

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

2.4 DNA techniques

2.4.1 Small scale isolation of plasmid DNA

TABLE 2.1

Bacterial strains used in this thesis

Strain	Characteristics	Source/Reference
Salmonella typhimurium LT2	*hsdL6 trpC2 nmlH1b fla-66 H2-enx rpsL120 xyl-404 lv-452 metE551 metA22 hsdA29*	I.Beacham
E246 (P9120)	*leu hsdL*(r^-m^+) *galE rpsL120 ilv452metE551 hsdA*(r^-m^+) *hsdB*(r^-m^+)*trpD2 metA22*	Bullas
P9029	Δ(*his -metG*)-338	H. Nikaido
Salmonella typhimurium (M2)		S. Dixon
Salmonella typhimurium (M5)		S. Dixon
Salmonella stanley (M15)		S. Dixon
Salmonella budapest (M16)		S. Dixon
Salmonella derby (M17)		S. Dixon
Salmonella agona (M117)		ICPMR
Salmonella heidelberg (M118)		ICPMR
Salmonella bredeney (M120)		ICPMR
Salmonella hessarek (M121)		ICPMR
Salmonella derby (M122)		ICPMR
Salmonella saintpaul (M123)		ICPMR

Salmonella typhimurium (M228)		J. Taplin
Salmonella paratyphi A IMVS1316 (M8)		S. Dixon
Salmonella paratyphi A (M107)		ICPMR
Salmonella paratyphi A (M108)		ICPMR
Salmonella paratyphi A (M109)		ICPMR
Salmonella paratyphi A (M110)		ICPMR
Salmonella paratyphi A (M111)		ICPMR
Salmonella paratyphi (M217)		J. Taplin
Salmonella paratyphi (M218)		J. Taplin
Salmonella paratyphi (M219)		J. Taplin
Salmonella nitra (M354)		L. Le Minor
Salmonella kiel (M355)		L. Le Minor
Salmonella paratyphi A var *durazzo* (M356)		L. Le Minor
Salmonella typhi Ty2 (M18)	*galE*	Germanier and Furer(1975)
Salmonella enteritidis (M125)		ICPMR
Salmonella dublin (M126)		ICPMR
Salmonella panama (M128)		ICPMR
Salmonella javiana (M129)		ICPMR
Salmonella victoria (M212)		J. Taplin
Salmonella eastbourne (M214)		J. Taplin
Salmonella enteritidis (M224)		J. Taplin

Escherichia coli K12		
C600	*thr1 leu6 tonA1 lacY1 supE44 thi1*	W.Arber
DH1 (P3154)	*supE44 gyrA96 recA endA1 thi-1 hsdR17 relA1*	D. Hanahan
JM101 (P2398)	*supE thi D(lac-proAB)* [F'*traD96 proAB lacI*^{q}*Z DM15*]	Yanish-Perron *et al.* (1985)
DS410(P2039)	*azi tonA lacY minA minB rpsL xyl mtl thi ara*	J. Reeve

The M and P numbers when in brackets are laboratory numbers.

TABLE 2.2

Plasmids used in this study

Cloning Vectors	Charateristics	Source/Reference
pBR325	amp^r tet^r cml^r	Prentki *et al* (1981)
pcos2EMBL	kan^r tet^r	Poustka *et al* (1984)
pGB2	str^r spe^r	Churchward *et al* (1984)
pHC79	amp^r tet^r	Honn and Collins (1980)
pPR328	cml^r	Quigley and Reeves (1987)
pUC8	amp^r	Vieira and Messing (1982)
pUC9	amp^r	Vieira and Messing (1982)
pUC18	amp^r	Norrander *et al* (1983)
pUC19	amp^r	Norrander *et al* (1983)
Bacteriophage M13 vectors		
M13mp18		Norrander *et al* (1983)
M13mp19		Norrander *et al* (1983)

Plasmids	
pPR429	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. paratyphi A* IMVS1316 (Fig 3.2)
pPR430	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. paratyphi A* IMVS1316 (Fig 3.2)
pPR431	amp^r, tet^r; vector, pBR325; 15.7 kb EcoRI fragment of *S. paratyphi A* IMVS1316 cloned from pPR430
pPR432	cml^r; vector, pPR328; 11.0 kb EcoRI *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR430
pPR434	amp^r, tet^r; vector, pBR325; EcoRI (6.8) to EcoRI(9.6) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR430
pPR436	amp^r, tet^r; vector, pBR325 EcoRI(14.2) to EcoRI(16.5) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR430
pPR437	amp^r, tet^r; vector, pBR325; EcoRI(12.4) to EcoRI(14.2) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR430
pPR440	amp^r, tet^r; vector, pBR325; 1.2 kb EcoRI *rfb* fragment of*S. paratyphi A* IMVS1316 cloned from pPR430
pPR441	amp^r, tet^r; vector, pBR325; 0.94 kb EcoRI *rfb* fragment of*S. paratyphi A* IMVS1316 cloned from pPR430
pPR443	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)

pPR444	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR445	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR446	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR447	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR448	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR449	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR450	amp^r; vector, pHC79; with cloned *rfb* DNA from *S. typhi* Ty21a (Fig 3.2)
pPR505	amp^r; vector, pUC19; EcoRI (0.0) to EcoRI (6.8) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR429
pPR506	kan^r; vector, pcos2EMBL; EcoRI (0.0) to EcoRI (6.8) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR429
pPR507	Sm^r,Spec^r; vector, pGB2; EcoRI (0.0) to EcoRI (9.57) *rfb* fragment of *S. typhimurium* cloned from pPR301 (Brahmbhatt *et al* 1988)
pPR508	Sm^r,Spec^r; vector, pGB2; same as pPR507 but insert is in opposite orientation
pPR615	amp^r; vector, pUC19; KpnI (2.47) to KpnI (14.55) *rfb* fragment cloned from *S. paratyphi A* IMVS1316

pPR616	amp^r; vector, pUC19; KpnI (2.47) to KpnI (8.95) *rfb* fragment cloned from *S. typhi* Ty21a
pPR617	amp^r; vector, pUC9; EcoRI(14.2) to EcoRI (16.5) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR436
pPR618	amp^r; vector, pUC9; EcoRI (8.6) to EcoRI (10.9) *rfb* fragment of*S. typhi* Ty21a cloned from pPR443
pPR619	amp^r; vector, pUC9; with KpnI (2.47) to EcoRI(10.9) *rfb* DNA of *S. typhi* Ty21a
pPR620	amp^r; vector, pUC9; EcoRI(8.6) to HindIII(9.7), HindIII cutdown of pPR618
pPR652	amp^r; vector, pUC9; EcoRI(8.6) to BglII(10.0), BamHI-BglII cutdown of pPR618
pPR653	amp^r; vector, pUC18; EcoRI(8.6) to EcoRI(10.9), *rfb* fragment of *S. typhi* Ty21a cloned from pPR618 in correct orientation for expression
pPR654	amp^r; vector, pUC18; EcoRI(8.6) to EcoRI(10.9), *rfb* fragment of*S. typhi* Ty21a cloned from pPR618 in opposite orientation to pPR653
pPR655	amp^r,tet^r; vector, pUC18; 2.8 kb EcoRI fragment from pcos2EMBL cloned into pPR653
pPR656	amp^r,tet^r; vector, pUC18; 2.8 kb EcoRI fragment from pcos2EMBL cloned into pPR654
pPR657	amp^r; vector,pUC18; HindIII(9.7) to EcoRI (10.9), HindIII cutdown of pPR653

pPR658	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR659	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR660	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR663	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR665	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR666	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR667	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPr670	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR673	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR675	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR677	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR680	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR681	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)
pPR681	amp^{r}; vector, pUC18; Bal31 deletion derivative of pPR655 (Fig 4.3)

pPR784	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR786	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR787	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR791	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR793	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR794	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR796	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR797	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR798	amp^r; vector, pUC18; Bal31 deletion derivative of pPR656 (Fig 4.3)
pPR816	vector, M13mp18; EcoRI (14.20) to BglII (15.60) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR617
pPR848	amp^r; vector, pUC18; EcoRI (9.6) to EcoRI (12.4) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR434
pPR849	vector, M13mp18; KpnI(14.55) to EcoRI(16.50) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR617

pPR850	vector, M13mp18; PstI(14.80) to EcoRI(16.50) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR617
pPR851	vector, M13mp19; EcoRI(14.20) to BglII(15.60) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR617
pPR852	vector, M13mp19; BglII(15.60) to EcoRI(16.50) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR617
pPR893	vector, M13mp18; NruI(11.10) to EcoRI(12.40) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR848
pPR894	vector, M13mp18; EcoRI(9.60) to NruI(11.10) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR848
pPR895	vector, M13mp19; NruI(5.50) to HpaI(5.80) *rfb* fragment of *S. paratyphi A* IMVS1316 cloned from pPR505
pPR908	amp^r; vector, pUC18; Bal31 deletion derivative of pPR848
pPR909	amp^r; vector, pUC18; Bal31 deletion derivative of pPR848
pPR927	cml^r; vector, pPR328; EcoRI(8.60) to EcoRI(10.90) *rfb* DNA fragment of *S. typhi* Ty21a cloned from pPR618 in correct orientation for expression
pPR928	cml^r; vector, pPR328; EcoRI(8.60) to EcoRI(10.90) *rfb* DNA fragment of *S. typhi* Ty21a cloned from pPR618 in opposite orientation to pPR927

pPR939	cml^{r}; vector, pPR328; EcoRI(8.60) to BglII (10.0) *rfb* DNA fragment of *S. typhi* Ty21a cloned from pPR618
pPR948	amp^{r}; vector, pUC9; 11.0 kb EcoRI fragment of *S. typhi* Ty21a cloned from pPR443

Small scale isolation of plasmid DNA was performed by the following two methods:

a) Large numbers of clones were screened by the method of Ish-Horowicz and Burke (1981). Although the method was very quick it was found that long storage resulted in partial or complete degradation of the plasmid DNA. The plasmid DNA isolated by this method was therefore extracted once with phenol/ether and precipitated with ethanol.

b) A better quality plasmid DNA was isolated by the Triton X-100 cleared lysate method as described by Kahn *et al* (1979).

2.4.2 Cesium Chloride (CsCl) purification of plasmid DNA

Large scale purification of plasmid DNA was carried out by the two-step cesium chloride gradient method of Garger *et al* (1983).

2.4.3 Isolation of chromosomal DNA

A modified method of Nakamura *et al* (1979) was used to isolate high molecular weight chromosomal DNA. A 20 ml overnight culture was centrifuged and the cells were washed in 10 ml of TES buffer (50 mM Tris-HCl [pH 8.0], 5 mM EDTA, 50 mM NaCl). The pellet was resuspended in 2 ml of 25% sucrose, 50 mM Tris-HCl [pH 8.0] and after adding 1 ml of 10 mg/ml lysozyme in 0.25 M EDTA (pH 8.0), the mixture was incubated on ice for 20 minutes. Lysis was obtained after the addition of 0.75 ml of TE buffer (10 mM Tris-HCl [pH 8.1], 1 mM EDTA) and 0.25 ml of the lysis solution (5% sarkosyl, 50 mM Tris-HCl [pH 8.0], 62.5 mM EDTA). Pronase (10 mg) was added to this mixture and incubated at 56° C for 1 hour. The solution was extracted three times with TE saturated phenol and twice with diethyl ether at 4° C. DNA was precipitated by adding 8 ml of cold 95% ethanol. The precipitated DNA was washed twice in 70% ethanol, dried under vacuum and dissolved in 1 ml of TE buffer (Tris-HCl, 10 mM; EDTA, 1 mM; pH 7.4). If the DNA was not dissolved completely, it was heated at 56° C for 20 to 30 minutes.

2.4.4 DNA quantitation

DNA concentration was determined by measurement of optical density (OD) at 260nm, assuming a OD_{260} of 1.0 corresponds to 50 µg DNA/ml (Maniatis *et al* 1982).

2.4.5 Restriction enzyme digestions

DNA digestions were usually performed using low salt, medium salt and high salt buffers as described by Maniatis *et al* (1982) and in cases where these did not apply, the buffer was made according to the manufacturer's specifications.

Usually 0.1 to 1 µg of the plasmid and chromosomal DNA were digested with 2 units and 10 to 20 units of enzyme respectively in a total volume of 20 µl at 37° C in a 1.5 ml Eppendorf tube. All digestions were terminated by heating the tubes at 65° C for 10 minutes. Double digests were carried out using the lower salt concentration requiring enzyme first, terminating the reaction and carrying out the second digestion after adjusting the salt concentration. Prior to agarose gel electrophoresis 2 μl of tracking dye (10X: 0.6% bromophenol blue, 15% ficoll, 100 μg/ml RNase) was added to the sample.

2.4.6 Agarose gel electrophoresis

Analysis of restriction enzyme digests of DNA was done by electrophoresis (7.5 V/cm) through horizontal 0.6% to 1.5% agarose gels in 0.5 X TBE buffer (10X: 0.89 M Tris; 0.89 M Borate; 0.025 M EDTA). Preliminary and rapid restriction enzyme analysis was performed using minigels, as described by Maniatis *et al* (1982). The gels were stained with ethidium bromide (1.0 μg/ml) and DNA bands were visualised and photographed under UV illumination. EcoRI-digested bacteriophage SPP1 DNA (Ratcliff *et al* 1979) and BglII-digested λ DNA were used as molecular size standards.

2.4.7 Estimation of relative molarities

When required, the relative amount of DNA in each band was estimated by scanning negatvies of gels with an LKB2202 Ultrascan Laser microdensitometer. Individual peaks (representing DNA bands) were excised, weighed and relative molarities were calculated according to the relationship:

$$\text{relative molarity} = \frac{(\text{weight of peak/total wt. of all peaks}) \text{ X estimated mol. wt. of DNA}}{\text{mol. wt. of fragment}}$$

2.4.8 Purification of the restriction fragments from plasmid DNA

Plasmid DNA digested with the appropriate restriction enzyme(s) was electrophoresed using a low gelling temperature (LGT) agarose gel containing 0.5 µg/ml of ethidium bromide in the buffer. Bands were visualised under UV and the segment(s) of the gel containing DNA of interest was cut out with a sterile scalpel blade. After this DNA was either extracted by the electroelution method, described by Maniatis *et al* (1982), or mixed with 1-2 ml of TE buffer and heated at 65°C for 30 minutes. This solution was then extracted twice with phenol at room temperature and once with ether at 4°C. DNA was ethanol precipitated and resuspended in TE.

2.4.9 DNA ligation

Digested and heat-inactivated DNA samples were usually ligated in a 20 µl volume. Sticky-end ligations were carried out at 4° C for 1-2 hours (approximately 50% ligation having occurred, as observed on a minigel) after which the mixture was diluted 10 to 20 fold with the addition of milliQ water and ligation buffer (10X: 0.66 M Tris-HCl, pH 7.5; 10 mM $MgCl_2$; 0.1 M dithiothreitol, 1 mM ATP) and the ligation continued for 12 to 16 hours. Blunt-end ligations were carried out in a total volume of 10 µl. For each ligation reaction the DNA concentration of vector and insert used were calculated using the method of Dugaiczyk *et al* (1975). T4 DNA ligase was used in all ligation reactions.

2.4.10 Dephosphorylation, end-filling, Bal31 exonuclease digestion and linker ligation

These techniques were performed essentially as described by Maniatis *et al* (1982). Linker DNA was phosphorylated by using PNK (polynucleotide kinase) according to the method of Maniatis *et al* (1982).

2.4.11 Cosmid cloning

Sau3A partial digests of chromosomal DNA were prepared by the method of Maniatis *et al* (1982). The cosmid used in this experiment was pHC79 (Honn and Collins, 1980). It was digested with BamHI and ligated with the partial Sau3A fragments in a final volume of 20 µl by mixing 2 µg of insert DNA with BamHI-digested and phosphatase-treated vector DNA in a molar ratio of 1:3. A 3 µl portion of this ligation mixture was packaged *in vitro* by the method of Sternberg *et al* (1977). DH1 was used as the host strain for infection and was plated out on ampicillin agar.

2.4.12 Bacterial transformation

Bacterial strains were made competent and transformed with plasmid DNA routinely by the method of Lederberg and Cohen (1974).

2.5 Hybridization techniques

2.5.1 Nick translation

The method of Rigby *et al* (1977) was employed for ^{32}P-labelling of DNA by nick translation of both plasmid DNA and DNA fragment extracted for gels.

2.5.2 Southern hybridization

Unidirectional Southern transfers and *in situ* DNA hybridizations were carried out using the method of Southern (1975) while bidirectional Southern transfers were done using the method of Smith and Summers (1980).

2.5.3 Colony hybridization

Colony hybridization was performed as described by Grunstein and Hogness (1975).

2.6 Preparation of heteroduplex DNA and electron microscopy

Heteroduplex formation and DNA spreading were performed using the method of Davis and Parkinson (1971) with minor modifications. Grids were prepared by the basic protein film technique as described by Kleinschmidt and Zahn (1959), and stained with uranyl acetate or rotary shadowed with platinum-palladium.

2.7 Bacterial agglutination

The agglutination was performed on glass slides as described by Leinonen (1985). A small amount of the bacterial culture was transferred into a drop of a suitable dilution of antiserum in 0.2% NaCl on a glass slide. After mixing well, the slide was tilted up and down for 2 minutes and examined against a dark background with an obliquely placed light source. Agglutination usually started in 10 to 20 seconds. As a control, 4% NaCl instead of antiserum was used to detect auto-agglutination. *Salmonella* somatic antisera of 0-2, 0-4 and 0-9 were obtained from IMVS, Adelaide, South Australia and Wellcome Diagnostic, England.

2.8 DNA sequence analysis

2.8.1 Preparation of oligonucleotide primers

0.2 μmoles of each oligonucleotide was made by using an "Applied Biosystems" 381A DNA synthesizer. DNA was diluted with 1X TE buffer to get a final concentration of 2.5 ng/ml. 1 μl DNA of this dilution was used as a primer for each reaction.

2.8.2 M13 techniques and DNA sequencing

Subcloning of DNA fragments into M13mp18 and M13mp19 (Norrander *et al* 1983) was performed, and the resultant recombinant phages were used for DNA sequencing according to the chain termination method of Sanger *et al* (1977).

2.8.3 Computer analysis

DNA sequences were compiled and analysed by Staden's programs (Staden 1980,1982; Staden and McLachlan 1982 and Staden 1984). Nucleotide and protein sequences were aligned using SEQA and SEQHP respectively (Kanehisa 1982). Secondary structure of protein sequences was predicted by the programs CHOUCOLOUR (MBIS) and PREDICT (Eliopoulos *et al* 1982); the MBIS suite of programs (Bucholtz and Reisner 1986) were used to screen the nucleic acid and protein databases for sequence homologies. All these programs were compiled by CSIRO, Division of Molecular Biology, North Ryde, New South Wales 2113, under a computer-based system termed MBIS [V 3.0, December 1986].

2.9 Protein methods

2.9.1 Minicell isolation and purification

Minicells were isolated and purified from parent cells using sucrose density gradient centrifugations and described by Clark-Curtiss and Curtiss III (1983).

2.9.2 L-[^{35}S] methionine labelling of proteins

Plasmid encoded proteins in minicells were labelled with L-[^{35}S] methionine using the method described by Clark-Curtiss and Curtiss III (1983). The labelled samples were frozen at -20° C overnight and then boiled for 5 minutes prior to SDS-PAGE gel electrophoresis.

2.9.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography

The labelled protein samples were analysed on 11% linear SDS-polyacrylamide gels as described by Lugtenberg *et al* (1975). Molecular weight markers were purchased from Pharmacia (Electrophoresis calibration kit, lot. no. 1040) which provided markers with molecular weights 94 kD, 67 kD, 43 kD, 30 kD, 20.1 kD and 14.4 kD. After electrophoresis gels were soaked in the fixer (40% methanol, 10% acetic acid) for 30 minutes, followed by staining overnight with Coomassie Brilliant Blue G250 (0.06%W/V) in 3.5% perchloric acid. The gels were soaked in destain (30% methanol, 7% acetic acid) with 3 to 4 changes every half hour. After sufficient destaining gels were transferred into a 25% ethanol solution for 30 minutes followed by drying under vacuum by using a Biolab gel drier. The gels were autoradiographed at room temperature for 2 to 5 days using Kodak X-ray film.

CHAPTER 3

Cloning and physical mapping of *rfb* regions from *S. typhi* and *S. paratyphi A*

3.1 Introduction

This project was running in parallel with those of two other Ph.D. students, H. Brahmbhatt and P. Wyk, who were analysing *rfb* genes of *S. typhimurium* LT2 in our laboratory. Some of the information from their work has been used for comparative studies with group D and group A *Salmonellae*. I have cited the reference(s) where information from their work has been used.

The aim of the work outlined in this chapter was to clone DNA from *S. typhi* Ty21a (group D) and *S. paratyphi A* IMVS1316 (group A) covering the whole *rfb* region, identify the areas of homology and nonhomology, and to look for the regions known to encode enzymes necessary for paratose and tyvelose biosynthesis. The following paragraphs describe the molecular cloning and restriction enzyme mapping of the *rfb* gene clusters of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. The cloning strategy used a cosmid vector and the Southern hybridization technique was used to detect DNA homology. The junction between the homologous and nonhomologous regions of both ends was localised by heteroduplex analysis.

3.2 Cosmid cloning

Cosmid banks were constructed from *S. typhi* Ty21a (M18) and *S. paratyphi A* IMVS1316 (M8) genomic DNAs, using pHC79 as a vector (Hohn and Collins 1980). A 2.25 kb KpnI fragment, which lies approximately in the middle of the *S. typhimurium rfb* operon (positions 9.95 to 12.20), was isolated from pPR300 (Brahmbhatt *et al* 1988). This fragment was used as a radioactive probe to screen cosmid libraries for the presence of homologous DNA. Eight clones in the *S. typhi* Ty21a library and two clones in the *S. paratyphi A* IMVS1316 library were identified with this probe (Fig 3.1). EcoRI digests of these clones were probed with pHC79 to show which fragments corresponded to *rfb* and which to vector DNA. DNA preparations from these ten cosmid clones were digested with EcoRI and run on agarose gels; they were found to have a number of common fragments, many of which also occur in *S. typhimurium*. These fragments were assumed to be homologous to those of *S. typhimurium*, and on this basis partial maps of the cosmid clones were deduced (Fig 3.2). Hybridization of the same 2.25 kb radioactive probe with blotted EcoRI digests of all cosmid clones and with chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 identified a single fragment in each case which, with the exception of pPR429, had the same mobility as the 11.0 kb EcoRI fragment of *S. typhimurium* DNA (positions 9.57 to 20.59) from which the probe was derived (Fig 3.3).

FIGURE 3.1

Identification of a cosmid clone using colony hybridization

Autoradiogram showing the identification of a cosmid clone during colony hybridization of the cosmid bank of *S. typhi* Ty21a in cosmid pHC79. A purified KpnI fragment (2.25 kb) of plasmid pPR300 (Brahmbhatt *et al* 1988) from *S. typhimurium* was used as a probe. The arrow indicates the clone that was identified by the probe used.

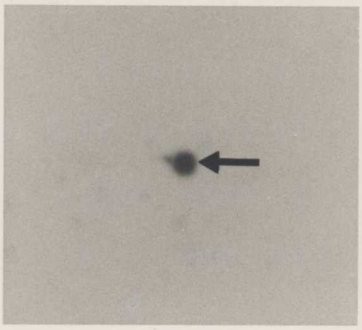


FIGURE 3.2

EcoRI restriction map of *rfb* DNA cloned in various cosmids from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316

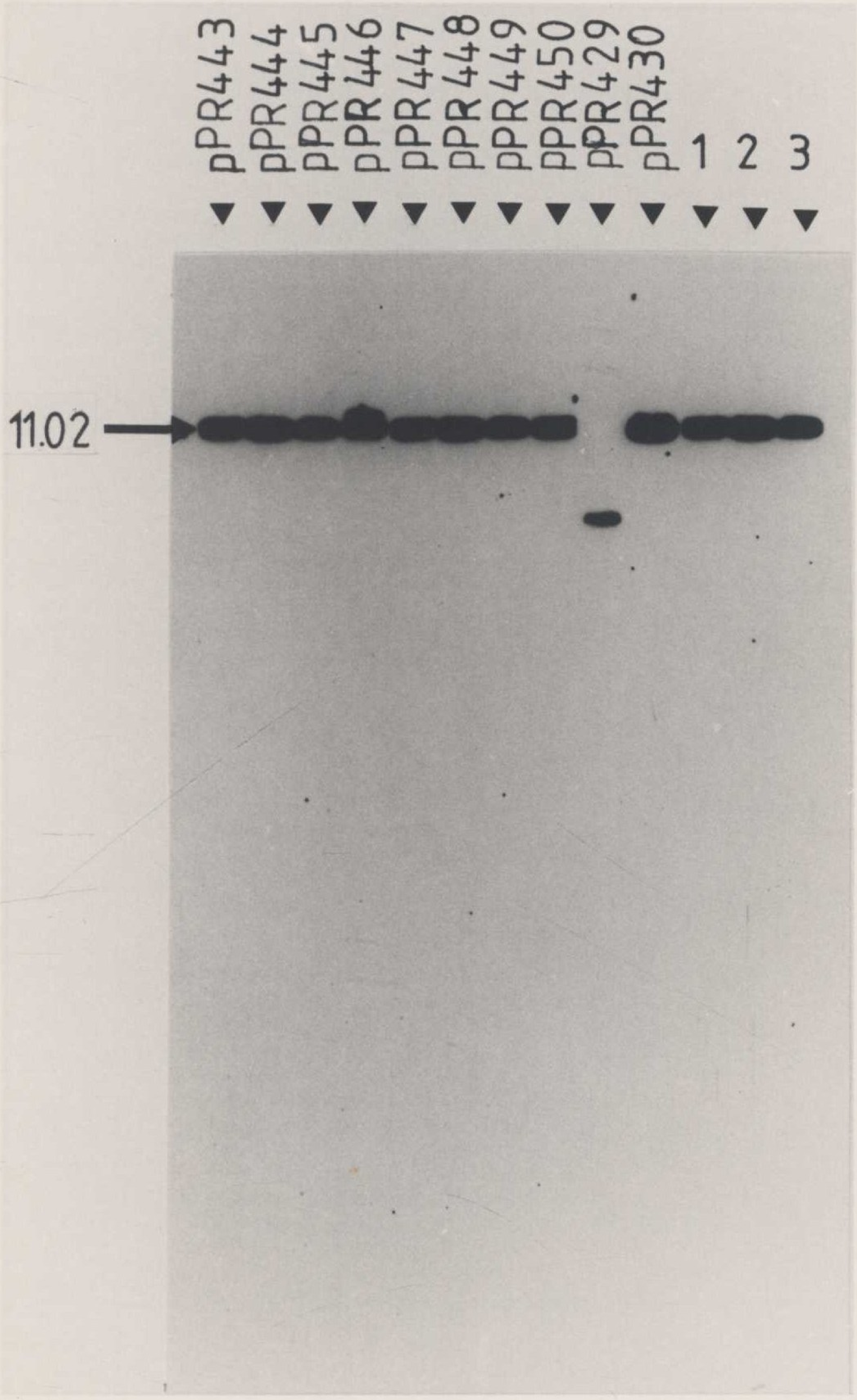
EcoRI restriction maps of cosmids carrying part of the *rfb* operons from *S. typhi* Ty21a (pPR443, pPR444, pPR445, pPR446, pPR447, pPR448, pPR449 and pPR450) and *S. paratyphi A* IMVS1316 (pPR430 and pPR429). The approximate extent of the *rfb* region of *S.typhimurium* LT2 is indicated by a heavy line. The other cosmids are aligned at the conserved EcoRI site at coordinate 9.57 of *S. typhimurium* LT2. The numbers 1, 2, 3 and 4 indicate the *rfb* DNA fragments that were subcloned to form plasmids pPR505, pPR434, pPR437 and pPR436, respectively.


0
9.57
20.59
S.typhimurium LT2
pPR443
pPR444
pPR445
pPR446
pPR447
pPR448
pPR449
pPR450
pPR430
pPR429
1
2
3
4

FIGURE 3.3

Confirmation of cosmid clones carrying *rfb* DNA

S. typhi Ty21a and *S. paratyphi A* IMVS1316 cosmid clones digested with EcoRI and probed with the 2.25 kb KpnI radioactive fragment of *S. typhimurium* LT2 *rfb* DNA (positions 9.95 to 12.20) from pPR300. Lanes 1, 2 and 3 contain *S. typhimurium* LT2, *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 chromosomal EcoRI digests, respectively.


pPR443
pPR444
pPR445
pPR446
pPR447
pPR448
pPR449
pPR450
pPR429
pPR430
1
2
3
11.02

3.3 Subcloning and restriction enzyme mapping

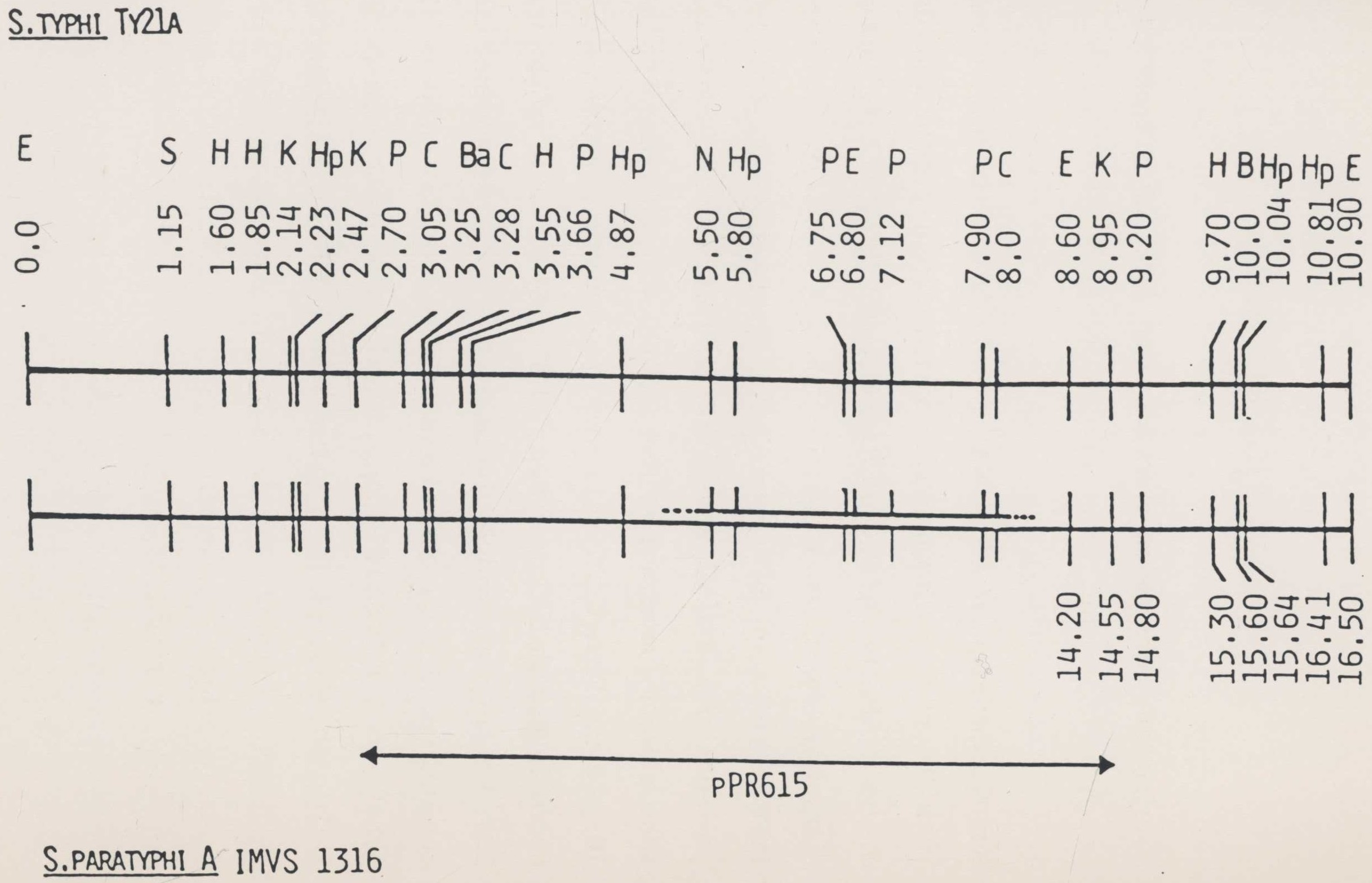
All EcoRI fragments from *S. typhi* and *S. paratyphi A* cosmid clones were subcloned into plasmid vectors pBR325 (Prentki *et al* 1981) and pUC9 (Vieira and Messing 1982) (Fig 3.2). The 11.0 kb fragments corresponding to the *S. typhimurium* fragment from positions 9.57 to 20.59, which covers most of the *rfb* gene cluster, were examined in detail. Restriction enzyme digests with HindIII, KpnI, ClaI, HpaI, PstI, XbaI and BglII gave identical patterns to those given with the corresponding *S. typhimurium* fragment. A SacI site which was absent in *S. typhimurium* was found in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 and mapped at a position corresponding to coordinate 12.6 of *S. typhimurium*, and a SacI site at coordinate 9.97 in *S. typhimurium* LT2 was absent in both *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. We conclude that DNA between the EcoRI sites at positions 9.57 and 20.59 in *S. typhimurium* LT2 is essentially homologous with the corresponding regions in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 DNA. This conclusion was supported by hybridizing radioactive *S. typhimurium* DNA from the homologous region with blotted DNA fragments from double enzyme digests of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 DNA within the same region.

Restriction site mapping was carried out on the 6.8, 2.3 and 1.8 kb EcoRI fragments from *S. typhi* Ty21a and 6.8, 2.8, 2.3 and 1.8 kb EcoRI fragments from *S. paratyphi A* IMVS1316. The order of these fragments was determined using them separately as probes to hybridize with double enzyme digests of chromosomal DNA. A restriction map of this region was then constructed from positions 0

FIGURE 3.4

Restriction map of a part of *rfb* DNA of *S.typhi* Ty21a and *S. paratyphi A* IMVS1316

Restriction map of *rfb* DNA of *S. typhi* Ty21a from positions 0 to 10.9 and *S. paratyphi A* IMVS1316 from positions 0 to 16.5. *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 have substantial similarity in their restriction maps. The *S. paratyphi A* IMVS1316 map is drawn such that sites present between positions 0 and 4.87 and positions 14.2 and 16.5 correspond to sites present between positions 0 and 4.87 and positions 8.6 and 10.9 on the *S. typhi* Ty21a map, respectively. The sites found to be triplicated in *S. paratyphi A* IMVS1316 (see text) are within a 2.8 kb region which includes the 2.5 kb segment from coordinates 5.5 to 8.0, indicated by double lines in the *S. paratyphi A* IMVS1316 map. Restriction enzyme sites: E, EcoRI; H, HindIII; K, KpnI; Hp, HpaI; P, PstI; C, ClaI; Ba, BamHI; N, NruI; B, BglII. Plasmid pPR615 is pUC19 carrying a 12.1 kb KpnI fragment which extends from positions 2.47 to 14.55 on the *S. paratyphi A* IMVS1316 map.


S.TYPHI TY21A
E S H H K Hp K P C Ba C H P Hp N Hp P E P P C E K P H B Hp Hp E
0.0
1.15
1.60
1.85
2.14
2.23
2.47
2.70
3.05
3.25
3.28
3.55
3.66
4.87
5.50
5.80
6.75
6.80
7.12
7.90
8.0
8.60
8.95
9.20
9.70
10.0
10.04
10.81
10.90
14.20
14.55
14.80
15.30
15.60
15.64
16.41
16.50
pPR615
S.PARATYPHI A IMVS 1316

to 10.9 of *S. typhi* Ty21a and positions 0 to 16.5 of *S. paratyphi A* IMVS1316 (Fig 3.4).

3.4 A 2.8 kb triplication

All restriction sites on the 6.8, 2.3 and 1.8 kb fragments from *S. paratyphi A* IMVS1316 were found to map at the same positions on the corresponding fragments of *S. typhi* Ty21a. The 2.8 kb EcoRI fragment, which stains with an intensity approximately twice that expected by comparison with adjacent bands, appears from the restriction map to duplicate the adjacent ends of the 6.8 and 1.8 kb *S. typhi* Ty21a EcoRI fragments (Fig 3.4). This suggests that a 2.8 kb region is triplicated (Fig 3.5). NruI, HpaI, EcoRI and ClaI digests all gave a 2.8 kb fragment, which was shown by densitometry of gel photographs to be present at twice the molarity of other fragments (Fig 3.5). This confirmed the 2.8 kb triplication and showed that one end lies between positions 4.87 and 5.50. Further confirmation was provided by the existence of the expected 12.1 kb KpnI fragment in the chromosome (pPR615) which extends from position 2.47 to position 14.55 (Fig 3.4). This was shown by probing KpnI-digested chromosomal DNA with ^{32}P-labelled pPR437 (positions 12.4 to 14.2 on *S. paratyphi A* IMVS1316 map). The probe hybridized with the 12.1 kb fragment (Fig 3.6), indicating that it was not a cloning artifact.

3.5 Identification of the region of nonhomology

The region of nonhomology between the three *Salmonella* species was identified using a purified 9.57 kb EcoRI *S. typhimurium*

FIGURE 3.5

Densitometric scan to show the presence of triplication of a 2.8 kb region

Plasmid pPR615 restriction fragments that were obtained after digestion with (A) ClaI, (B) KpnI-EcoRI, (C) NruI-KpnI and (D) HpaI and analysed by agarose gel electrophoresis, followed by staining with ethidium bromide to visualize the DNA fragments. Densitometer tracings are aligned with a photograph of the digest. The calculations of the relative molarity of each band are presented in Table 3.1.

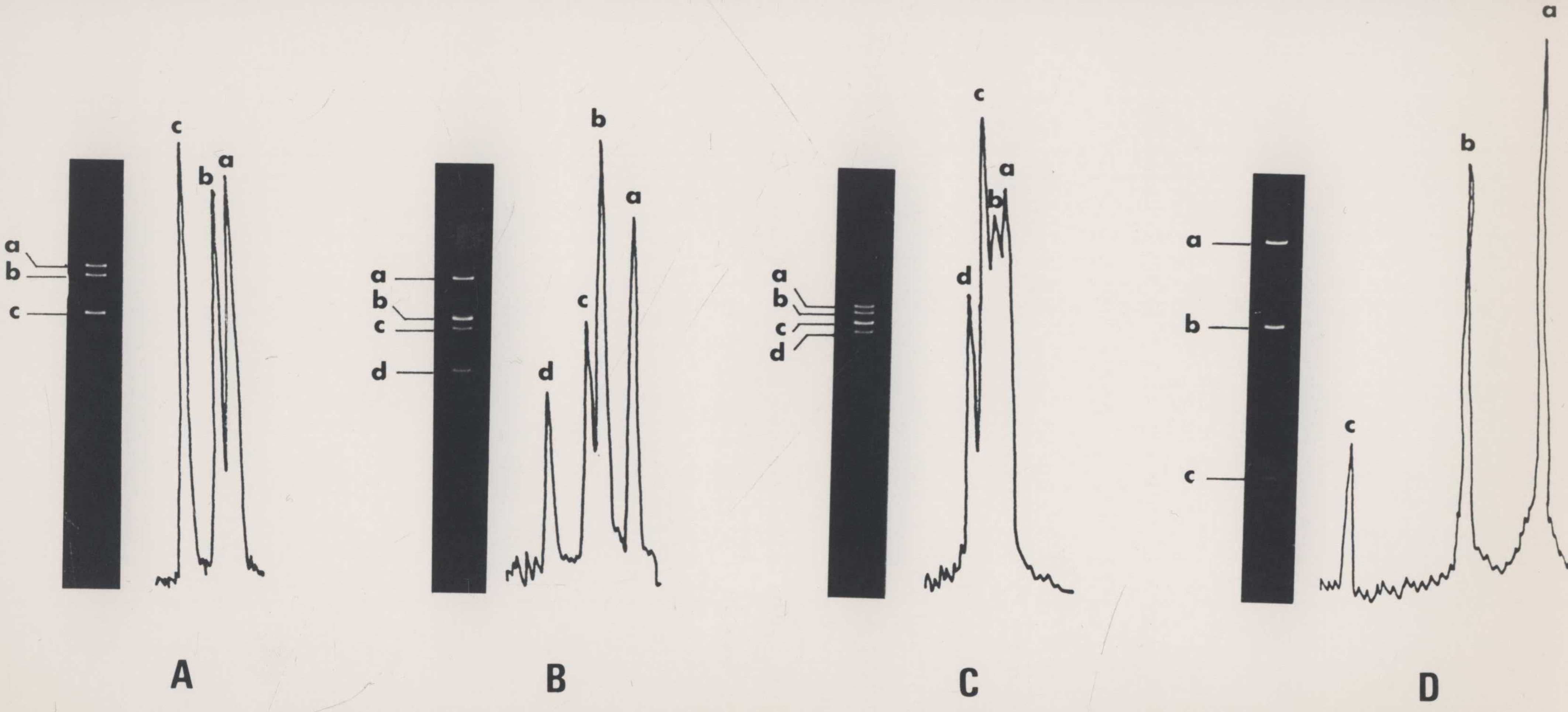

a
b
c
c
b
a
A
a
b
c
d
b
a
c
d
B
a
b
c
d
c
a
b
d
C
a
b
c
a
b
c
D

TABLE 3.1

Molecular weights and relative molarities

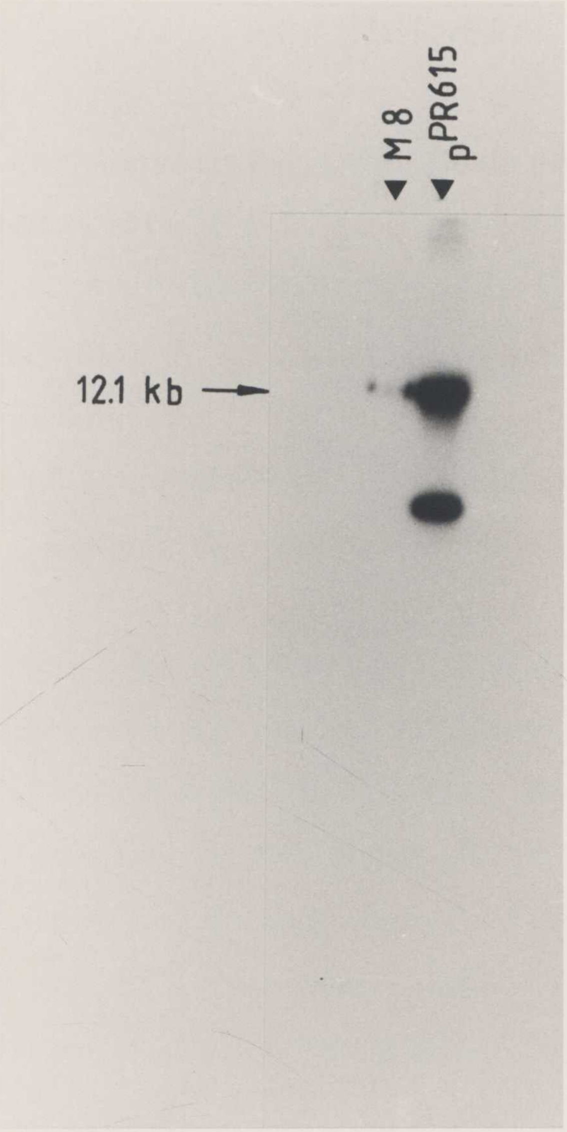
The calculated relative molarities for DNA fragments of each digest are listed here. Relative molarities are calculated by the formula given in section 2.4.7. Densitometric scan is shown in Fig 3.5.

Digest	Fragment	Mol.Wt.(kb)	Calculated Relative Molarity
ClaI	a	5.0	1.0
	b	4.3	1.0
	c	2.8	2.0
KpnI-EcoRI	a	4.5	0.96
	b	2.8	2.0
	c	2.6	1.0
	d	1.9	0.9
NruI-KpnI	a	3.45	1.0
	b	3.1	0.96
	c	2.8	2.0
	d	2.6	0.95
HpaI	a	8.0	1.0
	b	2.8	2.0
	c	0.96	0.85

FIGURE 3.6

Existence of the 12.1 kb KpnI fragment in the chromosome of *S. paratyphi A* IMVS1316

Chromosomal DNA from *S. paratyphi A* (M8) and plasmid pPR615 DNA were digested with KpnI and probed with ^{32}P-labelled pPR437 (positions 12.4 to 14.2 on *S. paratyphi A* IMVS1316 map) in Southern hybridization analysis. The arrow indicates position of 12.1 kb KpnI fragment.


M 8
pPR615
12.1 kb

DNA fragment from pPR301 (Brahmbhatt *et al* 1988) as a radioactive probe against EcoRI digests of chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. The probe, containing *S. typhimurium* DNA between map positions 0 and 9.57, hybridized strongly with the 6.8 kb and weakly with the 2.3 kb EcoR I fragments from these strains. In addition, the probe hybridized weakly with the 2.8 kb EcoRI fragment unique to the *S. paratyphi A* IMVS1316 digests (Fig 3.7). The common 1.8 kb fragment did not exhibit any detectable homology with the 9.57 kb *S. typhimurium* DNA probe.

3.6 Identification of the region of homology beyond *rfb* toward the *his* operon

A plasmid, pPR294 (Brahmbhatt *et al* 1986), was used as a molecular probe to determine whether the homology extended beyond *rfb* toward the *his* operon. This plasmid, encoding 3.35 kb directly to the left of the EcoRI site at position 0 (see Fig 3.14), hybridized with several restriction enzyme digests of chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 (Fig 3.8). The restriction enzyme sites toward the *his* operon for EcoRI, HpaI, BglII and PstI, mapped previously in *S. typhimurium* (Carlomagno *et al* 1983; Brahmbhatt *et al* 1986), were found to be conserved in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316.

3.7 Heteroduplex and electron microscopy

Southern hybridization and restriction enzyme analyses demonstrated that most of the sites from positions 0 to 5.5 and to

FIGURE 3.7

Identification of the region of nonhomology between *S. typhi* Ty21a, *S. paratyphi A* IMVS1316 and *S. typhimurium rfb* DNA

Southern hybridization analysis of EcoRI digests of *S. typhimurium* LT2, *S. typhi* Ty21a, and *S. paratyphi A* IMVS1316 chromosomal DNA and of various plasmids by using the EcoRI fragment of *S. typhimurium* LT2 *rfb* DNA from positions 0 to 9.57 as a radioactive probe. Lanes: 1, *S. typhimurium* LT2 chromosomal DNA; 2, *S. typhi* Ty21a chromosomal DNA; 3, *S. paratyphi A* IMVS1316 chromosomal DNA; 4, pPR505; 5, pPR434; 6, pPR436; 7, pPR437. The EcoRI subclones in lanes 4 to 7 carry the 6.8, 2.8, 2.3, and 1.8 kb *rfb* DNA fragments, respectively, from *S. paratyphi A* IMVS1316. The two bands of least intensity visible in lane 4 correspond to partial digestion products of pPR505.

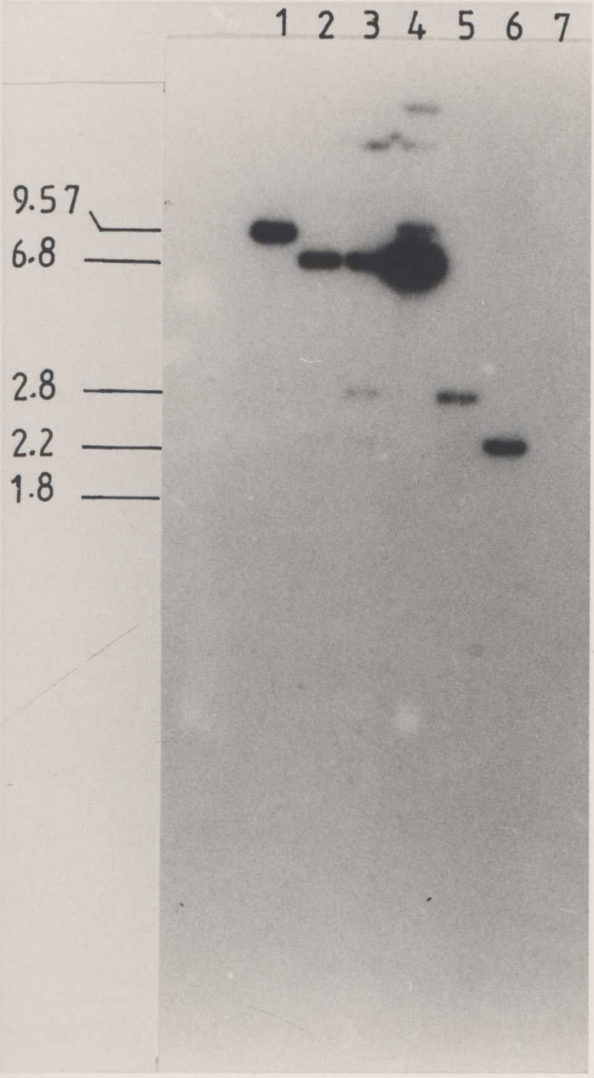
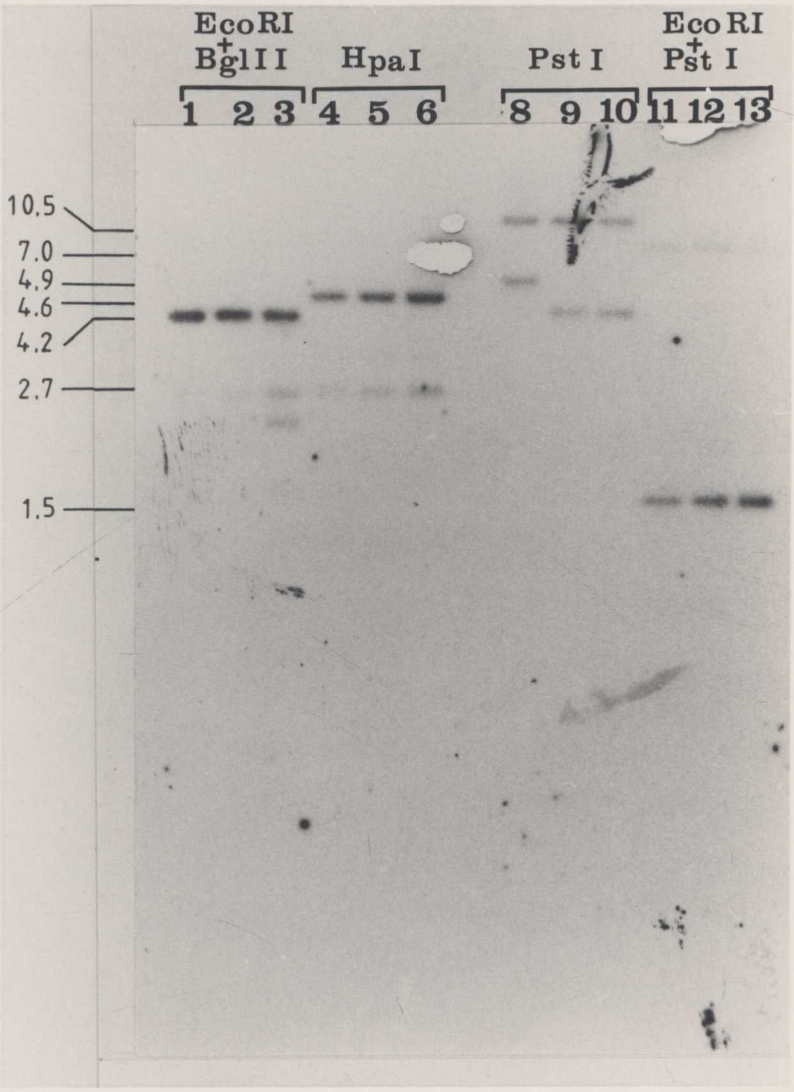

1
2
3
4
5
6
7
9.57
6.8
2.8
2.2
1.8

FIGURE 3.8

Identification of the region of homology beyond *rfb* toward the *his* operon

Southern hybridization analysis of restriction enzyme digests of *S. typhimurium, S. typhi* Ty21a and *S. paratyphi A* IMVS1316 chromosomal DNA probed with a 3.35 kb fragment from plasmid pPR294 (Brahmbhatt *et al* 1986) of *S. typhimurium* . Lanes: 1, 4, 8, and 11 = *S. typhimurium* ; 2,5,9, and 12 = *S. typhi* Ty21a; 3,6,10, and 13 = *S. paratyphi A* IMVS1316. Probe identified the similar size fragments in BglII-EcoRI, HpaI and PstI-EcoR I digests of chromosomal DNA from all three strains. Note that some bands resulted from partial DNA digests.


EcoRI
+
BglII
HpaI
Pst I
EcoRI
+
Pst I
1
2
3
4
5
6
8
9
10
11
12
13
10.5
7.0
4.9
4.6
4.2
2.7
1.5

the right of position 9.57 on the *S. typhimurium rfb* map were conserved. The EcoRI site of *S. typhimurium* at position 9.57 kb corresponds with the EcoRI sites at position 10.9 of *S. typhi* Ty21a and position 16.5 of *S. paratyphi A* IMVS1316. This suggests that the region of non-homology lies between positions 5.5 and 9.57. The junction between the homologous and nonhomologous regions of both ends was localised by heteroduplex analysis.

3.7.1 Localising the junction between homologous and nonhomologous regions in the vicinity of position 5.5

Plasmid pPR506 consists of pcos2EMBL (Poustka *et al* 1984) carrying the *S. paratyphi A* IMVS1316 6.8 kb EcoRI DNA fragment from positions 0 to 6.8. The plasmid construct pPR508 has a 9.57 kb EcoRI fragment (positions 0 to 9.57) from *S. typhimurium* cloned in pGB2 (Churchward *et al* 1984) (Fig 3.9). The vectors, pGB2 and pcos2EMBL, were chosen because they exhibit no homology with each other as predicted from their origins and confirmed by Southern blotting (data not shown). Plasmid pPR506 was digested with EcoRI while SalI was used to linearize pPR508. The digestion products were then allowed to hybridize in a heteroduplex reaction. Denaturation and renaturation of these plasmid DNA fragments gave rise to double-stranded linear molecules which branched into two single strands at one end (Fig 3.10). The double-stranded segments of these molecules were measured and the length of homologous DNA found to be about 5.8 ± 0.06 kb (Table 3.2A), indicating that the region of homology extends from position 0 to 5.8 between *S. paratyphi A* IMVS1316 and *S. typhimurium* LT2. Homoduplex

FIGURE 3.9

Plasmid constructs used for heteroduplex analysis

EcoRI restriction map of *S. typhimurium* LT2 and *S. paratyphi A* IMVS1316. Vertical bars represent the EcoRI restriction sites. Orientation of inserts with respect to the *lac* promoter is indicated by the arrowheads. Vectors used are shown in parentheses.

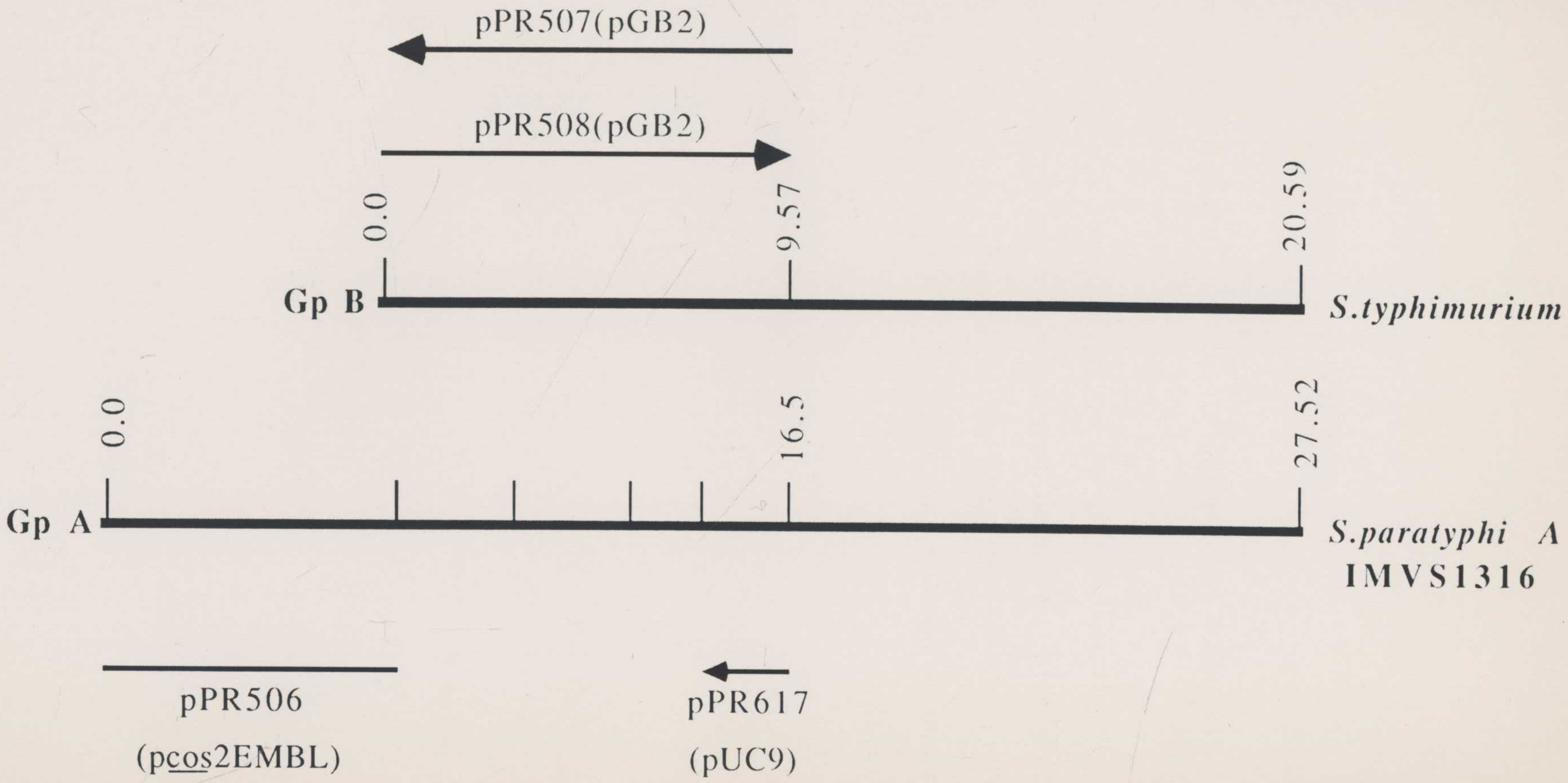
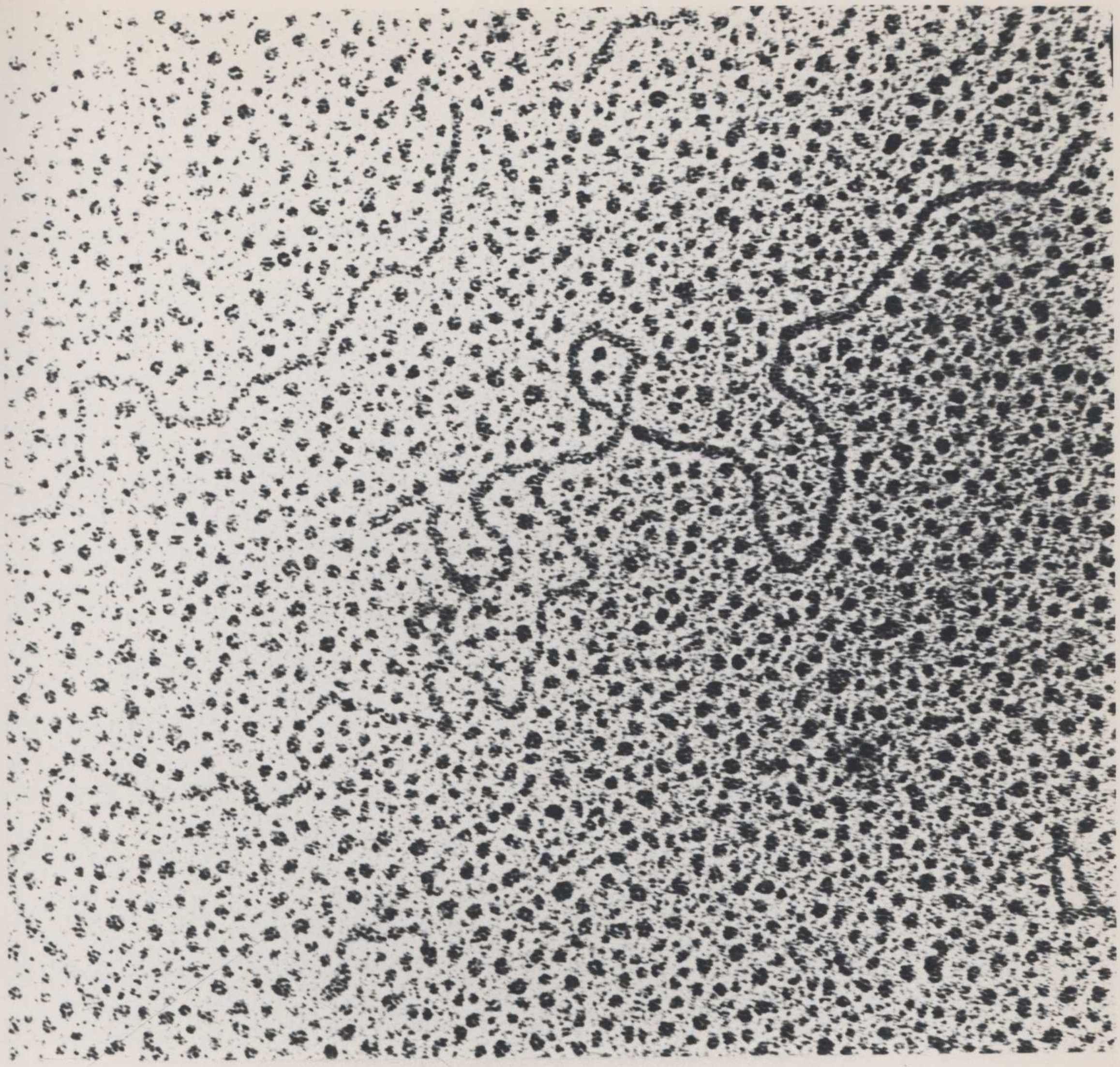
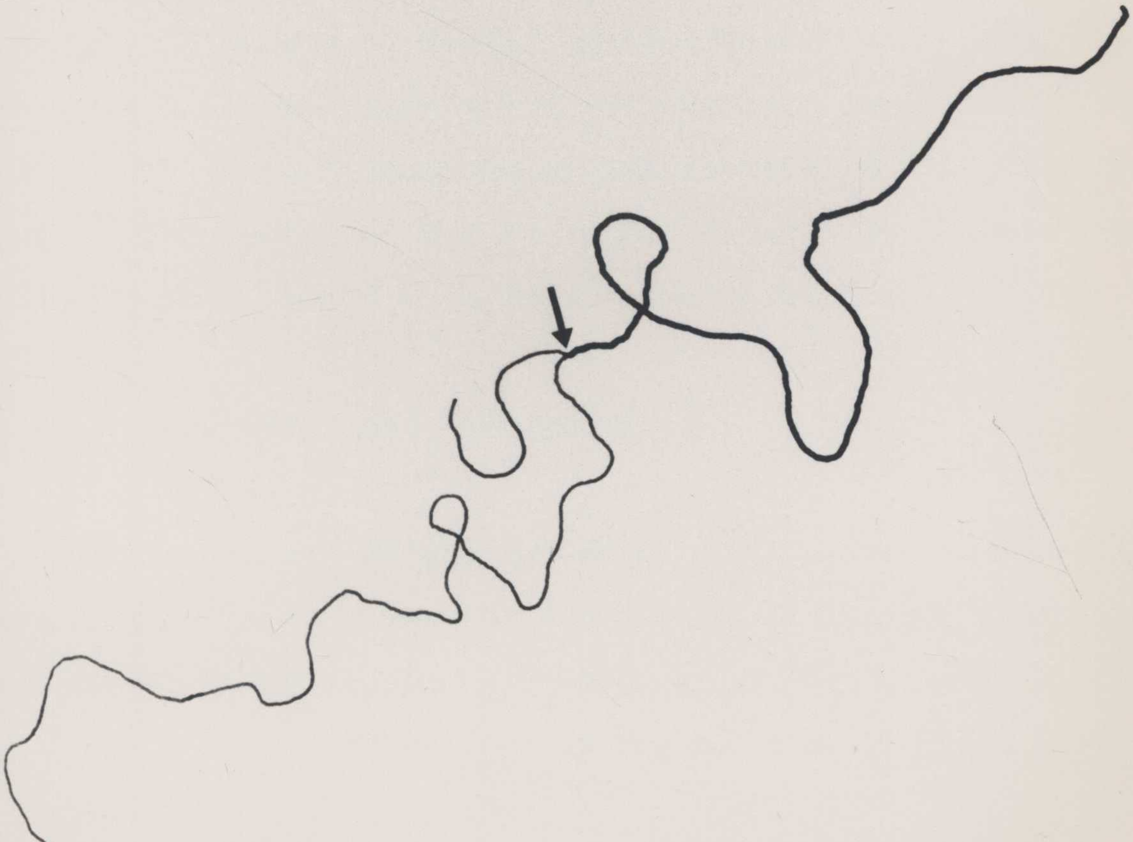

pPR507(pGB2)
pPR508(pGB2)
0.0
9.57
20.59
Gp B
S.typhimurium
0.0
16.5
27.52
Gp A
S.paratyphi A
IMVS1316
pPR506
(pcos2EMBL)
pPR617
(pUC9)

FIGURE 3.10

Electron micrograph of a heteroduplex molecule

Heteroduplex formed between a 6.8 kb EcoRI fragment from pPR505 (*S. paratyphi A* IMVS1316) and plasmid pPR508 (*S. typhimurium* LT2) linearized with SalI. The arrow indicates the end of the region of homology. Plasmid pcos2EMBL (6.1 kb) was used as the double-stranded DNA standard.





molecules were obtained after denaturation and renaturation of linearized pPR508 (Fig 3.11) as well as EcoRI digested pPR506 (Fig 3.12) in separate reactions. These were also used as standards to measure the length of double-stranded segments.

3.7.2 Localising the junction between homologous and nonhomologous regions in the vicinity of position 9.57 (*S. typhimurium* unit)

Plasmids pPR617 (containing the 2.3 kb EcoRI fragment of *S. paratyphi A* IMVS1316 cloned in pUC9) (Vieira and Messing 1982) and pPR507 (pGB2 carrying the 9.57 kb fragment of *S. typhimurium* (Fig 3.9) in the opposite orientation to that in pPR508) were used in a similar heteroduplex reaction to localise the other end of the nonhomologous region. pUC9 was also shown to have no homology with pGB2 (data not shown). pPR617 and pPR507 were linearized, with BamHI and SalI respectively, before being mixed in a heteroduplex reaction. Similar double-stranded linear molecules, branching into two single strands at one end, were formed (Fig 3.13). The double-stranded region was measured as 1.4 ± 0.05 kb (Table 3.2B), indicating that the region of homology extends from position 9.57 to about 8.17 on the *S. typhimurium* map.

3.8 Summary and discussion

The restriction analyses presented in this chapter show that most of the sites from positions 0 to 5.5 and to the right of position 9.57 on the *S. typhimurium rfb* map are conserved, with the 9.57 kb EcoRI site of *S. typhimurium* corresponding to the EcoRI sites at

FIGURE 3.11

Electron micrograph of a homoduplex molecule

Electron micrograph of a homoduplex, obtained after denaturation and renaturation of linearized pPR508.

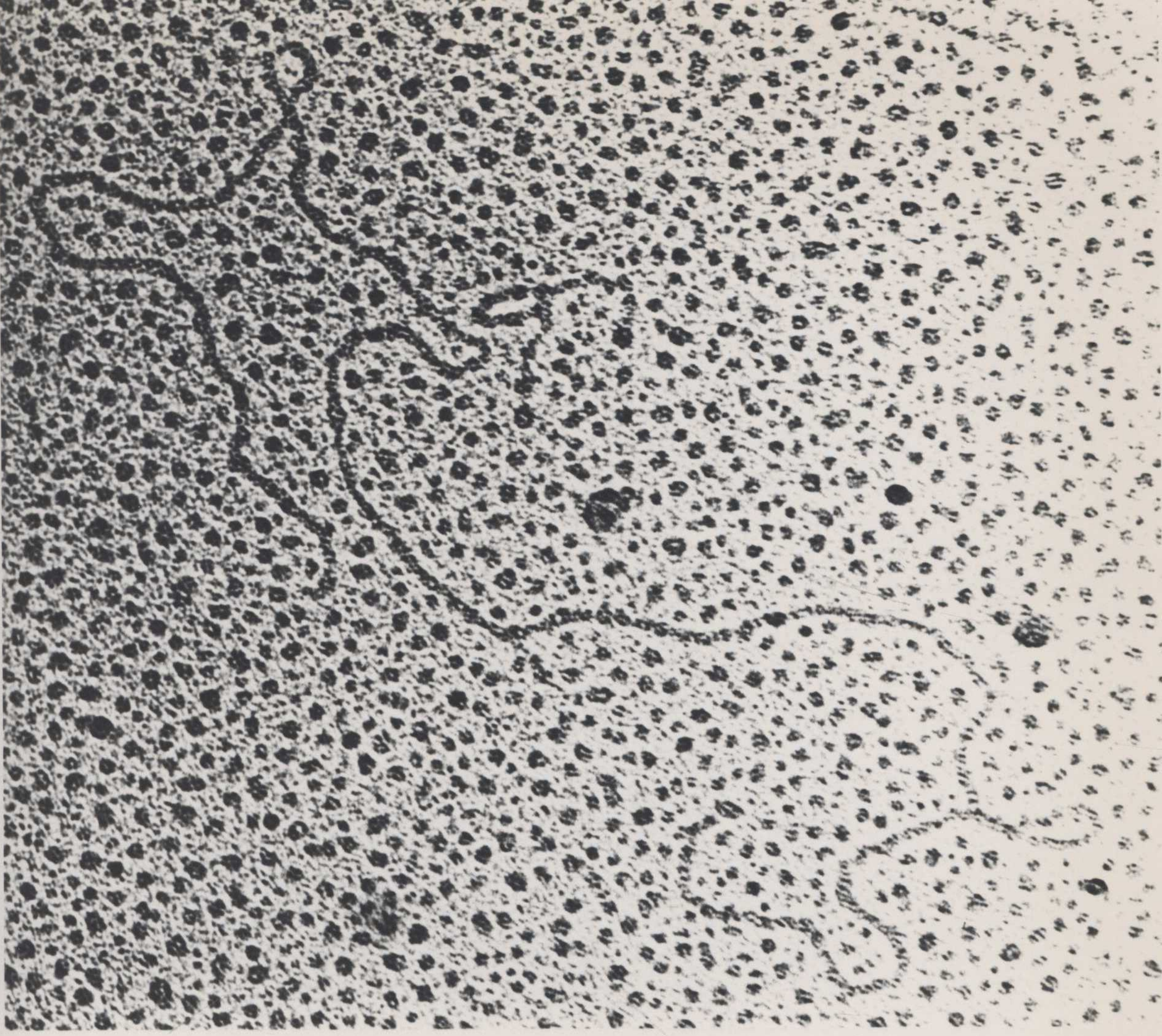
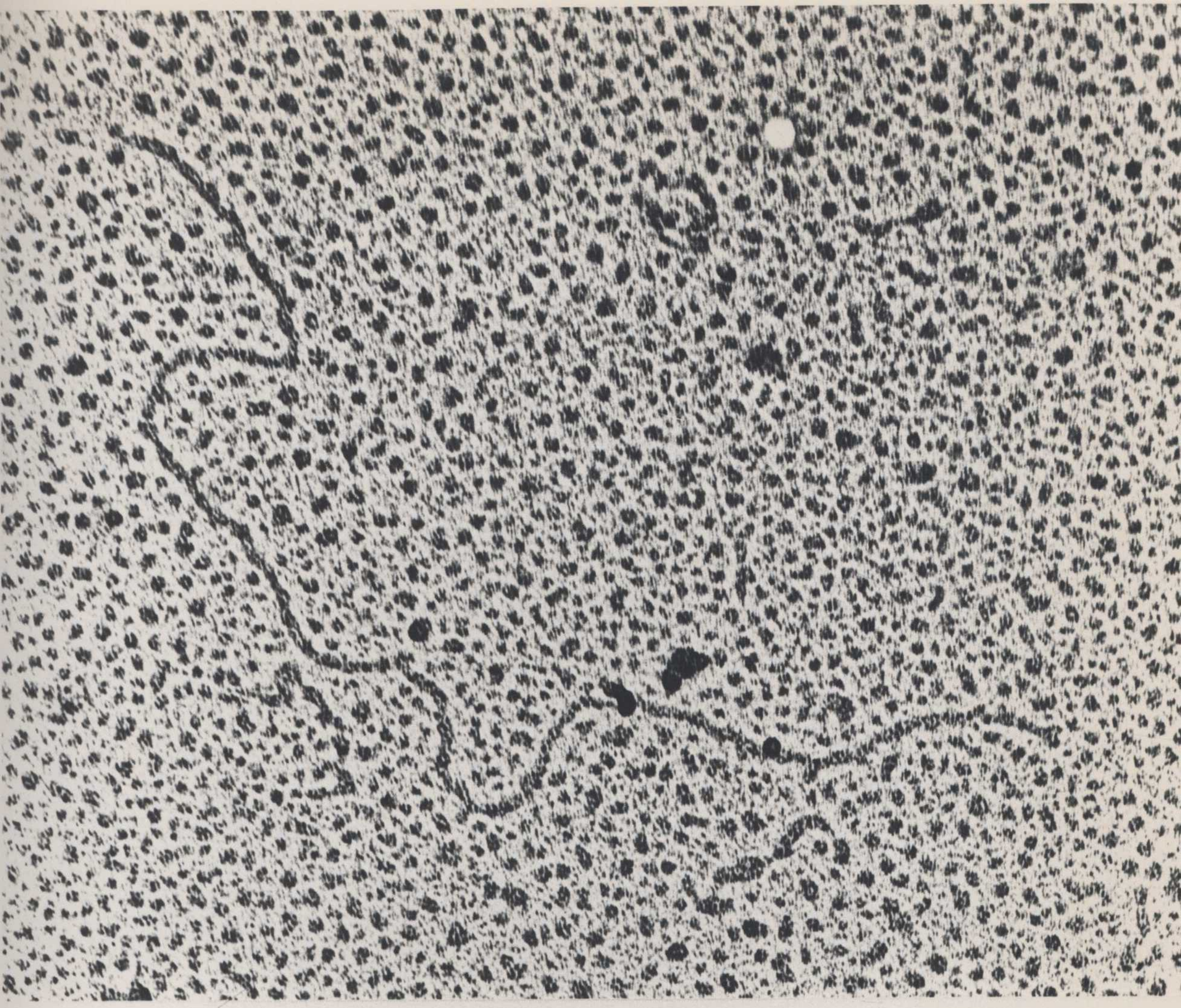


FIGURE 3.12

Electron micrograph of a homoduplex molecule

Electron micrograph showing a homoduplex molecule of 6.8 kb fragment formed after denaturation and renaturation of EcoR I digested pPR506.



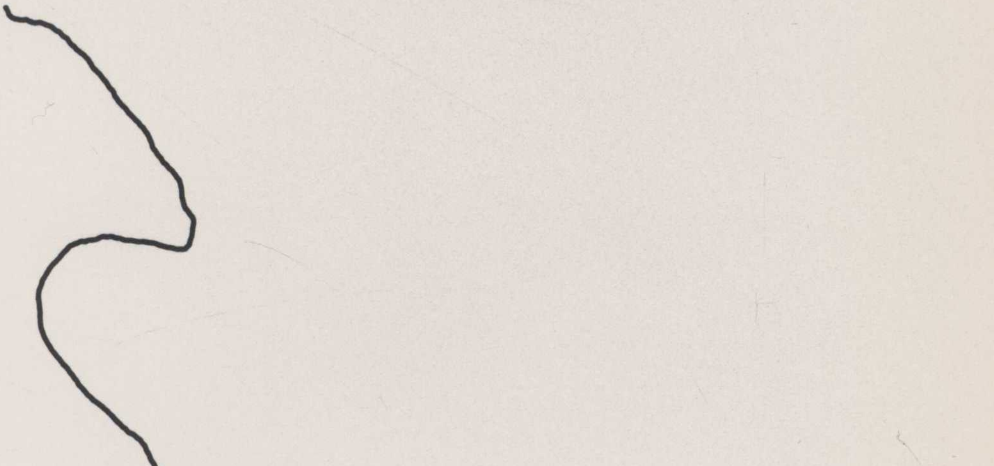


FIGURE 3.13

Electron micrograph of a heteroduplex molecule

Heteroduplex formed between plasmids pPR617 and pPR507 linearized with *Bam*HI and *Sal*I, respectively. The arrow indicates the end of the region of homology.

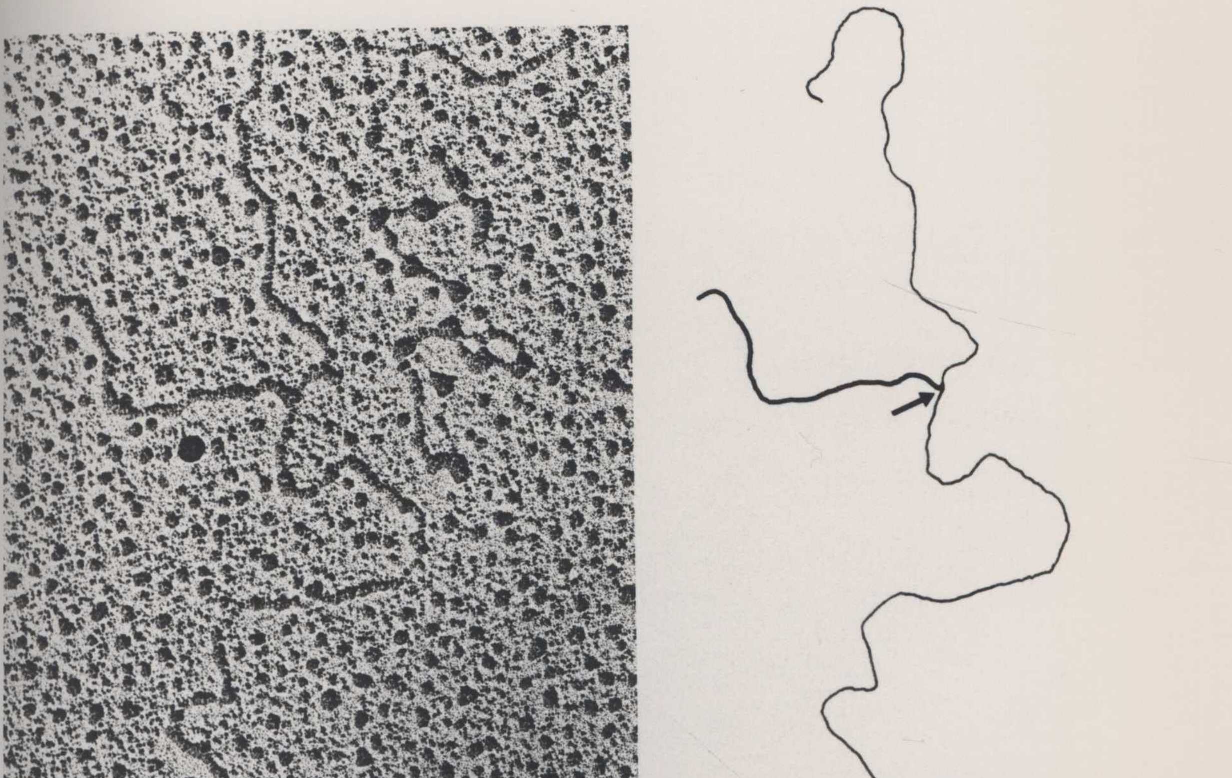


TABLE 3.2

Measurements of heteroduplexes formed

Measurements obtained from the analysis of heteroduplexes formed **A**) measurements of the double-stranded segment of heteroduplex molecules formed to localise the junction between homologous and nonhomologous regions in the vicinity of position 5.5. **B**) measurements of double-stranded segment of heteroduplex molecules formed to localise the junction at the other end of the region of nonhomology to the left of position 9.57 (*S. typhimurium*).

A

No. of molecules	Length of double-stranded segment measured (cms)	Calculated size of homologous DNA (kb)*
1	22.0	5.85
2	22.0	5.85
3	21.5	5.71
4	22.0	5.85
5	21.5	5.71
6	22.0	5.85
7	22.0	5.85
8	23.0	6.10
9	20.0	5.32
10	22.0	5.85

* mean size of homologous DNA = 5.79 ± 0.06

B

No. of molecules	Length of double-stranded segment measured (cms)	Calculated size of homologous DNA (kb)*
1	5.5	1.40
2	6.0	1.50
3	6.0	1.50
4	6.0	1.50
5	6.0	1.50
6	4.0	1.04
7	6.0	1.50
8	5.5	1.40

* mean size of homologous DNA = 1.41 ± 0.05

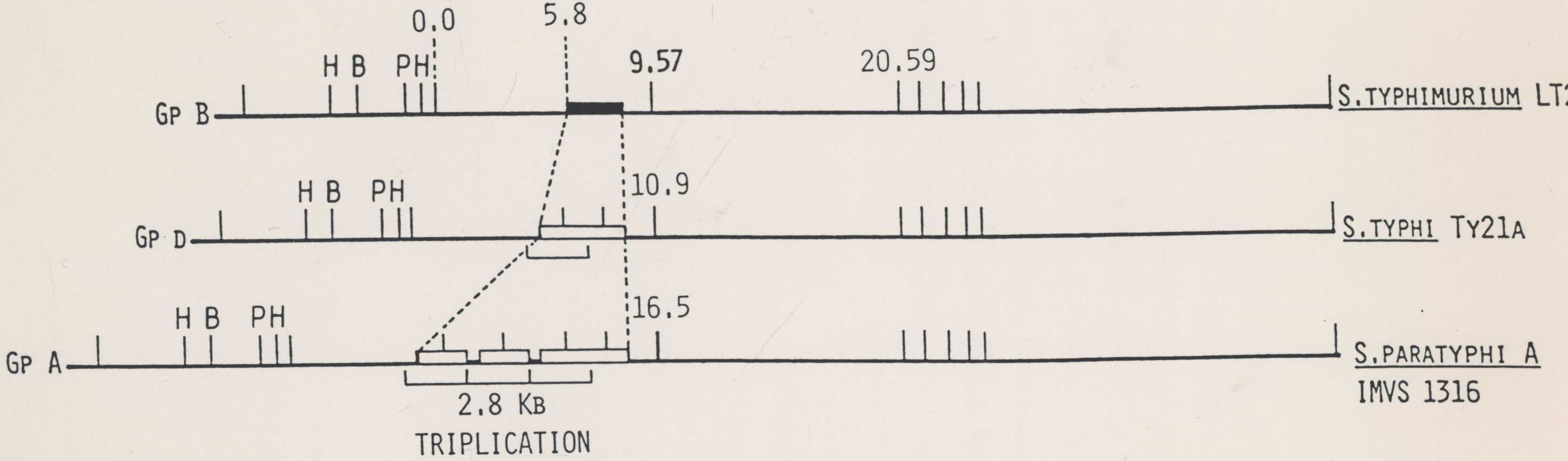
position 10.9 of *S. typhi* Ty21a and positon 16.5 of *S. paratyphi A* IMVS1316 (Fig 3.4). All restriction sites examined, with the exception of one SacI site, in the region from 9.57 (the first EcoRI site known after position 8.17) to 20.59 of *S. typhimurium*, were conserved in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 DNA. The DNA segment in *S. paratyphi A* IMVS1316 differs from that of *S. typhi* Ty21a only in having a 2.8 kb fragment triplicated, giving rise to 5.6 kb of additional DNA. This triplication was established by the presence of a 2.8 kb fragment at twice the molarity of other fragments in digests with four different enzymes. One end of the triplicated region lies between positions 4.87 and 5.5. Thus, the 2.8 kb repeated region includes a short piece of DNA homologous with *S. typhimurium* LT2, accounting for the homology of the 2.8 kb EcoRI fragment with pPR301.

The *S. typhimurium* LT2 regions from positions 0 to 5.8 and from 8.17 to 9.57 were shown to be homologous with *S. paratyphi A* IMVS1316 by heteroduplex analysis. However, despite the indicated homology in the region from position 8.17 to 9.57, little restriction site homology was detected. No attempts were made to analyse the 2.3 and 6.8 kb fragments from *S. typhi* Ty21a by heteroduplex technique. We assume that the same level of homology would be present in *S. typhi* Ty21a since both have been shown to have complete restriction site homology for those fragments used in heteroduplex reactions. The remaining region right to position 20.59, has not been examined in the same detail but presumably the same high level of homology exists throughout this conserved region (Fig 3.14).

FIGURE 3.14

Comparative map of *rfb* operon from *Salmonella* groups A, B and D

EcoRI sites are indicated by vertical bars. H, B and P are HpaI, BglII, and PstI recognition sites, respectively. Regions shown by restriction site homology and heteroduplex analysis to be homologous or unique are indicated as follows: —— , common to all; ▨, *S. typhimurium* LT2 specific; ▭, common to *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. Note that digestion at any site present only once in a triplicated interval produces a fragment of the same length as this interval and that the fragment is present at twice the molarity of other fragments generated by this enzyme.


0.0
5.8
H B PH
9.57
20.59
GP B
S.TYPHIMURIUM LT2
H B PH
10.9
GP D
S.TYPHI TY21A
H B PH
16.5
GP A
S.PARATYPHI A
IMVS 1316
2.8 KB
TRIPLICATION

Abequose is replaced by paratose and tyvelose in O-antigens of *S. paratyphi A* and *S. typhi* respectively. Other studies in our laboratory on *S. typhimurium* LT2, have shown that the genes for abequose synthesis (*rfbF, rfbG* and *rfbH*) (Brahmbhatt *et al* 1988) map in the general region of the structural difference between the three strains studied. Presumably genes encoded for biosynthesis of paratose and tyvelose may be present in this region. It seems that *S. paratyphi A* IMVS1316, which is expected to differ from *S. typhi* Ty21a in lacking CDP paratose-2-epimerase (Matsuhashi and Strominger 1965; Uchida *et al* 1974), nonetheless carries all of the DNA present in *S. typhi* Ty21a, but the gene for tyvelose synthesis may be inactive. Results presented in chapter 5 provide evidence for this.

CHAPTER 4

Localisation of DNA responsible for specificity in the biosynthesis of paratose and tyvelose, determination and analysis of its nucleotide sequence

4.1 Introduction

The 3, 6-dideoxyhexoses - paratose, abequose and tyvelose - are the immunodominant sugars of group A, B and D of *Salmonella.* The biochemical study of dideoxyhexoses revealed their close relationship in the biosynthetic pathway (Fig 1.4). A stereochemical reduction of an intermediate CDP-4-keto-3, 6-dideoxy-D-glucose at the last step leads to the formation of CDP-paratose or CDP-abequose. Matsuhashi and Strominger (1964, 1965, 1967) demonstrated that the interconversion of CDP-paratose and CDP-tyvelose, which are the donors of respective dideoxyhexoses in biosynthesis of O-polysaccharides, is catalysed by a single enzyme, CDP-paratose-2-epimerase.

Genetic studies in recombinants with exchanged *rfb* clusters have shown that recombinants containing the donor *rfb* as a rule produce O- specific units of the donor type (Nikaido *et al* 1966; Mäkelä 1966; Kochibe 1970; Kishi and Iseki 1973a, 1973b; Jousmies and Mäkelä 1974). These studies were based on conjugation systems that transferred large numbers of genes in any one cross, and so the effect of transfer of an individual gene could not be determined. We transformed smaller parts of *S. paratyphi A* and *S.*

typhi rfb DNA into a strain of group B and *S. typhi* DNA into a strain of group A, on the premise that, if they had enzyme function(s), they would substitute either paratose or tyvelose for the normally produced 3,6-dideoxyhexose. Most of the genes involved in abequose and paratose biosynthesis *rfbF*, *rfbG* and first two steps of '*rfbH* ' (genes for these two steps have been named *rfbH* and *rfbI* in *S. typhimurium*) are the same (Fig 1.3). The *rfbJ* gene, whose product (catalyses last step converting CDP-4-keto-3, 6-dideoxy-D-glucose to CDP-abequose in *S. typhimurium*) was expected to be identical with corresponding gene of *S. typhi* or *S. paratyphi A*. The *rfbJ* in *S. typhimurium* has been recently defined in our laboratory (P. Wyk, Ph.D. thesis, 1988). We identified and sequenced the gene(s) responsible for the enzyme(s) conferring specificity of the 3, 6-dideoxyhexose (paratose or tyvelose).

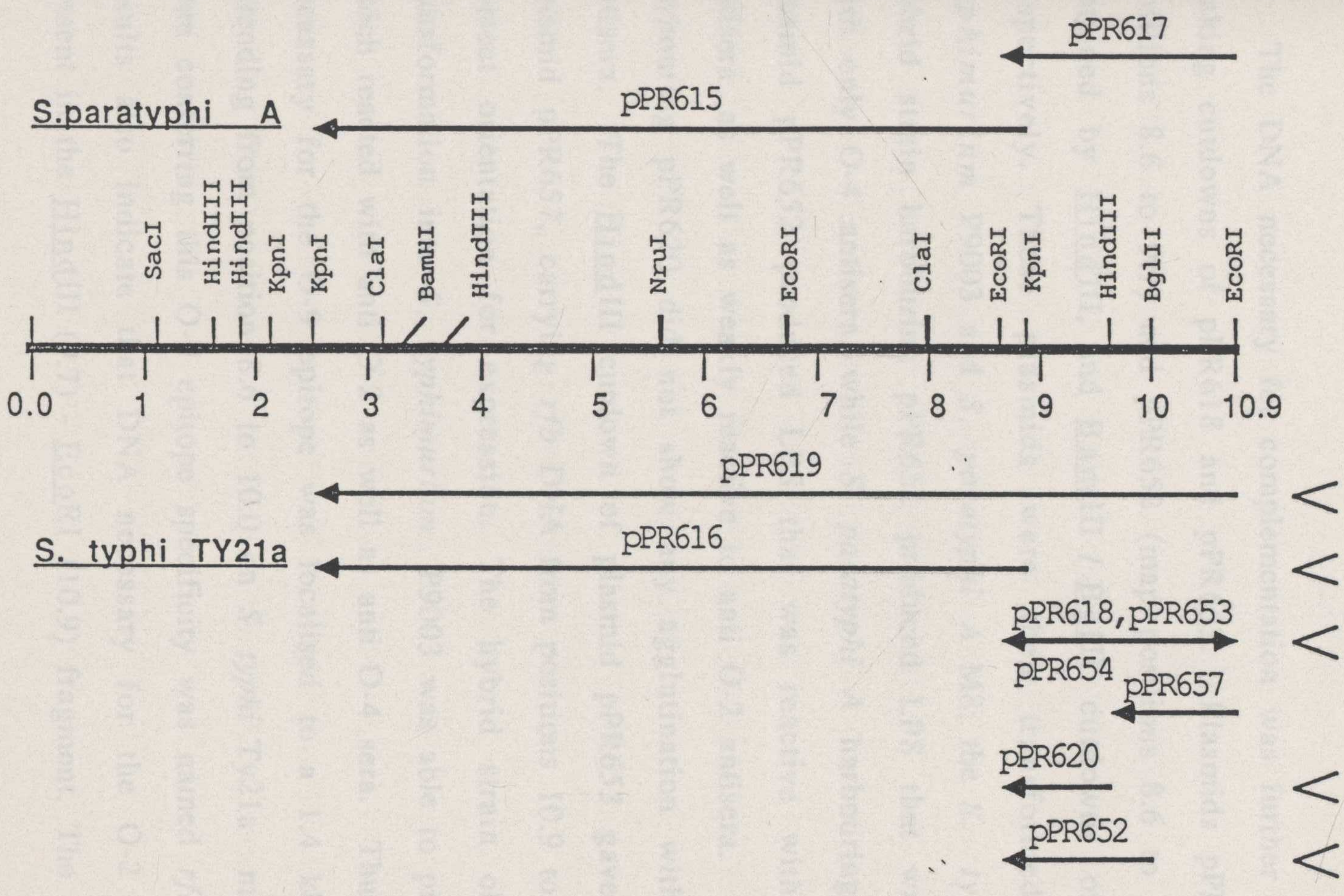
4.2 Expressing O-9 specific LPS in *Salmonellae* of groups A and B, and localising DNA that containing the O-9 specific genes

The *rfb* DNA of *S. typhimurium* LT2 from the *his* operon to position 5.8 and to the right of position 9.57 is shown to have good homology with that of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316, as described in chapter 3. It was thought that the genes responsible for the enzymes conferring the specificity of the 3,6-dideoxyhexose(s) would be encoded in this variable region of DNA, extending from positions 5.8 to 9.5 on the *S. typhi* Ty21a map and positions 5.8 to 15.1 on the *S. paratyphi A* IMVS1316 map. A plasmid, pPR619 (pUC9, carrying *S. typhi* Ty21a *rfb* DNA from positions 2.47 to 10.9, covering the whole variable region), was transformed into a representative strain of group B; *S. typhimurium* LT2 (P9003), and a representative strain of group A; *S. paratyphi A* IMVS1316 (M8). Transformants were purified and single colonies tested against anti O-9 antisera: all the transformants were agglutinated (Fig 4.1) indicating that complementation is possible in strains of *Salmonella* serogroups A, B and D. Hybrid strains, generated from the group B strain, could also agglutinate anti O-4 antisera indicating that two different O-antigens were being expressed in one strain. However, hybrid strains generated from the group A strain did not show any positive reaction with anti O-2 antisera. This suggested that the immunodominant sugar (paratose) had been converted into tyvelose (the immunodominant sugar of group D strains) by an extra enzyme encoded in the cloned DNA of pPR619; only one enzyme addition is required to make CDP-tyvelose from CDP-paratose (Fig 1.4). These results indicate that the genes responsible for making tyvelose (O-9 antigen) lie between *S. typhi* Ty21a map positions 2.47 and 10.9. Two other clones, pPR616 (map position 2.47 to 8.95) and pPR618 (map position 8.6 to 10.9), carrying insert DNA in the correct orientation for expression from the *lac* promoter, were transformed into *S. typhimurium* P9003 and *S. paratyphi A* IMVS1316 (M8) to locate the genes more precisely. These hybrid strains were checked for expression by the slide agglutination test. Only pPR618 conferred expression of O-9 in both *S. typhimurium* P9003 and *S. paratyphi A* M8; indicating that clone pPR618, containing insert DNA bounded by two EcoRI sites at 8.6 and 10.9, has sufficient DNA for complementation (Fig 4.1). This EcoRI piece was then endfilled and subcloned into the SmaI site of pUC18 (Norrander *et al* 1983) in both possible orientations (designated pPR653 and pPR654); the insert in plasmid pPR653

FIGURE 4.1

Localising DNA for specificity of 0 - 2 and 0 - 9 epitopes

Clones that express the antigenic factor 0-9 and 0-2 in *S. typhimurium* (P9003) and antigenic factor 0-9 in *S. paratyphi A* IMVS1316 (M8). Arrowheads indicate the direction of transcription from the external *lac* promoter. Arrowheads on both ends indicate that the insert has been cloned in both orientations. Each clone has been given a plasmid number. 0-2, 0-4 and 0-9 are paratose, abequose and tyvelose specific antisera, respectively, that were used in agglutination tests. S and W in parentheses indicates a strong or weak agglutination reaction. Restriction endonuclease sites shown by thick vertical bars indicate that these sites are repeated three times in *S. paratyphi A* IMVS1316.



Host	0-2	0-4	0-9
P9003	+(w)	+	−
P9003	−	+	+
M 8	−	−	+
P9003	−	+	−
M 8	+	−	−
P9003	−	+	+
M 8	−	−	+
P9003	+	+(s)	−
P9003	−	+	−
M 8	+	−	−
P9003	−	+	−
M 8	+(w)	−	+

being oriented such that the *lac* promoter would read from 10.9 to 8.6. These clones were transformed into *S. typhimurium* P9003 and *S. paratyphi A* M8 and shown to produce anti O-9 reactive LPS. The weak positive agglutination against O-9 antisera in *S. typhimurium* harbouring pPR654 indicates that a weak promoter may be present within the cloned Eco RI fragment and this will be discussed in a later part of this chapter.

The DNA necessary for complementation was further defined by making cutdowns of pPR618 and pPR653. Plasmids pPR620 (map positions 8.6 to 9.7) and pPR652 (map positions 8.6 to 10.0) were obtained by HindIII, and BamHI / BglII cutdowns of pPR618, respectively. These plasmids were then transformed into *S. typhimurium* P9003 and *S. paratyphi A* M8; the *S. typhimurium* hybrid strain harbouring pPR652 produced LPS that was reactive with only O-4 antisera, while *S. paratyphi A* harbouring the same plasmid pPR652 produced LPS that was reactive with anti O-9 antisera, as well as weakly reactive to anti O-2 antisera. The strains harbouring pPR620 did not show any agglutination with anti O-9 antisera. The HindIII cutdown of plasmid pPR653 gave rise to a plasmid pPR657, carrying *rfb* DNA from positions 10.9 to 9.7 in the correct orientation for expression. The hybrid strain obtained on transformation into *S. typhimurium* P9003 was able to produce LPS which reacted with anti O-2 as well as anti O-4 sera. Thus the DNA necessary for the O-9 epitope was localised to a 1.4 kb fragment extending from position 8.6 to 10.0 on *S. typhi* Ty21a map and the gene conferring this O-9 epitope specificity was named *rfbE*. These results also indicate that DNA necessary for the O-2 epitope is present in the HindIII (9.7) - EcoRI (10.9) fragment. The name *rfbS* is allocated to the gene conferring specificity for O-2 epitope. Complementation for O-2 epitope is further discussed in section 4.3.

4.3 Expressing O-2 specific LPS in *Salmonella* of group B and localising DNA containing the O-2 specific gene

In order to localise DNA that specifies the O-2 specific gene from *S. paratyphi A*, plasmids pPR615 (*S. paratyphi A* IMVS1316 map positions 2.47 to 14.55) and pPR617 (*S. paratyphi A* IMVS1316 map positions 14.2 to 16.5) were transformed into *S. typhimurium* P9003. Resultant strains harbouring pPR617 were able to produce anti O-2 as well as anti O-4 reactive LPS; that is, two types of LPS were being expressed in the one strain (Fig 4.1). However, hybrid strains carrying pPR615 did not show any positive agglutination reaction with anti O-2 antisera. These results suggest that the gene for the O-2 epitope lies within this 2.3 kb (14.2 to 16.5) EcoRI fragment. No attempt was made to do cutdowns of plasmid pPR617 since it showed complete restriction site homology with the corresponding 2.3 kb (8.6 to 10.9) EcoRI fragment from *S. typhi* Ty21a (Fig 3.4). It was presumed that the *rfbS* gene would lie within the region (from positions 15.3 to 16.5 on *S. paratyphi A* IMVS1316 map) corresponding to its location (between positions 9.7 to 10.9) on *S. typhi* Ty21a map (see section 4.2). It seems that the *rfbE* gene in *S. paratyphi A* IMVS1316 is present but nonfunctional because pPR617 essentially carries all the DNA present in pPR618 from *S. typhi* Ty21a.

4.4 Cloning strategy for use in DNA sequencing

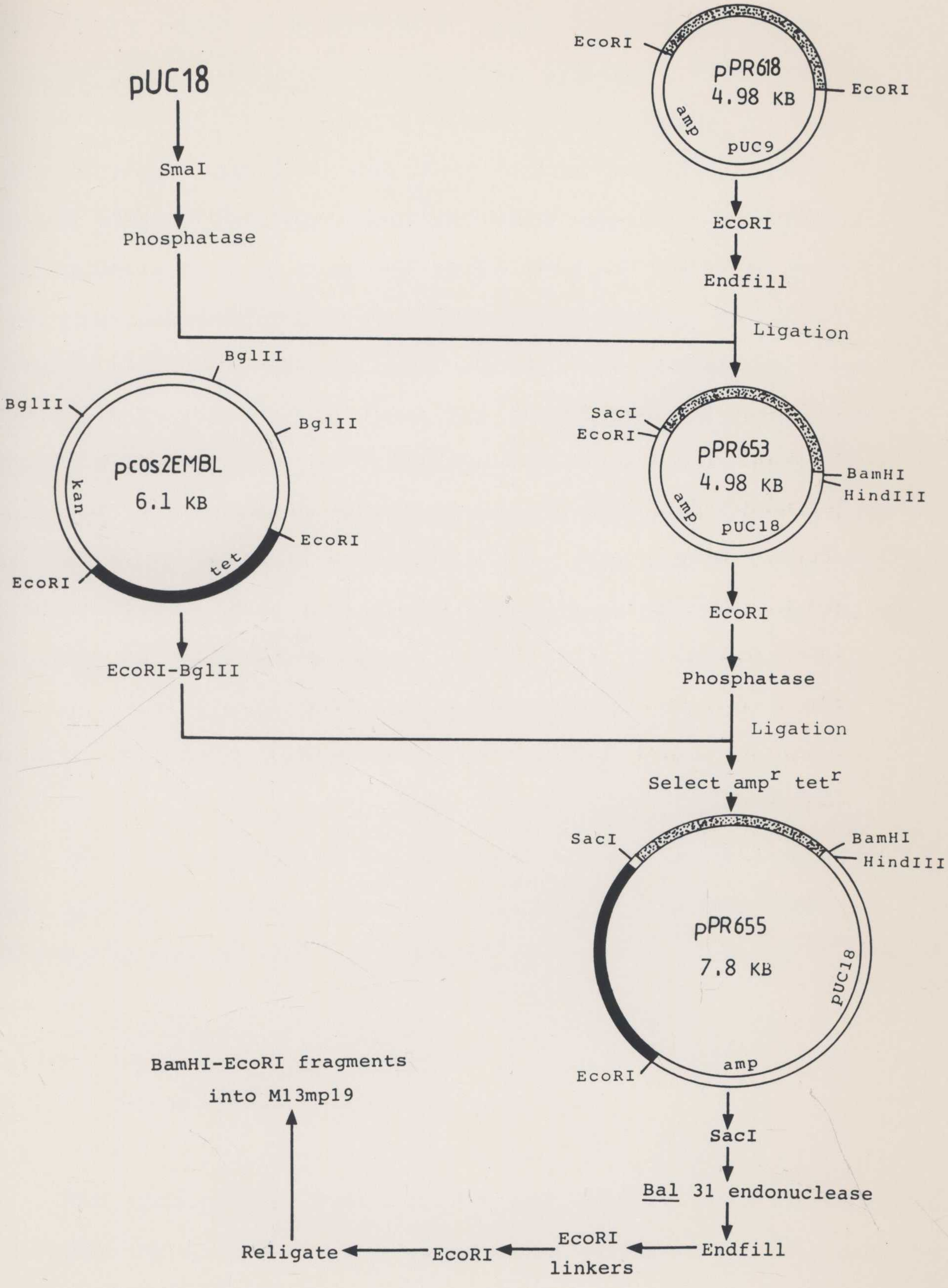
We sequenced the 2.3 kb DNA fragment of *S. typhi* Ty21a bounded by EcoRI sites at positions 8.6 and 10.9 because, as described in sections 4.2 and 4.3, the genes responsible for the enzyme conferring specificity of the 3,6-dideoxyhexose (paratose or tyvelose) are encoded within this region. The EcoRI site at position 10.9 was a convenient restriction site from which to begin sequencing to the left. This site corresponds to the EcoRI site at position 9.57 on the *S. typhimurium* map (Fig 3.14). The gene for an enzyme conferring abequose specificity (*rfbJ*) was also localised left of this position (9.57) in *S. typhimurium* LT2 (P. Wyk, Ph.D. thesis, 1988).

The cloning strategy used for sequencing is shown in Fig 4.2. The 2.3 kb EcoRI *rfb* fragment (positions 8.6 to 10.9) from pPR618 was endfilled and cloned into the SmaI site of pUC18 (Norrander *et al* 1983) in both orientations. The two plasmids were designated pPR653 and pPR654; plasmid pPR653 being orientated such that the *lac* promoter would read from 10.9 to 8.6. A 2.8 kb EcoRI fragment carrying the tetracycline resistance gene from pcos2EMBL was cloned into the EcoRI site of both plasmids as a stuffer segment of DNA. This 2.8 kb EcoRI fragment was obtained by digesting plasmid pcos2EMBL with EcoRI and BglII restriction enzymes (BglII being used to destroy other EcoRI fragment). This digest was then mixed with EcoRI digested and phosphatased pPR653 or pPR654 in a ligation reaction. Tetracycline and ampicillin resistant colonies were screened for the presence of the 2.8 kb fragment, and the selected clones designated pPR655 and pPR656 respectively. These plasmids

FIGURE 4.2

Cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing

The dotted section of the plasmid represents the 2.3 kb EcoR I fragment of *S. typhi* Ty21a used in sequencing. Only the relevant restriction endonuclease sites are shown on each plasmid. A 2.8 kb EcoRI fragment (indicated by the dark section) from pcos2EMBL was inserted as a stuffer fragment at the EcoRI site of pPR653. The resultant plasmid, pPR655, was linearized with SacI and then treated with Bal31 endonuclease for a different periods of time. The inserts from plasmid deletion clones derived by this method were cloned into M13mp19.


pUC18
SmaI
Phosphatase
EcoRI
EcoRI
pPR618
4.98 KB
amp
pUC9
EcoRI
Endfill
Ligation
BglII
BglII
BglII
pcos2EMBL
6.1 KB
kan
tet
EcoRI
EcoRI
SacI
EcoRI
pPR653
4.98 KB
BamHI
HindIII
amp
pUC18
EcoRI-BglII
EcoRI
Phosphatase
Ligation
Select amp^r tet^r
SacI
BamHI
HindIII
pPR655
7.8 KB
pUC18
amp
EcoRI
SacI
Bal 31 endonuclease
Endfill
EcoRI
linkers
EcoRI
Religate
BamHI-EcoRI fragments
into M13mp19

have a convenient unique *Sac*I site, between the 2.3 kb *rfb* DNA and the 2.8 kb stuffer DNA (neither of which contains *Sac*I sites), that can be used to linearize the plasmids. Linearized DNA was then treated with *Bal*31 nuclease. Aliquots were collected at different time intervals, endfilled with DNA polymerase and ligated with *Eco*RI linkers. Resultant clones were then digested with *Eco*RI and self-ligated to remove remaining stuffer DNA. A family of deletions was generated starting from the *Eco*RI site at position 10.9 (Fig 4.3). Selected deletion derivatives were digested with *Bam*HI and *Eco*RI (unique restriction enzyme sites) and cloned into the *Bam*HI and *Eco*RI sites on M13mp19 (Norrander *et al* 1983) for sequencing. A series of M13 clones in reverse orientation were also obtained from pPR656 using the same method (Fig 4.3). The dideoxy sequencing method (Sanger *et al* 1977, 1980) was used to determine the DNA sequence of the cloned region.

4.5 Sequence analysis

The DNA sequence of the various clones was collated using the DB system of Staden (Staden 1982, 1986). Sequence data was thoroughly checked and the sequence obtained is presented in Fig 4.4.

4.5.1 Open reading frames

Two open reading frames of 837 and 1044 bp were identified, corresponding to 279 and 348 amino acids, respectively. The open reading frame of 837 bp was equivalent to the *rfbJ* gene of *S. typhimurium* LT2 (P. Wyk, Ph.D. thesis, 1988) and has been named

FIGURE 4.3

Clones used in sequencing

A restriction endonuclease map of the 2.3 kb EcoRI fragment extending from positions 8.6 to 10.9 from *S. typhi* Ty21a (M18) is shown. Lines represent deletion clones derived from strategy shown in Fig 4.2. Plasmid clones and M13 clones (in parentheses) are shown above each line. The arrows denote the direction in which sequence was obtained.

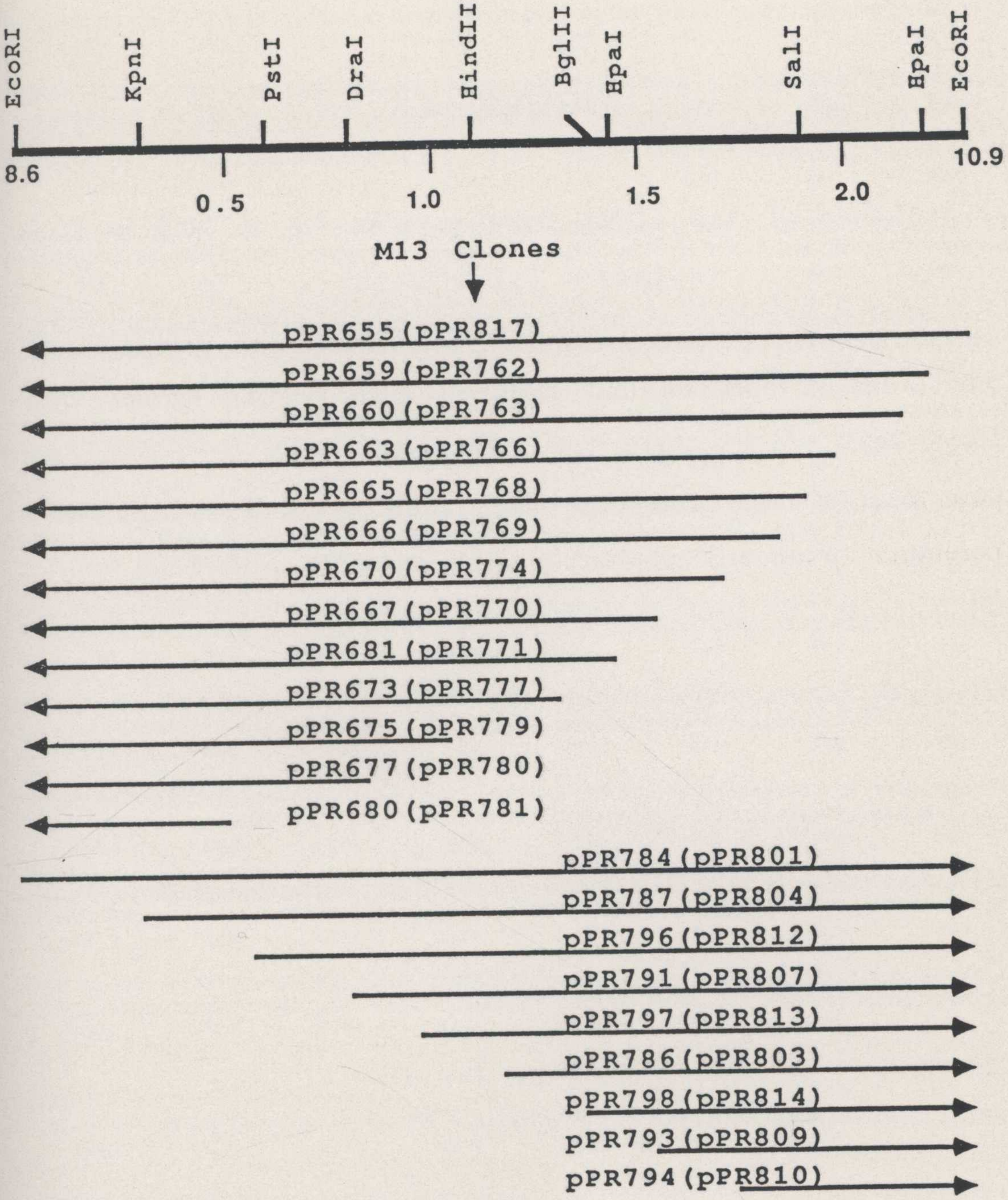

EcoRI
KpnI
PstI
DraI
HindIII
BglII
HpaI
SalI
HpaI
EcoRI
8.6
0.5
1.0
1.5
2.0
10.9
M13 Clones
pPR655(pPR817)
pPR659(pPR762)
pPR660(pPR763)
pPR663(pPR766)
pPR665(pPR768)
pPR666(pPR769)
pPR670(pPR774)
pPR667(pPR770)
pPR681(pPR771)
pPR673(pPR777)
pPR675(pPR779)
pPR677(pPR780)
pPR680(pPR781)
pPR784(pPR801)
pPR787(pPR804)
pPR796(pPR812)
pPR791(pPR807)
pPR797(pPR813)
pPR786(pPR803)
pPR798(pPR814)
pPR793(pPR809)
pPR794(pPR810)

FIGURE 4.4

Sequence of *rfb* DNA from position 8.6 to 10.9 in *S. typhi* Ty21a (M18)

The sequence obtained and proposed open reading frames with amino acids are presented here. The amino acids are denoted by their international three-letter code, stops are indicated by three asterisks. Proposed ribosone-binding sites are underlined.

10.9

```
GAATTCCTCGAATTACCAGAAGCAACAGAGAAATCAGATCCATCCTGGTTTGGCTTCCCTATCACCCTGAAAGAAACTAGCGTGTTAACC
        10        20        30        40        50        60        70        80        90

GTGTCGAACTGGTGAAATTCCTTGATGAAGCAAAAATCGGTACACGTTTACTGTTTGCTGGAAATCTGATTCGCCAACCGTATTTTGCTA
       100       110       120       130       140       150       160       170       180

ATGTGAAATATCGTGTAGTGGGTGAGTTGACAAATACCGACCGTATAATGAATCAAACGTTCTGGATTGGTATTTATCCAGCTTGACTAC
       190       200       210       220       230       240       250       260       270

AGAGCATTTAGATTATGTAGTTAGCAAGTTTGAAGAGTTCTTTGGTTTGAATTTCTAATTTTTGAGGGGGGGGATTCCCTCTATGATTT
       280       290       300       310       320       330       340       350       360

 MetLysIleLeuIleMetGlyAlaPheGlyPheLeuGlySerArgLeuThrSerTyrPheGluSerArgHisThrValIleGlyLeuAla
CATGAAAATTCTAATAATGGGAGCGTTTGGGTTCCTTGGATCACGACTTACATCCTACTTCGAAAGTCGACATACTGTGATTGGCTTAGC
       370       380       390       400       410       420       430       440       450

 ArgLysArgAsnAsnGluAlaThrIleAsnAsnIleIleTyrThrThrGluAsnAsnTrpIleGluLysIleLeuGluPheGluProAsn
AAGGAAGAGGAACAATGAAGCTACCATAAATAATATTATTTACACGACAGAAAATAATTGGATCGAAAAAATACTAGAATTTGAACCGAA
       460       470       480       490       500       510       520       530       540

 IleIleIleAsnThrIleAlaCysTyrGlyArgHisAsnGluProAlaThrAlaLeuIleGluSerAsnIleLeuMetProIleArgVal
TATTATTATTAACACTATTGCTTGCTATGGAAGACATAACGAACCTGCAACAGCTTTAATAGAAAGCAATATTCTTATGCCTATCAGAGT
       550       560       570       580       590       600       610       620       630

 LeuGluSerIleSerSerLeuAspAlaValPheIleAsnCysGlyThrSerLeuProProAsnThrSerLeuTyrAlaTyrThrLysGln
ATTAGAATCTATCTCATCACTTGATGCAGTATTCATAAATTGTGGAACATCACTGCCACCAAATACGAGTTTATATGCATATACTAAACA
       640       650       660       670       680       690       700       710       720

 LysAlaAsnGluLeuAlaAlaAlaIleIleAspLysValCysGlyLysTyrIleGluLeuLysLeuGluHisPheTyrGlyAlaPheAsp
AAAAGCAAATGAACTCGCCGCCGCCATTATAGATAAAGTTTGCGGTAAATATATAGAGTTAAAATTGGAGCATTTCTATGGAGCTTTTGA
       730       740       750       760       770       780       790       800       810

 GlyAspAspLysPheThrSerMetValIleArgArgCysLeuSerAsnGlnProValLysLeuThrSerGlyLeuGlnGlnArgAspPhe
TGGAGACGATAAGTTTACCAGTATGGTTATTAGACGTTGTTTAAGTAACCAGCCAGTAAAGTTAACATCTGGTTTGCAACAGAGAGATTT
       820       830       840       850       860       870       880       890       900

 LeuTyrIleLysAspLeuLeuThrAlaPheAspCysIleIleSerAsnValAsnAsnPheProLysPheHisSerIleGluValGlySer
CTTGTATATAAAAGATCTACTAACAGCGTTCGATTGTATTATAAGTAATGTTAATAATTTCCCCAAATTTCATAGTATTGAAGTTGGTAG
       910       920       930       940       950       960       970       980       990

 GlyGluAlaIleSerIleArgGluTyrValAspThrValLysAsnIleThrLysSerAsnSerIleIleGluPheGlyValValLysGlu
TGGAGAGGCGATATCAATTCGTGAATATGTAGATACTGTTAAAAATATCACAAAAAGCAATTCTATAATTGAATTTGGCGTGGTCAAAGA
      1000      1010      1020      1030      1040      1050      1060      1070      1080

 ArgValAsnGluLeuMetTyrSerCysAlaAspIleAlaGluLeuGluLysIleGlyTrpLysArgGluPheSerLeuValAspAlaLeu
AAGAGTAAATGAATTGATGTATAGTTGTGCTGATATAGCAGAACTTGAAAAAATAGGATGGAAAAGAGAGTTCTCTCTTGTTGATGCATT
      1090      1100      1110      1120      1130      1140      1150      1160      1170

                           MetLysLeuLeuIleThrGlyGlyCysGlyPheLeuGlySerAsnLeuAlaSerPheAlaLeu
 ThrGluIleIleGluGluGluGlyLys***
AACTGAAATAATTGAAGAGGAAGGAAATGAAGCTTTTAATTACCGGTGGATGTGGCTTCCTTGGGAGTAATCTTGCCTCCTTTGCTTTA
      1180      1190      1200      1210      1220      1230      1240      1250      1260

SerGlnGlyIleAspLeuIleValPheAspAsnLeuSerArgLysGlyAlaThrAspAsnLeuHisTrpLeuSerSerLeuGlyAsnPhe
AGTCAAGGGATTGATTTAATTGTATTCGATAATCTATCACGTAAAGGTGCAACAGATAATTTACATTGGTTATCCTCCTTAGGAAACTTT
      1270      1280      1290      1300      1310      1320      1330      1340      1350

GluPheValHisGlyAspIleArgAsnLysAsnAspValThrArgLeuIleThrLysTyrMetProAspSerCysPheHisLeuAlaGly
GAGTTTGTACATGGTGATATTCGCAACAAAAATGATGTTACAAGATTAATAACTAAGTATATGCCTGATAGCTGTTTTCATCTTGCAGGT
      1360      1370      1380      1390      1400      1410      1420      1430      1440
```

```
GlnValAlaMetThrThrSerIleAspAsnProCysMetAspPheGluIleAsnValGlyGlyThrLeuAsnLeuLeuGluAlaValArg
CAAGTGGCAATGACTACATCTATTGACAATCCTTGTATGGATTTTGAAATTAATGTAGGTGGAACTTTAAATTTACTTGAGGCAGTACGG
      1450      1460      1470      1480      1490      1500      1510      1520      1530

GlnTyrAsnSerAsnCysAsnIleIleTyrSerSerThrAsnLysValTyrGlyAspLeuGluGlnTyrLysTyrAsnGluThrGluThr
CAGTATAATTCAAATTGTAATATAATTTATTCATCAACAAATAAAGTATACGGCGATCTTGAGCAATATAAATACAATGAAACAGAAACT
      1540      1550      1560      1570      1580      1590      1600      1610      1620

ArgTyrThrCysValAspLysProAsnGlyTyrAspGluSerThrGlnLeuAspPheHisSerProTyrGlyCysSerLysGlyAlaAla
AGATACACTTGTGTAGATAAGCCTAATGGATATGATGAGAGCACACAATTAGATTTCCACTCACCATATGGTTGTTCAAAAGGTGCTGCA
      1630      1640      1650      1660      1670      1680      1690      1700      1710

AspGlnTyrMetLeuAspTyrAlaArgIlePheGlyLeuAsnThrValValPheArgHisSerSerMetTyrGlyGlyArgGlnPheAla
GACCAATACATGCTTGATTATGCAAGGATTTTTGGTTTGAATACAGTGGTGTTCAGGCATTCATCAATGTATGGTGGAGACAGTTTGCT
      1720      1730      1740      1750      1760      1770      1780      1790      1800

ThrTyrAspGlnGlyTrpValGlyTrpPheCysGlnLysAlaValGluIleLysAsnGlyIleAsnLysProPheThrIleSerGlyAsn
ACTTATGATCAAGGCTGGGTAGGTTGGTTTTGTCAAAAAGCGGTTGAAATTAAAAATGGTATTAATAAACCCTTCACTATTTCTGGTAAT
      1810      1820      1830      1840      1850      1860      1870      1880      1890

GlyLysGlnValArgAspValLeuHisAlaGluAspMetIleSerLeuTyrPheThrAlaLeuAlaAsnValSerLysIleArgGlyAsn
GGTAAGCAAGTTAGGGATGTTTTGCATGCTGAAGATATGATTTCGTTATATTTCACTGCCTTGGCAAATGTATCAAAAATTAGGGGGAAC
      1900      1910      1920      1930      1940      1950      1960      1970      1980

AlaPheAsnIleGlyGlyThrIleValAsnSerLeuSerLeuLeuGluLeuPheLysLeuLeuGluAspTyrCysAsnIleAspMetArg
GCTTTTAATATTGGTGGTACCATTGTCAACAGCCTATCATTACTTGAATTATTCAAATTGCTTGAAGATTATTGCAACATAGATATGAGG
      1990      2000      2010      2020      2030      2040      2050      2060      2070

PheThrAsnLeuProValArgLysValIleSerValPheLeuLeuGlnIleLeuLysLysSerLeuMetGlnLeuThrGlyAlaArgLys
TTCACTAATTTACCTGTAAGGAAAGTGATCAGCGTGTTTTTGTTGCAGATATTAAAAAAATCACTAATGCAATTGACTGGAGCCCGAAAG
      2080      2090      2100      2110      2120      2130      2140      2150      2160

SerArgGlnLysMetValSerArgLysCysMetIleGlyLeuValLeuTyrAspValSerAlaArgGluSerGlyHisAsn***
TCTCGGCAAAAGATGGTGTCCAGAAAATGTATGATTGGACTAGTTCTATATGATGTTTCTGCCCGCGAAAGCGGGCATAATTAAAGAATG
      2170      2180      2190      2200      2210      2220      2230      2240      2250

AAATATTTTTATAATTAAAAGATGAAGCTGACGTGAGGAAACTGAGGTTGGTTAGAATTC
      2260      2270      2280      2290      2300      2310
```

8.6

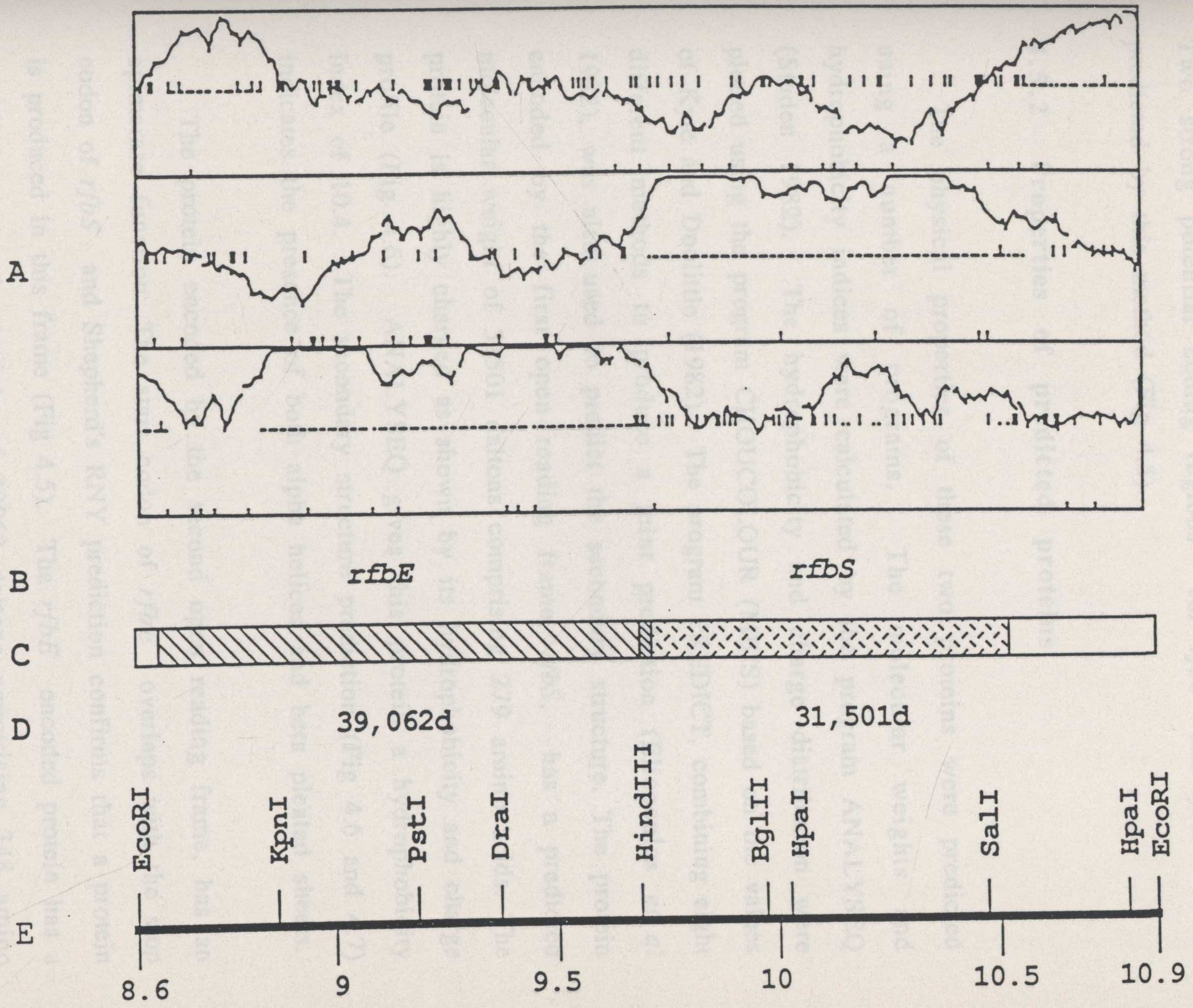
rfbS , while the second open reading frame of 1044 bp is present at the expected location of the epimerase gene, *rfbE*. The first 361 bases from the EcoRI site (position 10.9) may also encode part of the carboxy terminus of a protein, but the length of open-reading frame within this region is too short to confirm the existence of a functional gene.

The sequence was checked for potential ribosome binding sites; that is, sequences showing complementation with the 16S ribosomal RNA sequence (UAAGGAGGUGA) hypothesized to mediate ribosomal binding to mRNA (Shine and Dalgarno 1974). For both open reading frames a probable ribosome-binding site could be found suitably spaced from a methionine codon. These probable ribosome-binding sites are indicated by underlined bases in Fig 4.4. The spacings between start codon and predicted ribosome binding site for *rfbS* and *rfbE*, are 4 and 7 nucleotides, respectively. The median spacing between these two known elements (start codon and ribosome binding site) so far reported in the literature is 7 nucleotides (Gold *et al* 1981, Kozak *et al* 1983) so we are not confident of the ribosome binding site prediction for *rfbS*. We expected two open reading frames from the data shown in section 4.3. Thus these two open reading frames must be those predicted to be present in this region. The gene *rfbS* ends with a single stop codon (TGA) and the start codon (ATG) of *rfbE* overlaps with stop codon of *rfbS*. The overlap of a few nucleotides where one coding region stops and another starts has previously been identified in sequence from other bacterial genes (Yanofsky *et al* 1981). Shepherd's RNY preference method (Shepherd *et al* 1981) was also used to predict genes in this region. Shepherd has found that many genes have a preference for

FIGURE 4.5

Coding regions in the DNA sequence

Shepherd's RNY preference method was used to predict proteins (**A**). three windows in the plot represent the three possible reading frames of the sequence, the uppermost window using the first base of the sequence as the first base of the triplet, the second window uses the second base as the first base of the triplet, and the third window uses the third base as the first base of the triplet. Short vertical lines in the middle of the windows indicate the termination codons. The extent of open reading frames with respect to start and stop codons are shown in section **C**. Section **B** represents the names used for the open reading frames. The estimated molecular weights are listed in Section **D**. **E** shows the restriction endonuclease map of the 2.3 kb EcoRI fragment extending from positions 8.6 to 10.9 from *S. typhi* Ty21a (M18).


A
B
rfbE
rfbS
C
D
39,062d
31,501d
E
EcoRI
KpnI
PstI
DraI
HindIII
BglII
HpaI
SalI
HpaI
EcoRI
8.6
9
9.5
10
10.5
10.9

the use of codons in form RNY where R = purine, Y = pyrimidine, and N = any base (attributed to remnants of a primitive genetic code). Two strong potential coding regions for *rfbS* and *rfbE* were predicted by this method (Fig 4.5).

4.5.2 Properties of predicted proteins

The physical properties of these two proteins were predicted using a number of programs. The molecular weights and hydrophobicity indices were calculated by the program ANALYSEQ (Staden 1982). The hydrophobicity and charge distribution were plotted using the program CHOUCOLOUR (MBIS) based on the values of Kyte and Doolittle (1982). The program PREDICT, combining eight different methods to produce a joint prediction (Eliopoulos *et al* 1982), was also used to predict the secondary structure. The protein encoded by the first open reading frame, *rfbS,* has a predicted molecular weight of 31501 daltons comprising 279 amino acids. The protein is highly charged as shown by its hydrophobicity and charge profile (Fig 4.6). ANALYSEQ gives this protein a hydrophobicity index of 10.4. The secondary structure prediction (Fig 4.6 and 4.7) indicates the presence of both alpha helices and beta pleated sheets.

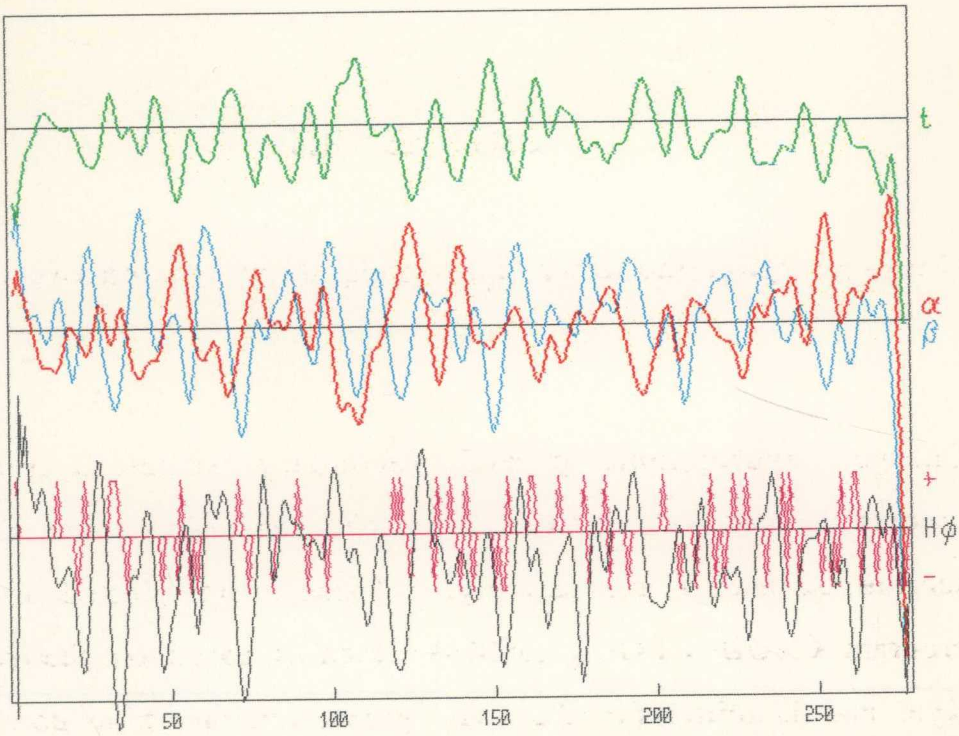
The protein encoded by the second open reading frame, has an epimerase function. The start codon of *rfbE* overlaps with the stop codon of *rfbS* and Shepherd's RNY prediction confirms that a protein is produced in this frame (Fig 4.5). The *rfbE* encoded protein has a predicted molecular weight of 39062 daltons comprising 348 amino acids, with a hydrophobicity index of 12.5. The charge/hydrophobicity profile shows that the molecule consists of

FIGURE 4.6

Charge, hydrophobicity and secondary structure plots of predicted proteins

Charge, hydrophobicity and secondary structure plots for predicted proteins encoded by *rfbS* and *rfbE* are shown in section **A** and **B** respectively. These were plotted using the program CHOUCOLOUR (MBIS) which is based on the values of Kyte and Doolittle (1982). The green, red and blue colour plots represent turn, alpha helix and beta strand prediction. Charge distribution (positive charges point up, negative charges point down) and hydropathy plots are shown by maroon and black colour respectively. The computation for α-helix, β-sheet and turn prediction is based on Chou and Fasman (1978) parameters and algorithm. The method and parameters of Rose and Roy (1980) are used for computation of hydropathy profiles.

A



B

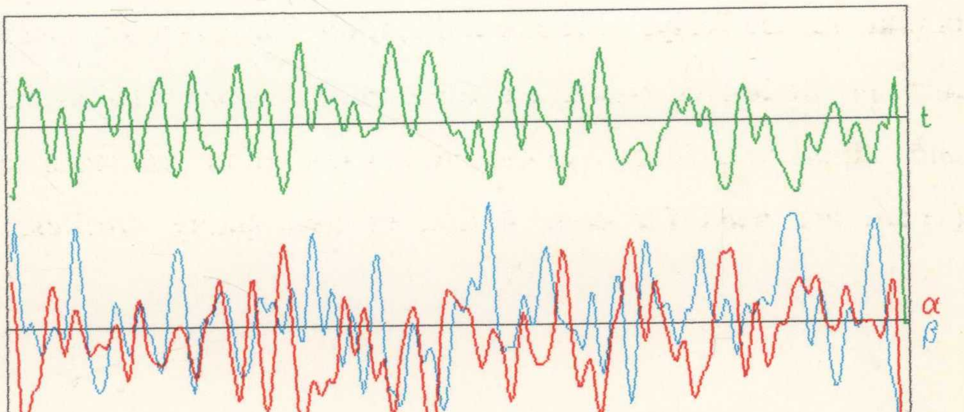
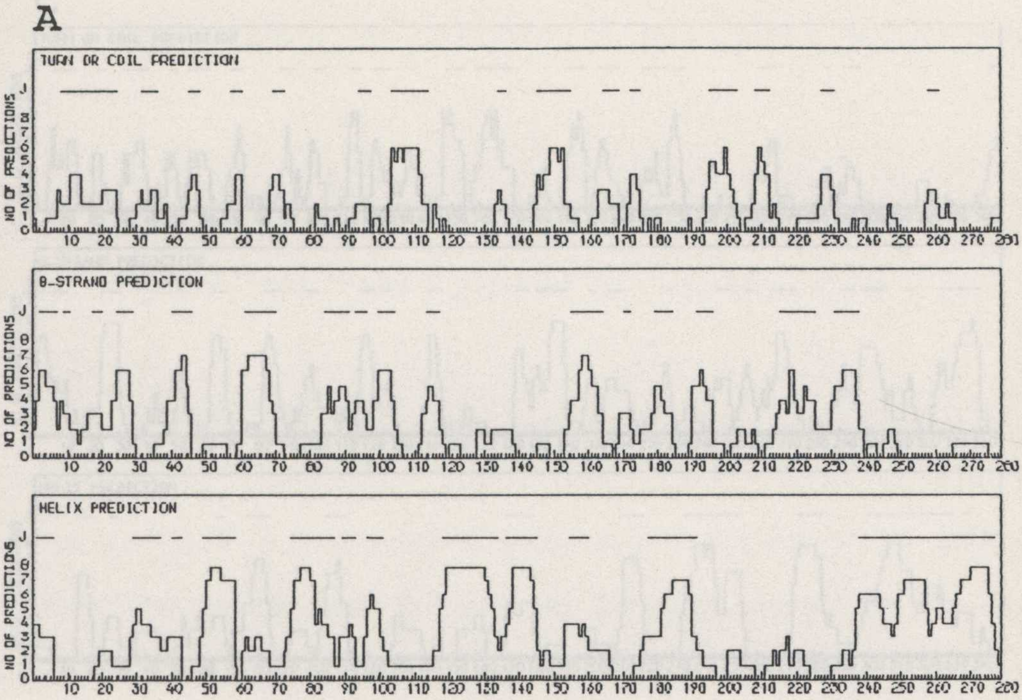
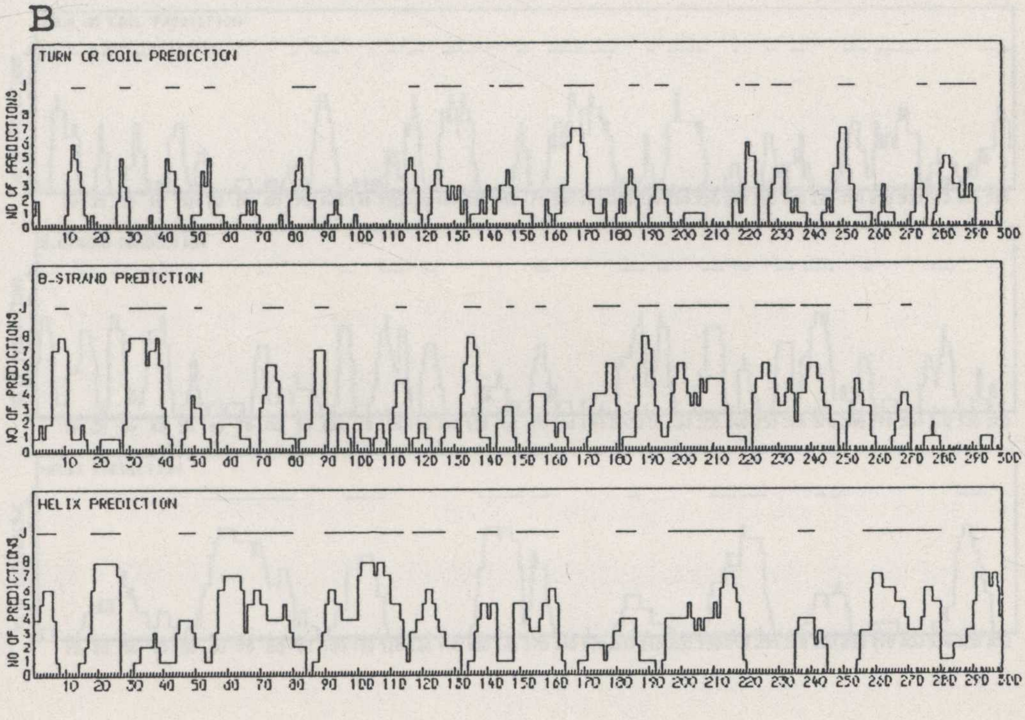


FIGURE 4.7

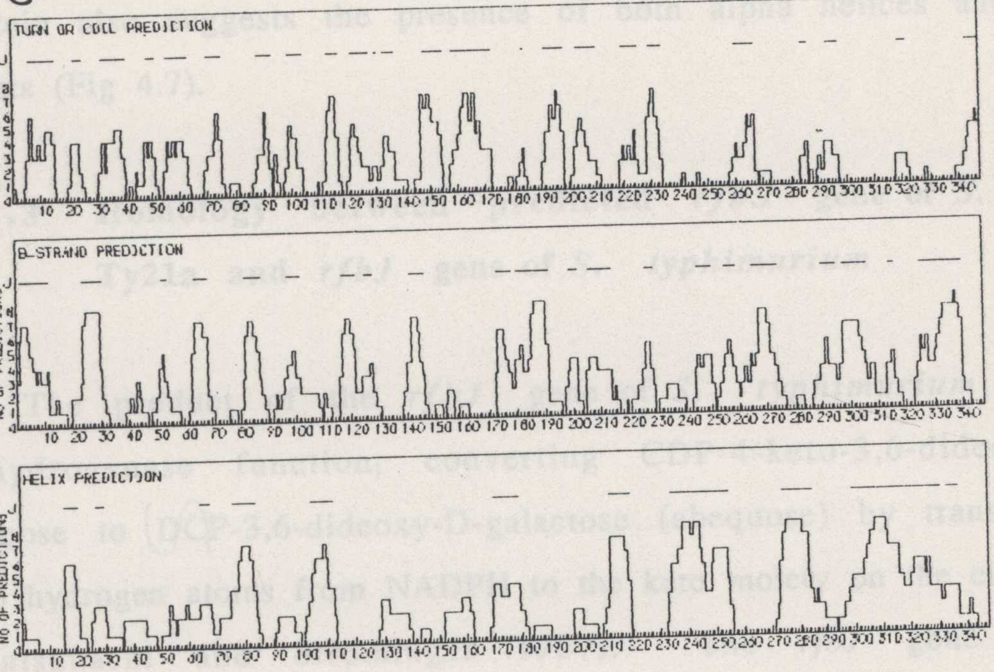
Possible secondary structures of predicted proteins

Section **A**, **B**, **C** and **D** show secondary structure prediction plots for the proteins encoded by *rfbS* (*S. typhi* Ty21a), *rfbJ* (*S. typhimurium* LT2) *rfbE* (*S. typhi* Ty21a) and *galE* (*E. coli*), respectively. The program PREDICT (Eliopoulos *et al* 1982) was used to predict secondary structure; this gives a consensus of the eight different methods to produce a joint prediction. The joint prediction is in the form of a broken line across the top of each window. Each protein has three plots labelled as turn or coil, beta-strand and helix prediction. The number of predictions is shown by the numbers on left the vertical line and the bottom horizontal line shows the number of amino acids.

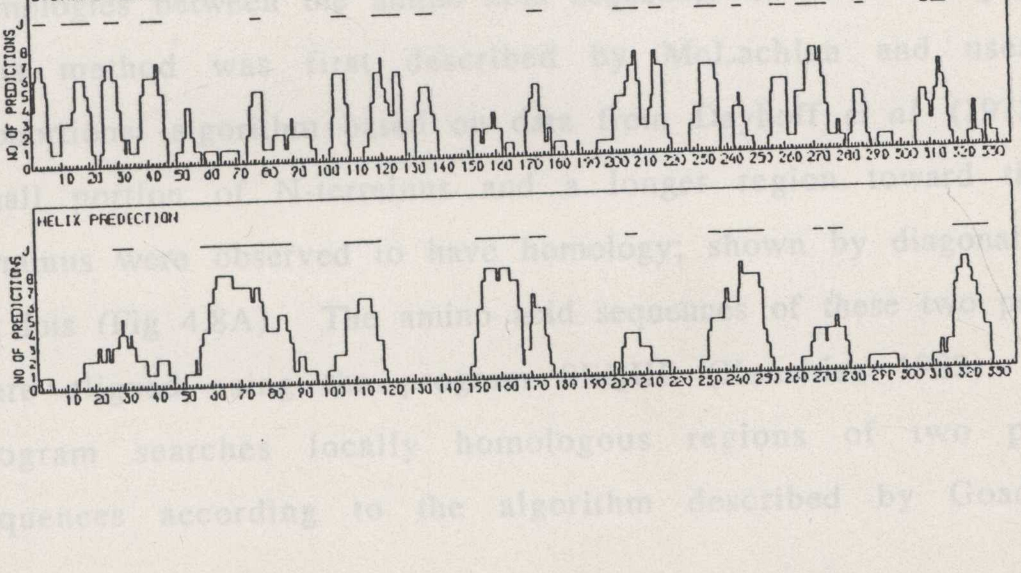

A
TURN OR COIL PREDICTION
B-STRAND PREDICTION
HELIX PREDICTION
NO OF PREDICTIONS
J
0 1 2 3 4 5 6 7 8
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280


B
TURN OR COIL PREDICTION
B-STRAND PREDICTION
HELIX PREDICTION
NO OF PREDICTIONS
J
0 1 2 3 4 5 6 7 8
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300

C


TURN OR COIL PREDICTION
B-STRAND PREDICTION
HELIX PREDICTION
NO OF PREDICTIONS
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340

D


TURN OR COIL PREDICTION
B-STRAND PREDICTION
HELIX PREDICTION
NO OF PREDICTIONS
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330

charged amino acids. The predicted secondary structure of this protein also suggests the presence of both alpha helices and beta sheets (Fig 4.7).

4.5.3 Homology between predicted *rfbS* gene of *S. typhi* Ty21a and *rfbJ* gene of *S. typhimurium*

The product of the *rfbJ* gene of *S. typhimurium* has a dehydrogenase function, converting CDP-4-keto-3,6-dideoxy-D-glucose to DCP-3,6-dideoxy-D-galactose (abequose) by transferring two hydrogen atoms from NADPH to the keto moiety on the carbon 4 (Matsuhashi and Strominger 1967). The *rfbJ* gene of *S. typhimurium* was sequenced recently in our laboratory (P. Wyk, Ph.D. thesis, 1988). The product of *rfbS* has a similar dehydrogenase function using the same substrate, CDP-4-keto-3,6-dideoxy-D-glucose, for the formation of CDP-3.6-dideoxy-D-glucose (paratose), and therefore we expected it to have homology with the protein encoded by *rfbJ* in *S. typhimurium.*

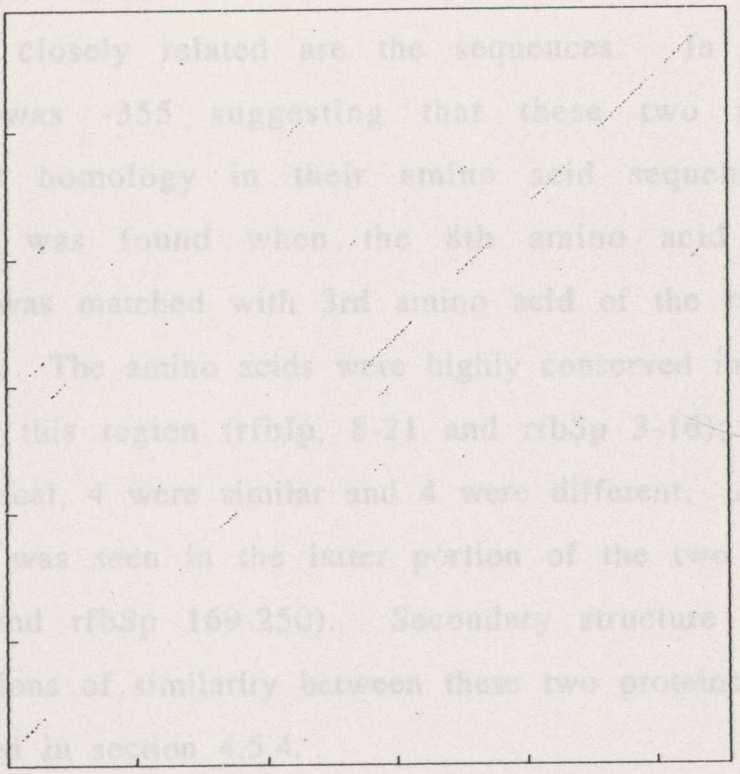
We used a program, DIAGON (Staden 1982), to search for homologies between the amino acid sequences of these two proteins. This method was first described by McLachlan and uses the proportional algorithm based on data from Dayhoff *et al* (1978). A small portion of N-terminus and a longer region toward the C-terminus were observed to have homology; shown by diagonal rows of dots (Fig 4.8A). The amino acid sequences of these two proteins were aligned using the program SEQHP (Kanehisa 1982). This program searches locally homologous regions of two protein sequences according to the algorithm described by Goad and

FIGURE 4.8

Relationship of the amino acid sequence of proteins encoded by the *rfbS* gene of *S. typhi* Ty21a and *rfbJ* gene of *S. typhimurium*

A shows the output of the program DIAGON (Staden 1982) used to show the similarity of amino acid sequence of these two proteins. This method was first described by McLachlan and uses the proportional algorithm based on the data of Dayhoff *et al* (1978). The diagram consists of two dimensional plot in which the X-axis represents amino acids of rfbSp and the Y-axis the amino acid sequence of rfbJp of *S. typhimurium.* A span length of 11 amino acids and percent proportional score of 132 were used. **B** shows the alignment of amino acid sequence of these two proteins (8 - 297 of rfbJp and 3 - 275 of rfbSp) made by the program SEQHP. Distance (maximum homology score) refers to the relationship between the proteins. Maximum homology score is sum of weights of all matches, mismatches, and deletions. The weight for a deletion and a value to be added to each element of the similarity matrix were 8 and 0, respectively. The numbers in the parentheses denote the number of gaps added to the sequence to make the best alignment.

A



B

```
ALIGNMENT   1,  DISTANCE  -355,     8- 297 ( 1),     3- 275 (18)

 10        20        30        40        50        60        70        80
VIVSGASGFIGKHLLEALKKSGISVVAITRDVIKNNSNALANVRWCSWDNIELLVEELSIDSALIGIIHLATEYGHKTSS
    :: :: :  :          :    :    ::     :       :       :      : :   :  ::
ILIMGAFGFLGSRLTSYFESRH TVIGLARK  RNNEATINNIIYTT EN NWIEKILEFEPNII INTIAC YGRHNEP
      10        20         30          40          50        60          70
 90       100       110       120       130       140       150       160
LINIEDANVIKPLKLLDLAIKYRADIFLNTDSFFAKKDFNYQHMRPYIITKRHFDEIGHYYANMHDISFVNMRLEHVYGP
       :   :   :   :      : :           :     :   :                      ::: ::
ATALIESNILMPIRVLE SISSLDAVFINCGTSLPPNTSLY    AYTKQKAN ELAAIIDKVCG KYIELKLEHFYGA
   80        90        100       110           120        130        140
170       180        190       200       210       220       230       240
GDGENKFIPYIID CLNKKQSCVKCTTGEQIRDFIFVDDVVNAYLTILENRKEVPSYTEYQVGTGAGVSLKDFLVYLQNT
 ::  ::    :  ::   :  :: : : : :::    :   :   :  :    :      :: :   :         :
FDGDDKFTSMVIRRCLSN QP VKLTSGLQQRDFLYIKDLLTAFDCIISNVNNFPKFHSIEVGSGEAISIREYVDTVKN
         160         170       180       190       200       210       220
 250       260       270       280       290
MMPGSSSIFEFGAIEQRDNEIMFSVANNKNLKAMGWKPNFDYKKGIEELLK
    : :: :::    : :: : : :    :   :::  :       :
ITK SNSIIEFGVVKERVNELMYSCADIAELEKIGWKREFSLVDALTEIIE
   230       240       250       260       270
```

Kanehisa (1982); the lower the distance number (see Fig legend 4.8), the more closely related are the sequences. In this case, the distance was -355 suggesting that these two proteins share substantial homology in their amino acid sequences. The best alignment was found when the 8th amino acid of the rfbJp sequence was matched with 3rd amino acid of the rfbSp sequence (Fig 4.8B). The amino acids were highly conserved in the N-terminal region; in this region (rfbJp, 8-21 and rfbSp 3-16); 6 amino acids were identical, 4 were similar and 4 were different. A high level of homology was seen in the latter portion of the two proteins (rfbJp 189-272 and rfbSp 169-250). Secondary structure prediction also shows regions of similarity between these two proteins (Fig 4.7); this is discussed in section 4.5.4.

The program SEQA (Kanehisa 1982) was used to search for homologous regions between nucleotide sequences of *rfbJ* and *rfbS*. This program finds global homology according to the algorithm described by Needleman and Wunsch (1970). Best alignment of these two sequences was achieved by matching the regions which showed homology in SEQHP for this program (Fig 4.9A). The program found 482 matches, 341 replacements and 79 deletions or insertions; of these deletions/insertions, 17 were in the *rfbJ* and 62 were found in the *rfbS* nucleotide sequence. The program DIAGON was also used to check homology between the nucleotide sequences of *rfbJ* and *rfbS* genes (Fig 4.9B). Scattered regions of homology are seen as shown by the diagonal lines of dots along the diagonal axis.

FIGURE 4.9

Relationship of the DNA sequence of the genes *rfbS* and *rfbJ*

A) The nucleotide sequence of both genes *rfbS* and *rfbJ* are aligned using the program SEQA (Kanehisa 1982). Only one alignment is shown here. The values obtained by alignments are shown by the bold letters at the bottom line.

B) Shows the output of the program DIAGON. The X-axis represents the nucleotide sequence of *rfbS* and the Y-axis represents the nucleotide sequence of *rfbJ*. A span length of 11 nucleotides and percent proportional score of 8 were used. Diagonal rows of dots are used to show homology.

A

```
   20        30        40        50         60         70        80        90
GAATATGTAATTGTCAGTGGGGCTTCCGGCTTTATTGG-TAAGCATTTAC-TCGAAGCGCTAAAAAAATCGGGGATTTCA
    :  :  :  : :  :: :: :  :: ::  :::: : :  : :::: ::  :   ::   ::: :: :  : : :
ATGAAAATTCTAATAATGGGAGCGTTTGGGTTCCTTGGATCACGACTTACATCCTA---CTTCGAAAGTC-GACA-TACT
        10        20        30        40        50           60         70
    100       110       120       130       140       150       160       170
GTTGTCGCAATCACTCGAGATGTAATAAAAAATAATAGTAATGCATTAGCTAATGTTAGATGGTGCAGTTGGGATAATAT
::  : :   :  :  : :: :  :  :: ::: : ::  :  :: ::  :::: ::  ::  : ::   :  : :: ::
GTGATTGGCTTAGCAAG-GAAG--AGGAACAATGA-AGCTA-CCA-TAAATAATATT--AT-TTACA--CGACAGAAAAT
   80        90           100        110        120        130          140
    180       190       200       210       220       230       240       250
CGAATTATTAGTCGAGGAGTTATCAATTGATTCTGCATTAATTGGTATCATTCATTTGGCAACAGAATATGGGCATAAAA
  ::::   : ::::  :   :: : : :: : :   : ::   : :: ::: : ::   :::  : :::  :: :
--AATT-GGA-TCGA--AAAAAT-ACTAGAAT-T---TGAA-CCGAAT-ATT-A-TTATTAAC--ACTAT-TGC-T--TG
       150          160         170           180          190          200
    260       270       280       290       300       310       320        330
CATCATCTCTCATAAATATTGAAGATGCAAATGTTATAAAACCATTAAAGCTTCTTGATTTGGCAAT-AAAATATCGGGC
: : :  :   : : :::  :::  :::::  : : ::: :  :   ::  ::::: :  :: : :: : : :::   :
C-T-A--TGGAAGACATAACGAACCTGCAACAGCTTTAATAGAAAGCAATATTCTT-A--TGCCTATCAGAGTAT-TAGA
         210       220       230       240       250         260       270
     340       350       360       370       380       390       400         410
GGATATCTTTTTAAATACAGATAGTTTTTTTGCCAAGAAAGATTTTAATTATCAACATATGCGGCCTTATAT-AAT-TAC
   :::: :  : : :  :   ::: ::  :   :     ::   : : :  :: :: :: ::   ::::::  :: :::
ATCTATC-TCATCACTTGATGCAGTATTCAT-AAATTGTGGAACATCACTGCCACCAAATACGAGTTTATATGCATATAC
   280        290       300        310       320       330       340       350
       420       430         440        450       460       470       480
TAAAAGACACTTTGATGAAATTG--GGCATTATTAT-GCTAATATGCATGACATTTCATTTGTAAACATGCGATTAGAGC
::::  : :     ::::: : :  : :   ::::: : ::: : :  :: :  :  :: : :: :  :   ::: ::::
TAAACAAAAAGCAAATGAACTCGCCGCCGCCATTATAGATAA-A-GTTTG-CGGTAAATATATAGAGTTAAAATTGGAGC
     360       370       380       390         400        410       420       430
490       500       510       520       530         540       550       560
ATGTATATGGGCCTGGGGATGGTGAAAATAAATTTATTCCATA-CATTA-TCGACTGCTTAAATAAAAAACAGAGTTGCG
:: : :::::  ::   ::::: ::  :::: ::::  :  ::   ::: : ::: : ::   :::  :::  ::    :
ATTTCTATGGAGCTTTTGATGGAGACGATAAGTTTA-CCAGTATGGTTATTAGAC-G-TTGTTTAAGTAAC-CAG-CCAG
        440       450       460        470       480         490        500
  570       580       590       600       610       620       630       640
TGAAATGTACAACAGGCGAACAGATAAGAGACTTTATTTTTGTAGATGATGTGGTAAATGCTTATTTAACTATATTAGAA
: :: :  ::: : ::    ::    ::::: ::  : : : :: : ::: :  :::  ::   ::  : : ::::: ::
TAAAGTTAACATCTGGTTTGCAACAGAGAGATTTCTTGTATATAAAAGATCTACTAACAGC--GTTCGATTGTATTATAA
   510       520       530       540       550       560         570       580
   650       660       670        680       690       700       710       720
-AATAGAAAAGAAGTACCTTCATATACTGAGTA-TCAAGTTGGAACTGGTGCTGGGGTAAGTTTGAAAGATTTTCTGGTT
  :  :  :: :: : ::   :  : :  :::: : ::::::: : ::: :  : : ::    :    :: : : : :
GTAATGTTAATAATTTCCCCAAATTTCATAGTATTGAAGTTGGTAGTGGAGAGGCGATATCAATTCGTGAATATGTAG-A
     590       600       610       620       630       640       650       660
    730       740       750       760       770       780       790       800
TATTTGCAAAATACTATGATGCCAGGTTCATCGAGTATATTTGAATTTGGTGCGATAGAGCAAAGAGATAATGAAATAAT
::  ::  ::: : ::: :  : :    ::     :::: :::::::::: : : :  :  ::::::  :::::: : ::
TA-CTGTTAAA-AATATCA--CAAAAAGCA-ATTCTATAATTGAATTTGGCGTGGTCAAAGAAAGAGTAAATGAATTGAT
       670          680        690        700       710       720       730
     810       820        830       840       850         860        870
GT-TCTCTGTAGCAAATA-ATAAAAATTTAAAAGCAATGGGCTGGAAACCA-AATT-TCGATT-ATAAAAAAGGAATTG-
:: :   ::: ::  ::: :  : :: :: :::  ::: :: ::::::  : : :: ::  ::  : :   :  :: ::
GTATAGTTGT-GCTGATATAGCAGAACTTGAAA-AAATAGGATGGAAAAGAGAGTTCTCTCTTGTTGATGCATTAACTGA
 740        750       760         770       780       790       800       810
```

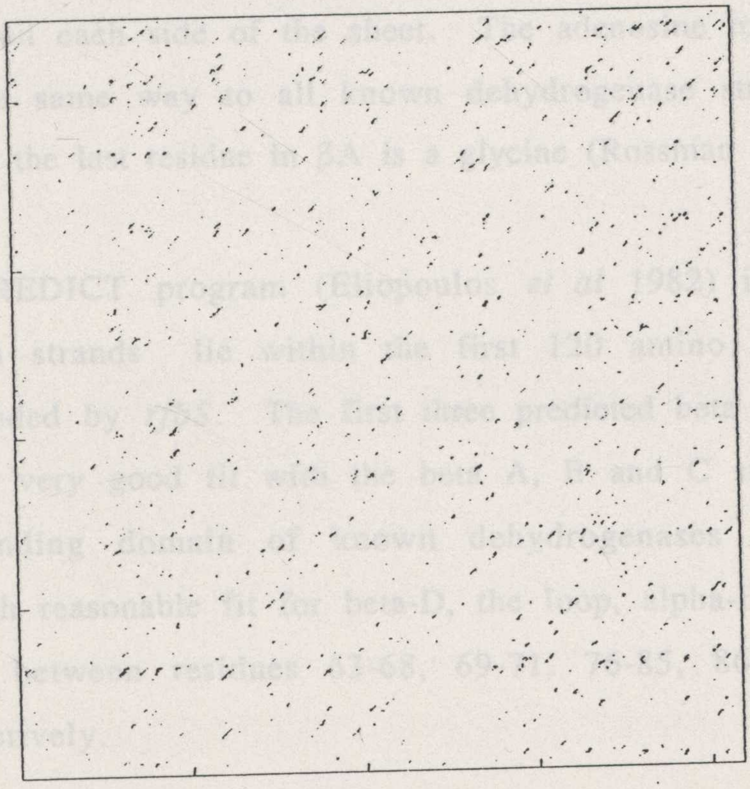
```
880        890        900
AAGAACT--ACTGAAACGGTTATGA
:: :: :   :   : :: ::  ::::
AATAATTGAAGAGGAAGGGAAATGA
  820        830        840

MATCHES    489
  AA  197    CC    39    GG    84    TT  169
REPLACEMENTS    331
  AC    56    AG    73    CG    34    AT    85    CT    40    GT    43

DELETIONS IN SEQUENCE 1:        20
DELETIONS IN SEQUENCE 2:        65

MATCHES 482    REPLACEMENTS 341    DEL/INS 79  (Seq1.17;  Seq2.  62)
```

B



4.5.4 Structural similarity of the dehydrogenases with other known dehydrogenases

The function of the protein encoded by *rfbS* is similar to that of the protein encoded by *rfbJ* of *S. typhimurium*. The enzyme is a nicotinamide linked dehydrogenase. Dinucleotide binding proteins, such as NAD^+-linked dehydrogenases of the citric acid cycle in eukaryotes, have been extensively studied by X-ray crystallographic techniques. There is a striking similarity in the structure of the coenzyme (NAD) binding domains of all the NAD^+-linked dehydrogenases (Table 4.1), though the catalytic domains have very different structures. The main structural elements of the NAD^+ binding domain are six strands forming a parallel sheet (βA, βB, βC, βD, βE, and βF) and four helices (αB, αC, αE, and α1F). Two helices are located on each side of the sheet. The adenosine moiety (AMP) binds in the same way to all known dehydrogenase structures, and in each case the last residue in βA is a glycine (Rossman *et al* 1975).

The PREDICT program (Eliopoulos *et al* 1982) indicates that several beta strands lie within the first 120 amino acids of the protein encoded by *rfbS*. The first three predicted beta strands (1 to 47) show a very good fit with the beta A, B and C strands of the NAD^+ binding domain of known dehydrogenases (Table 4.1). Regions with reasonable fit for beta-D, the loop, alpha-E, beta-E and beta-F, lie between residues 63-68, 69-71, 76-85, 86-92 and 98-105, respectively.

The protein encoded by the *rfbJ* gene of *S. typhimurium* (P. Wyk, Ph.D. thesis, 1988) also has beta strands and alpha helices that

TABLE 4.1

Amino acid sequence comparison based on structural alignments

The alignment of amino acids of proteins encoded by *rfbJ* of *S. typhimurium, rfbS* (*S. typhi* Ty21a), *rfbE* (*S. typhi* Ty21a) and *galE* (*E. coli*) is shown in relation to amino acids from the NAD^+ binding domain in structurally known dehydrogenases. The alignment of amino acid sequences of Dogfish LDH (lactate dehydrogenase), Pig GADPH (glyceraldehyde-3-phosphate dehydrogenase), Lobster GADPH, Horse LADH (liver alcohol dehydrogenase), Rat LADH and Bovine GluDH (glutamate dehydrogenase) is adapted from Table IV of Rossman *et al* (1975). The alignment of rfbJp of *S. typhimurium* LT2 is adapted from P. Wyk Ph.D. thesis (1988). Each amino acid is designated by its International single-letter code and the amino acids which are similar to those of proposed function in the table published by Rossman *et al* (1975) are boxed. The underlined letters H, S and F denote the amino acids are present opposite an alpha helix, opposite a beta strand or have substrate binding function respectively.

						beta-A													beta-B											beta-C		
Dogfish LD	N	K	I	T	V	V	G	C	B	A	V	G	D	E	V	A	L	V	D	V	M	E	D	K	A	K	I	V	S	G	K	D
Pig GADPH	V	K	V	G	V	D	G	F	G	R	I	G	D	I	V	A	I	N	D	P	F	I	D	L	K	A	I	T	I	F	Q	E
Lobster GADPH	S	K	I	G	I	D	G	F	G	R	I	G	Q	V	V	A	V	N	D	P	F	I	A	L	K	K	I	T	Y	F	N	E
Yeast GADPH	V	R	V	A	I	D	G	F	G	R	I	G	Z	V	V	A	S	B	B	P	F	I	B	L	K	K	I	A	T	Y	Q	E
Horse LADH	S	T	C	A	V	F	G	L	G	G	V	G	A	R	I	I	G	V	D	I	N	K	D	K	G	A	T	E	C	I	N	P
rat LADH	S	T	C	A	V	F	G	L	G	G	V	G	A	K	I	I	A	V	D	I	N	K	D	K	G	A	T	D	C	I	N	P
Bovine GluDH	K	T	F	A	V	Q	G	F	G	N	V	G	K	C	V	A	V	G	E	S	D	G	S	I								
rfbJp	E	Y	V	I	V	S	GA	S	G	F	I	G	S	V	V	A	I	T	R	V	I	K	N	N	A	L	I	G	I	I	H	L
rfbSp	M	K	I	L	I	M	GA	F	G	F	L	G	T	V	I	G	L	A	R	K	R	N	N	E	N	N	I	I	Y	T	T	E
rfbEp	M	K	L	L	I	T	GG	C	G	F	L	G	I	D	L	I	V	F	D	N	L	S	R	K								
galEp	M	R	V	L	V	T	GG	S	G	Y	I	G	H	D	V	I	I	L	D	N	L	C	N	S								
			H		H		F					S			H		H		F								H		H		H	

			beta-D					loop									alpha-e									beta-E							beta-F				
Dogfish LDH	K	L	V	V	I	T	A	G	A	R	K	F	I	I	P	N	I	V	K	H	S	P	D	C	I	L	E	L	H	P	H	R	I	I	G	S	G
Pig GAPDH	A	Y	V	V	E	S	T	G	V	F	M	E	K	A	G	A	H	L	K		G	G	A	K	R	V	I	I	S	A	L	K	I	V	S	N	A
Lobster GAPDH	E	Y	I	V	E	S	T	G	V	S	I	E	K	A	S	A	H	F	K		G	G	A	K	K	V	V	I	S	A	M	T	V	V	S	N	A
Yeast GAPDH	V	I	A	I	R	S	T	G	V	S	L	D	T	A	Q	K	H	L	K		A	G	A	K	K	V	V	I	T	A	L	K	I	V	S	N	A
Horse LADH	D	F	X	F	E	V	I	G			L	D	T	M	V	T	A	L	S	C	C	Q	Y	G	V	S	V	I	V	G	R	T	W	K	G	A	I
Rat LADH	D	F	X	F	E	V	I	G			L	D	T	M	A	X	X	L	L	S	C	H	C	G	V	S	V	I	V	G	R	T	W	K	G	A	I
rfbJp	I	I	H	L	A	T	Y	G	H		L	I	N	I	E	D	A	N	V	I				K	P	L	K	L	L	D			I	F	I	WT	D
rfbSp	I	N	T	I	A	C	Y	G	R		A	T	A	L	I	E	S	N	I	L				M	P	I	R	V	L	E	D	A	V	F	I	NC	G
			H		H			F				S	S			S	S									H		H					H		H		H

fit the proposed NAD+ binding domains of several other dehydrogenases (Table 4.1). The first three beta strands (βA, βB and βC) also match well with the protein encoded by *rfbS*. This, and the similar function of these two proteins, suggests that the first 105 amino acids of protein encoded by *rfbS* and the first 140 amino acids of protein encoded by *rfbJ* may form an NADPH binding domain. The latter portions of rfbSp (105-279) and rfbJp (140-299) are probably involved in binding and modifying the substrate; both enzymes require the same substrate (4-keto-3,6-dideoxy-D-glucose) for the reaction.

4.5.5 Comparison of CDP-paratose-2-epimerase (*rfbE*) with UDP-galactose-4-epimerase (*galE*)

The function of the *rfbE* gene product is not chemically well characterized. The enzyme is an epimerase acting on carbon 2 to convert CDP-3,6-dideoxy-D-glucose (paratose) to CDP-3,6-dideoxy-D-mannose (tyvelose), using NAD as a coenzyme. The predicted protein product for *rfbE* was compared with an enzyme of similar function, UDP-galactose-4-epimerase which acts on carbon 4 in the same way. This enzyme is encoded by the *galE* gene which has recently been sequenced in *E. coli* (Lemaire and Müller-Hill 1986).

The protein encoded by the gene *rfbE* comprises 348 amino acids, which is quite similar in length to the protein encoded by *galE* (338 amino acids). The program DIAGON (Staden 1982) was used to search for homology between the amino acid sequences of these two proteins. Small regions of homology are seen at the N-terminus portion as shown by the diagonal rows of dots (Fig 4.10A). We used

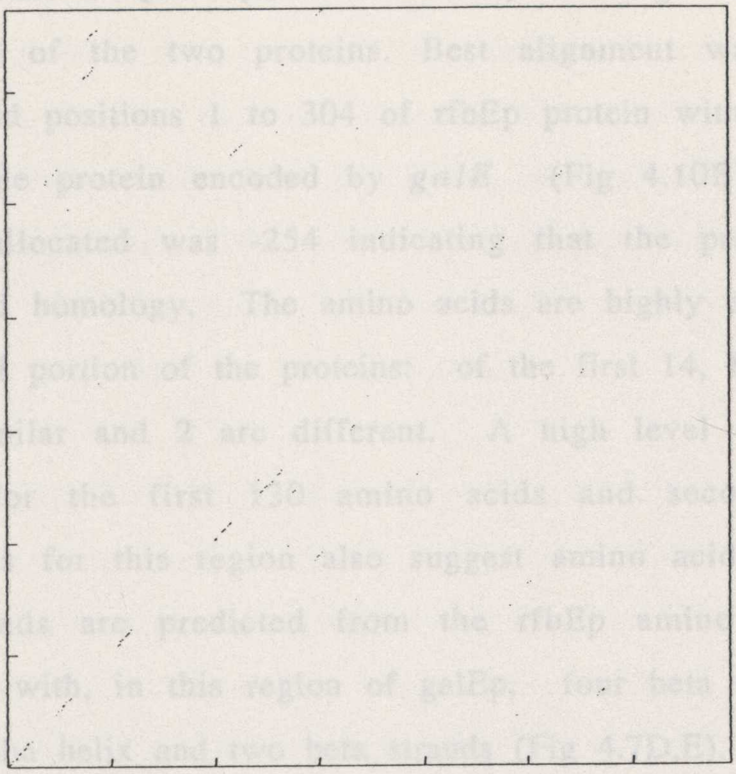
FIGURE 4.10

Relationship of the amino acid sequence of proteins encoded by *rfbE* and *galE*

A) An outcome of Diagon plot is shown. The X-axis represents the amino acid sequence of the protein predicted from *rfbE* and the Y-axis represents the amino acid sequence of the protein encoded by *galE*. A span length of 11 amino acids and percent proportional score of132 were used. Diagonal rows of dots indicate the regions of homology.

B) The alignment of the amino acid sequence of proteins preidcted from *rfbE* (from positions 1 to 304) and *galE* (from positions 1 to 296) is presented here. Distance refers to the relationship between the two proteins. The numbers in parentheses denote the number of gaps added to the sequence to make the best alignment.

A



B

```
DISTANCE  -254,      1- 304 ( 7),      1- 296 (15)
        10        20        30          40        50        60        70
MKLLITGGCGFLGSNLASFALSQGIDLIVFDNL  SRKGATDNLHWLSSLGNFEFVHGDIRNKNDVTRLITKYMPDSCFH
:  : ::: :  ::      :  : : :  :::  :          :       :: :::::    :        :   :
MRVLVTGGSGYIGSHTCVQLLQNGHDVIILDNLCNSKRSVLPVIERLGG KHPTFVEGDIRNEALMTEILHDHAIDTVIH
        10        20        30        40         50        60        70
80        90        100       110       120       130       140       150
LAGQVAMTTSIDNPCMDF EINVGGTLNLLEAVRQYNSNCNIIYSSTNKVYGDLEQYKYNETETRYTCVDKPNGYDESTQ
 ::  :   :   :       :  ::: :  : :  :   : : ::   ::::     :   :           :  :
FAGLKAVGESVQKP LEYYDNNDNGTLRLISAMRAANVK NFIFSSSATVYGDQPKIPY  VES FPTGTPQSPYGKSKL
         90        100       110        120       130          140       150
160       170       180       190       200       210       220          230
LDFHSPYGCSKGAADQYMLDYARIFGLNTVVFRHSSMYGGRQFATYDQGWVGWFCQKAVEIKNG  I NKPFTISGNGKQ
          :   :       : :           : :                : ::       :          :
MVEQILTDLQKAQPD WSIALLRYFN PVGAHPSGDM GEDP QGIPNNLMPYIAQVAVGRRDSLAIFGNDYP TEDGTG
   160        170        180        190        200       210       220
   240       250       260        270       280       290       300
VRDVLHAEDMISLYFTALANVSKIRG NAFNIGGTIVNSLSLLELFKLLEDYCNIDMRFTNLPVRKVISVF
:::  :  :       :        :    : :    ::  :          :           :     :
VRDYIHVMDLADGHVVAMEKLANKPGVHMYNLGAGVGNSV L DVVNAFSKACGKPVNY HFAPRRE GAF
        240       250       260        270        280        290
```

the program SEQHP (Kanehisa 1982) to align the amino acid sequences of the two proteins. Best alignment was shown from amino acid positions 1 to 304 of rfbEp protein with positions 1 to 296 of the protein encoded by *galE* (Fig 4.10B). The distance number allocated was -254 indicating that the protein sequences have good homology. The amino acids are highly conserved in the N-terminal portion of the proteins; of the first 14, 8 are conserved, 4 are similar and 2 are different. A high level of homology is present for the first 130 amino acids and secondary structure predictions for this region also suggest amino acid similarity. Six beta strands are predicted from the rfbEp amino acid sequence compared with, in this region of galEp, four beta strands followed by an alpha helix and two beta strands (Fig 4.7D,E). The amino acid sequences of epimerases encoded by *rfbE* and *galE* were compared with sequences for other dehydrogenases where secondary structures have been published. The first two beta strands (βA and βB) of CDP-paratose-2-epimerase (rfbEp) and UDP-galactose-4-epimerase (galEp) show a good fit with corresponding strands of known dehydrogenases as well as with dehydrogenases encoded by *rfbJ* and *rfbS* (Table 4.1). The other beta stands predicted in the secondary structure of rfbEp and galEp could not be located in these comparisons.

4.5.6 Homology with other proteins

GENBANK Nucleic Acids Release 45.0 22 Sept 1986, NBRF (National Biomedical Research Foundation) Nucleic Acids Release 29, 18 Sept 1986 and EMBL (European Molecular Biology Laboratory) Nucleic Acid database release 9, Sept 1986 were searched for homology with the nucleotide sequences of *rfbS* and *rfbE* . Three different protein databases NBRF (National Biomedical Research Foundation) Protein database Release 10, 13 Aug 1986, NEWAT Protein database April 1985, and Kyoto University (PROSEQ) Protein database September 1986 were searched for homologies between the amino acid sequences of the predicted proteins and known amino acid sequences for other proteins.

We used the MBIS suite of programs (Bucholtz and Reisner 1986) to search these databases using quasi correlation coefficient (Qr) (Lipman and Pearson 1985) and a program written by Kanehisa *et al* (1984) based on the algorithm developed by Wilber and Lipman (1983). A third method FASTN and FASTP was used for searching the PIR and KYOTO databases based on the algorithm and software of Lipman and Pearson (1985). However, none of the protein and nucleotide sequences on these databases were found to have significant homology with our protein and nucleic acid sequences.

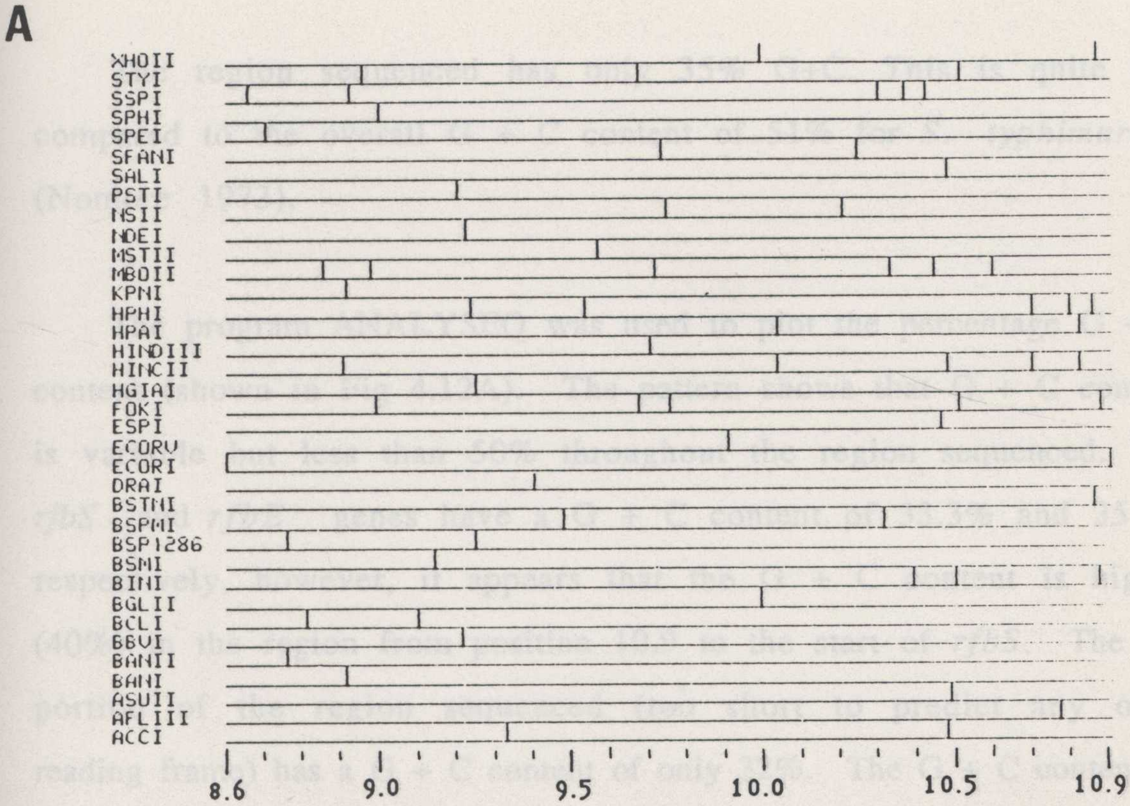
4.5.7 Search for restriction endonuclease sites

Restriction endonuclease sites were located (Fig 4.11A) using the program ANALYSEQ (Staden 1984). Restriction endonucleases whose sequences were not present are also listed in Fig 4.11B. All the restriction sites mapped earlier to produce a map of this region (Chapter 3) match with the sites predicted by the program ANALYSEQ.

FIGURE 4.11

Search for restriction endonuclease sites in the sequenced region

Restriction enzyme sites were predicted using the program ANALYSEQ. **A)** Plot of restriction endonucleases with six base specificity that was found by ANALYSEQ is shown in this figure. **B)** Those enzymes which were not present are listed here.



B <u>NAMES WHOSE SEQUENCES DO NOT MATCH</u>

AAT II
ACY I
AFL II
AHA II
APA I
APAL I
ASU I
AVA I
AVA II
AVR II
BAL I
BAM HI
BGL I

BSPM II
BSSH II
BSTE II
BSTX I
CLA I
DRA II
DRA III
EAE I
ECA I
ECO47 III
ECOB
ECOK
FNUD2

FSP I
HAE I
HAE II
HAE III
HGA I
HGID I
HHA I
HINP I
HINP II
MLU I
MST I
NAE I
NAR I

NC II
NCO I
NHE I
NOT I
NRU I
PAL I
PSS I
PVU I
PVU II
RSP I
RSR II
SAC I
SAC II

SAU96 I
SCA I
SFI I
SMA I
SNAB I
SPL I
SST I
SST II
STU I
THI I
XBA I
XHO I
XMN I

4.5.8 G + C content

The region sequenced has only 35% G+C. This is quite low compared to the overall G + C content of 51% for *S. typhimurium* (Nomore 1973).

The program ANALYSEQ was used to plot the percentage G + C content (shown in Fig 4.12A). The pattern shows that G + C content is variable but less than 50% throughout the region sequenced. The *rfbS* and *rfbE* genes have a G + C content of 33.3% and 35.5% respectively, however, it appears that the G + C content is higher (40%) in the region from position 10.9 to the start of *rfbS*. The tail portion of the region sequenced (too short to predict any open reading frame) has a G + C content of only 32%. The G + C content of intergenic spaces usually diverges more from 50% than that of coding sequences (Muto and Osawa 1987). The tail portion may then be an intergenic space. Local deviation in base composition, dinucleotide composition, trinucleotide composition and negentropy were plotted (Fig 4.12B,C,D,E) using the program ANALYSEQ, to check for patterns emerging in G+C content of the region sequenced. None of the plots showed any significant boundary or cut off mark from which a conclusion could be drawn.

4.5.9 Codon usage and amino acid composition

Codon usage for the predicted proteins of gene *rfbS* and *rfbE* is shown in Table 4.2A,B. The codon usage in those genes of *S. typhimurium* that have been sequenced and entered in the GenBank genetic sequence database, is also shown in Table 4.2C

FIGURE 4.12

G + C content plot and associated parameters

Five different plots based on the base composition of the sequenced DNA, were plotted using ANALYSEQ. **A** shows the variation in G+C content. The line through the middle of the plot represents 50% G+C content. A span length of 21 nucleotides is used. **B** plots local deviation in base composition. Local deviation indicates unususal or similar regions in the sequence. The base composition of a defined window length is compared to the base composition of the overall sequence. The maximum value calculated is 32.102161, the minimum is 0.023309. **C** shows the plot of local deviation in dinucleotide composition, maximum and minimum values are 43.12155 and 2.53526 respectively. **D** is a plot of local deviation in trinucleotide composition. The maximum value is 141.69623, and minimum is 26.03979. **E** shows a negentropy plot. Negentropy is a measure of the variation of the composition of each base, the program compares the base composition of the whole region (T = 0.317, C = 0.144, A = 0.333, G = 0.206) with the base composition of the span in question. The program assigns a value based on the similarity, to the whole sequence composition. The given maximum and minimum values are 0.005684 and 0.002268. A span length of 121 nucleotides is used in plots B, C, D and E.

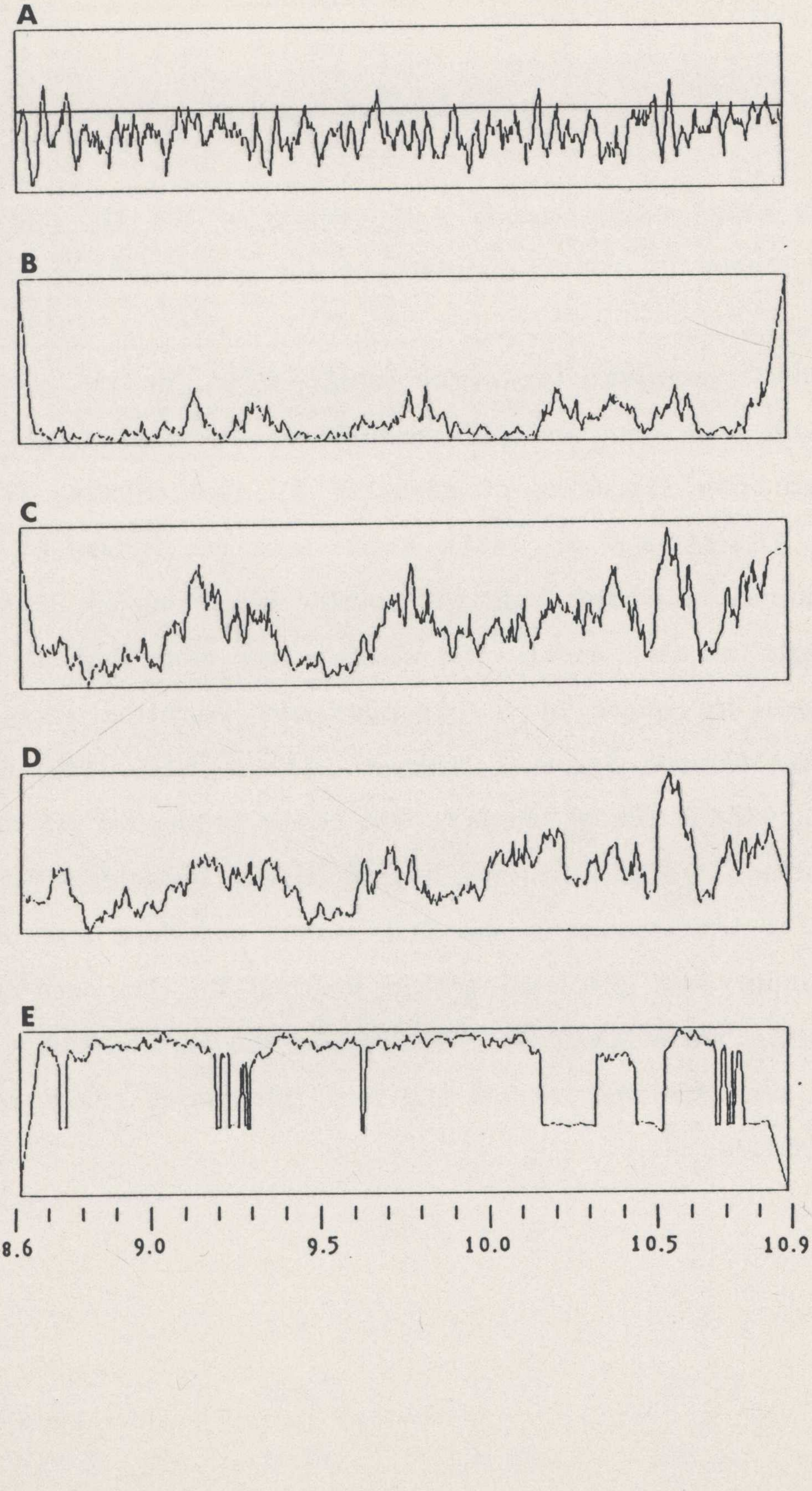

A
B
C
D
E
8.6
9.0
9.5
10.0
10.5
10.9

TABLE 4.2

Codon usage tables and G + C content of the 1st, 2nd and 3rd bases in coding triplets of the predicted proteins

A and **B** represent the codon usage table for the proteins predicted from *rfbS* and *rfbE* , respectively. A codon usage table derived from sequences of genes of *S. typhimurium* is also shown (C) (Aota *et al* 1988). Amino acids are denoted by their international one-letter code and number of codons are listed to the right of each codon. The codon usage table derived from sequences of genes of *S. typhimurium* tabulated from the GenBank Genetic Sequence Database (Release 50.0, 1987) (Aota *et al* 1988) is also shown here. The codon frequencies are shown as codon usage per thousand codons in this instance only. **D** shows G + C content in the first, second and third position of codons for each predicted protein encoded by *rfbS* and *rfbE*. The values are indicated as percentages of the base content. 1, 2 and 3 represent first, second and third position of triplet codon respectively.

A

```
----------------------------------------------
F TTT    6. S TCT    4. Y TAT    8. C TGT    4.
F TTC    8. S TCC    1. Y TAC    2. C TGC    2.
L TTA    8. S TCA    5. * TAA    0. * TGA    0.
L TTG    4. S TCG    0. * TAG    0. W TGG    2.
----------------------------------------------
L CTT    6. P CCT    2. H CAT    4. R CGT    2.
L CTC    1. P CCC    1. H CAC    0. R CGC    0.
L CTA    4. P CCA    3. Q CAA    2. R CGA    2.
L CTG    1. P CCG    1. Q CAG    2. R CGG    0.
----------------------------------------------
I ATT   16. T ACT    5. N AAT   16. S AGT    8.
I ATC    4. T ACC    2. N AAC    4. S AGC    2.
I ATA   14. T ACA    7. K AAA   15. R AGA    6.
M ATG    5. T ACG    2. K AAG    3. R AGG    2.
----------------------------------------------
V GTT    6. A GCT    5. D GAT   10. G GGT    3.
V GTC    1. A GCC    3. D GAC    1. G GGC    2.
V GTA    5. A GCA    7. E GAA   20. G GGA    8.
V GTG    2. A GCG    3. E GAG    5. G GGG    2.
----------------------------------------------
```

B

```
==============================================
F TTT   10. S TCT    4. Y TAT   12. C TGT    8.
F TTC    8. S TCC    4. Y TAC    4. C TGC    1.
L TTA   15. S TCA   11. * TAA    0. * TGA    0.
L TTG    7. S TCG    1. * TAG    0. W TGG    3.
==============================================
L CTT    9. P CCT    4. H CAT    6. R CGT    1.
L CTC    0. P CCC    1. H CAC    1. R CGC    2.
L CTA    5. P CCA    1. Q CAA   10. R CGA    1.
L CTG    0. P CCG    0. Q CAG    3. R CGG    2.
==============================================
I ATT   16. T ACT   10. N AAT   21. S AGT    2.
I ATC    1. T ACC    2. N AAC    5. S AGC    5.
I ATA    4. T ACA    7. K AAA   14. R AGA    4.
M ATG   11. T ACG    0. K AAG    6. R AGG    6.
==============================================
V GTT    6. A GCT    5. D GAT   18. G GGT   15.
V GTC    1. A GCC    4. D GAC    2. G GGC    3.
V GTA    9. A GCA    7. E GAA    8. G GGA    6.
V GTG    6. A GCG    1. E GAG    4. G GGG    5.
==============================================
```

C

```
==============================================
F TTT 22.0 S TCT  9.2 Y TAT 16.8 C TGT  4.9
F TTC 14.8 S TCC 12.1 Y TAC 13.6 C TGC  6.3
L TTA 11.6 S TCA  6.0 * TAA  2.3 * TGA  1.0
L TTG 11.6 S TCG  7.9 * TAG  0.4 W TGG 12.4
==============================================
L CTT 10.2 P CCT  6.7 H CAT 13.8 R CGT 18.9
L CTC  8.2 P CCC  5.5 H CAC 11.9 R CGC 21.9
L CTA  3.6 P CCA  5.7 Q CAA 12.6 R CGA  3.9
L CTG 51.9 P CCG 24.4 Q CAG 32.9 R CGG  6.7
==============================================
I ATT 30.2 T ACT  8.7 N AAT 17.2 S AGT  5.9
I ATC 25.4 T ACC 27.5 N AAC 20.7 S AGC 15.1
I ATA  4.4 T ACA  5.3 K AAA 35.6 R AGA  1.4
M ATG 23.9 T ACG 17.6 K AAG  9.8 R AGG  1.1
==============================================
V GTT 11.9 A GCT 15.8 D GAT 34.8 G GGT 21.0
V GTC 19.6 A GCC 29.8 D GAC 20.1 G GGC 39.9
V GTA 10.4 A GCA 13.1 E GAA 32.8 G GGA  7.8
V GTG 25.8 A GCG 41.6 E GAG 22.1 G GGG 11.9
==============================================
```

D

	G+C		
	1	2	3
rfbS	40.65	35.90	28.19
rfbE	41.96	36.21	28.44

TABLE 4.3

Amino acid content of the proteins predicted from open reading frames *rfbS* and *rfbE*

A and B represent amino acid content table for proteins encoded by *rfbS* and *rfbE*, respectively. Amino acids are denoted by their international one-letter code. A measure of the variation from the average amino acid composition of proteins found in the EMBL sequence library is given by the O - E% values.

A

```
          A     C     D     E     F     G     H     I     K     L
         18.    6.   11.   25.   14.   15.    4.   34.   18.   24.
O-E %   -25.  -26.  -28.   49.   39.  -36.  -28.  171.   -2.   16.

          M     N     P     Q     R     S     T     V     W     Y
          5.   20.    7.    4.   12.   20.   16.   14.    2.   10.
O-E %     5.   67.  -52.  -63.  -12.    2.   -6.  -24.  -45.    5.
```

B

```
          A     C     D     E     F     G     H     I     K     L
         17.    9.   20.   12.   18.   29.    7.   21.   20.   36.
O-E %   -43.  -11.    4.  -43.   44.   -1.    1.   34.  -13.   40.

          M     N     P     Q     R     S     T     V     W     Y
         11.   26.    6.   13.   16.   27.   19.   22.    3.   16.
O-E %    86.   74.  -67.   -4.   -6.   11.  -10.   -4   -34.   35.
```

(Aota *et al* 1988). Codons with low G+C content are present in greater numbers than normally found in *S. typhimurium* due to low G + C content of the coding DNA of the *rfbS* and *rfbE* genes. Codons with T or A in the third position are strongly favoured; for example, for protein encoded by *rfbS* , of the 34 Isoleucine codons, 30 are ATT or ATA and of the 25 glutamate codons, 20 are GAA. A similar situation is observed in *rfbE* ; 95% of the Isoleucine codons are ATT or ATA and 90% of aspartate codons are GAT. The third position preference for A or T is manifested by a higher A + T content in the third codon position than in the first or second (Table 4.2D). A similar trend for the G+C contents of the first, second and third bases of the coding triplets was also observed by Muto and Osawa (1987).

The observed amino acid composition for each of the proteins predicted was compared to the average amino acid composition of all coding sequences in the EMBL sequence library (Staden 1984) using the program ANALYSEQ. Analysis of the amino acids present in the predicted open reading frames is shown in the Table 4.3. Four amino acids (alanine, glycine, proline and arginine) which have codons exclusively with G or C occur in lower amounts in these two proteins, while amino acids (such as isoleucine, asparagine and phenylalanine) encoded by triplets where all three bases are either A or T are present in greater contents. This is clear indication of a shift favouring amino acids encoded by A + T rich triplets.

4.5.10 Determination of nucleotide sequence in the vicinity of position 5.5 (*S. typhi* Ty21a unit) to localise the boundary between DNA homologous and nonhomologous to that of *S. typhimurium* DNA

The region of homology between the three strains studied extends from position 0 to 5.8 as shown by heteroduplex and restriction enzyme analysis (Chapter 3). We sequenced the region in the vicinity of position 5.5 in *S. typhi* Ty21a to localise more precisely the junction between homologous and nonhomologous DNA. The nucleotide sequence throughout this region of *S. typhimurium* LT2 was also being studied in our laboratory at that time (P. Wyk, Ph.D. thesis, 1988).

Plasmid pPR616 is pUC19 carrying the KpnI fragment of *rfb* DNA from *S. typhi* Ty21a (position 2.45 to 8.95). A SacI site in the polylinker was used to clone a 4.53 kb EcoRI - SacI fragment from pPR616 into M13mp18 for sequencing; this M13 clone was designated as pPR924. We were able to utilise a few of the oligonucleotides (listed in Table 4.4) used in sequencing the same region of *S. typhimurium* LT2. The nucleotide sequence of a 1.284 kb region around the 5.5 map position was deduced, in one direction only, using these oligonucleotides (Fig 4.13). Attempts were not made to sequence in the other direction since this sequence was only prepared to look for homology with the corresponding already known sequence of *S. typhimurium*. SEQA, used to align both nucleotide sequences, showed homology extending all the way through from position 4.55 to about 5.71 but there seemed to be a

TABLE 4.4

Synthetic oligonucleotides used in determining the sequence in the vicinity of position 5.5 (*S. typhi* Ty21a unit)

The oligonucleotides used as primers are listed in this table. 'rcm' denotes the reverse compliment of the sequence. The position of each oligonucleotide on the sequence of *rfb* DNA of *S. typhimurium* (P. Wyk Ph.D. thesis, 1988) and on the DNA sequence shown in Fig 4.13 (in parentheses) is shown.

Oligonucleotide	Orientation	Position
TAA TGG TCT AAG TCG CGA	rcm	4158-4175 (410-427)
TAC ATT ATA AGT TTT AGC	rcm	4425-4442 (677-694)
CGC GTA GGC TTT AGA ATA	rcm	4725-4745 (679-997)
CCG CCA ACC AAT TGC CAA	rcm	4996-5013 (1248-1265)
TAA CAT CTG CGT CCA GAG	rcm	5254-5271

FIGURE 4.13

DNA sequence obtained in the vicinity of position 5.5 (*S. typhi* Ty21a unit)

DNA sequence obtained using oligonucleotides (Table 4.3) is presented. Amino acids are shown by their international three-letter code. The bases in the coding region that differ from the corresponding sequence of *S. typhimurium* are underlined. All the replacements are listed in the accompanying table (Table 4.4). An NruI site present at position 5.5 in both *S. typhimurium* (Brahmbhatt *et al* 1988) and *S. typhi* Ty21a is shown by italics (approximate base 412 in this sequence).

```
TTTTTACTTGTATAATATTTGCATTGATTATGAATGATTTAATGGTGATTTTATAATAATTACAAGCTTGCTTATATTTATTTGGATAT
        10        20        30        40        50        60        70        80        90

GATTCCTTATTCATTTTTAATTATGATCCGTGGTTCATTTTAATGAAAATGTTTCCTAATAACTATTGGTCGTTCTTGTGATAATCGCTA
       100       110       120       130       140       150       160       170       180

                                  MetIleValAsnLeuSerArgLeuGlyLysSerGlyThrGlyMetTrpGlnTyrSer
TACAAGCATGAATGGAAAAACGGAGAGCTATTCAATGATCGTTAACCTATCACGTTTAGGTAAAAGCGGTACGGGGATGTGGCAATACTC
       190       200       210       220       230       240       250       260       270

 IleLysPheLeuThrAlaLeuArgGluIleAlaAspValAspAlaIleIleCysSerLysValHisAlaAspTyrPheGluLysLeuGly
GATTAAATTTTTAACGGCTCTGCGAGAAATAGCTGATGTTGACGCAATAATCTGTAGCAAGGTACACGCTGATTATTTTGAAAAGCTCGG
       280       290       300       310       320       330       340       350       360

 TyrAlaValValThrValProAsnIleValSerAsnThrSerLysThrSerArgLeuArgProLeuValTrpTyrValTyrSerTyrTrp
TTATGCAGTAGTTACTGTTCCGAATATTGTTAGCAACACATCAAAAACATCGCGACTTAGACCATTAGTATGGTATGTATATAGTTACTG
       370       380       390       400       410       420       430       440       450

 LeuAlaLeuArgValLeuIleLysPheGlyAsnLysLysLeuValCysThrThrHisHisThrIleProLeuLeuArgAsnGlnThrIle
GCTTGCGCTGAGGGTTTTAATTAAGTTTGGTAATAAAAAATTGGTGTGTACTACGCATCACACTATTCCCTTACTGAGAAACCAAACGAT
       460       470       480       490       500       510       520       530       540

 ThrIleHisAspIleArgProPheTyrTyrProAspSerLeuIleGlnLysValTyrPheArgPheLeuLeuLysMetSerValLysArg
AACCATACATGATATAAGACCTTTTTATTATCCAGATAGTCTTATTCAGAAAGTGTATTTTCGCTTTTTATTAAAAATGTCTGTTAAGCG
       550       560       570       580       590       600       610       620       630

 CysLysHisValLeuThrValSerTyrThrValLysAspSerIleAlaLysThrTyrAsnValAspSerGluLysIleSerValIleTyr
ATGTAAGCATGTTTTAACGGTATCTTATACCGTTAAAGATAGCATTGCTAAAACTTATAATGTAGATAGTGAGAAAATATCAGTAATTTA
       640       650       660       670       680       690       700       710       720

 AsnSerValAsnLysSerAspPheIleGlnLysLysGluLysGluAsnTyrPheLeuAlaValGlyAlaSerTrpProHisLysAsnIle
TAACAGCGTTAATAAATCTGATTTTATACAAAAAAAAGAAAAAGAGAATTACTTTTTAGCTGTTGGAGCAAGTTGGCCACATAAAAATAT
       730       740       750       760       770       780       790       800       810

 HisSerPheIleLysAsnLysLysValTrpSerAspSerTyrAsnLeuIleIleValCysGlyArgThrAspTyrAlaMetSerLeuGln
TCATTCATTCATAAAAAATAAAAAAGTTTGGTCTGACTCTTATAATTTAATTATTGTATGTGGTCGTACCGACTATGCAATGTCTCTCCA
       820       830       840       850       860       870       880       890       900

 GlnMetValValAspLeuLysLeuThrAspLysValThrPheLeuHisGluValSerPheAsnGluLeuLysValLeuTyrSerLysAla
ACAAATGGTCGTTGATCTGAAACTAACAGATAAAGTGACTTTTTTACATGAAGTCTCATTTAATGAATTAAAAGTTTTATATTCTAAAGC
       910       920       930       940       950       960       970       980       990

 TyrAlaLeuValTyrProSerIleAspGluGlyPheGlyIleProProIleGluAlaMetAlaSerAsnThrProValIleValSerAsp
CTACGCGCTTGTTTATCCATCTATTGATGAGGGTTTTGGCATACCTCCTATTGAAGCGATGGCATCAAATACTCCAGTTATAGTGTCCGA
      1000      1010      1020      1030      1040      1050      1060      1070      1080

 IleProValPheHisGluValLeuThrAsnGlyAlaLeuTyrValAsnProAspAspGluLysSerTrpGlnSerAlaIleGluAsnIle
TATACCAGTATTTCATGAAGTGTTAACCAATGGTGCATTATATGTGAATCCGGATGATGAAAAAAGCTGGCAGAGTGCAATTGAAAATAT
      1090      1100      1110      1120      1130      1140      1150      1160      1170

 GlyGlnLeuProAspAlaIleSerArgPheAsnAsnTyrValAlaArgTyrAspPheAspAsnMetLysGlnMetValGlyAsnTrpLeu
AGGGCAGTTGCCTGATGCAATTTCCAGATTTAACAACTATGTCGCACGGTATGACTTTGATAATATGAAGCAGATGGTTGGCAATTGGTT
      1180      1190      1200      1210      1220      1230      1240      1250      1260

 AlaGluSerLys***
GGCGGAATCAAAATAAATGAAAA
      1270      1280
```

clear cut off mark at position 5.71 which could also be seen clearly in the Diagon plot (Fig 4.14).

One open reading frame was seen in the *S. typhi* Ty21a sequence extending from base position 215 to 1276 (Fig 4.13) and from position 4.6 to 5.7 on the *S. typhi* Ty21a map (Fig 3.4). This open reading frame corresponds to that of ORF353 in *S. typhimurium* (P. Wyk, Ph.D. thesis, 1988) but the function of the protein predicted by ORF353 is not yet known. ORF353 of *S. typhimurium* and the corresponding open reading frame in *S. typhi* Ty21a were then aligned using the program SEQA. No deletion or insertion was found in either sequence. However, there were 21 replacements; of these, 13 were at the third base position, 2 were at second base position and 6 were at first base position in the codons used (Table 4.5) showing an overall 2% difference between these corresponding regions.

4.5.11 Other structures

The entire 2.3 kb sequence was searched for the presence of promoters using ANALYSEQ (Staden 1984) which employs the algorithm of Hawley and McClure (1983). This program failed to find any promoter, however genuine promoters may be obscured by the low G+C content of the region sequenced. Physical data given in section 4.2 shows that there is possibly a promoter before the start of the *rfbS* gene. The sequence was also searched for evidence of any transposon event having occurred but none was found. This search does not entirely rule out the possibility of a rearrangement of this kind having occurred but no relevant clues were found. A G+C

FIGURE 4.14

Localising the junction of homologous DNA and nonhomologous DNA between *S. typhi* Ty21a and *S. typhimurium* LT2 in the vicinity of position 5.5

Diagon plot obtained using the program DIAGON is shown here. The X-axis represents the nucleotides sequence of the region in the vicinity of position 5.5 of *S. typhi* Ty21a. The Y-axis represents the nucleotide sequence of the corresponding region (position 3748 - 5031) of *S. typhimurium* LT2 which was adapted from P. Wyk, Ph.D. thesis (1988). A span length of 11 nucleotides and percent proportional score of 8 were used. The diagonal line of dots shows the region of homology.

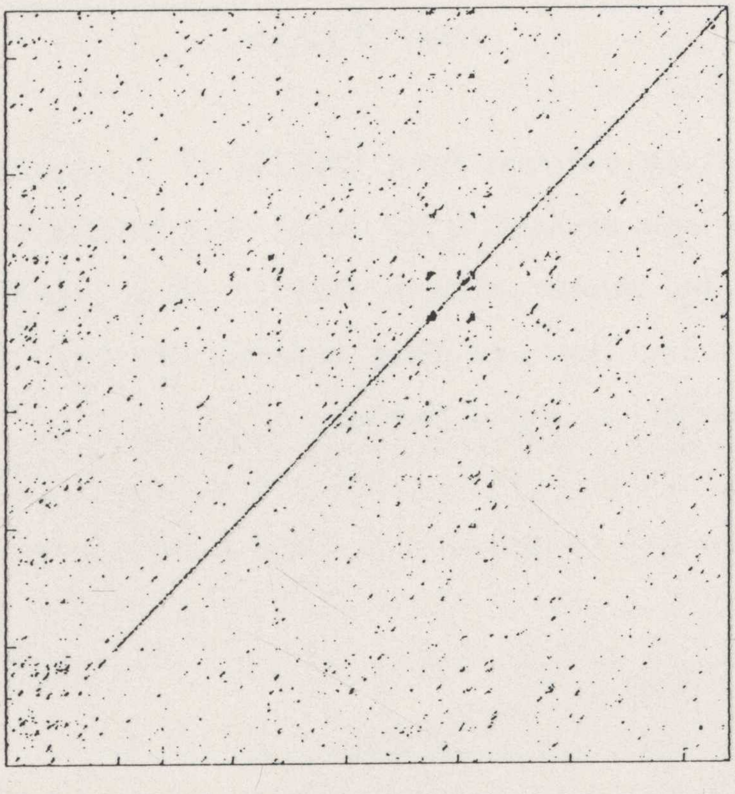


TABLE 4.5

Replacements at first, second and third base positions of codons of ORF353 equivalent coding sequence from *S. typhi* Ty21a

Codons that are different from ORF353 of *S. typhimurium* are shown with their encoded amino acid. The base in a codon that is replaced by another base in ORF353 of *S. typhimurium* is underlined. Ist, IInd and IIIrd represent the first, second and third position of the triplet codons.

	Ist		IInd		IIIrd
Ile	A T A	Thr	A C A	Ser	A G C
Val	G	Cys	A		T
Leu	C T T	Gly	G G G	Val	G T T
Phe	T	Glu	A		A
Lys	A A A			Gly	G G G
Glu	G				A
Val	G T T			Ala	G C T
Ile	A				A
Glu	G A A			Thr	A C G
Lys	A				A
Arg	A G A			Ile	A T T
	C				C
				Ser	T C T
					C
				Asn	A C C
					T
				Ser	A G C
					T
				Gly	G G A
					T
				Thr	A C C
					T
				Lys	A A A
					G
				Gly	G G C
					T

rich region of dyad symmetry lies 22 bp before the stop codon (between positions (2220-2037) of the *rfbE* coding sequence (Fig 4.4). It has been reported that a region of dyad symmetry plays a role in transcription termination (Rosenberg and Court 1979). We could not conclude that the transcription would terminate here, since mRNA for this region has not yet been analysed.

4.6 Summary and discussion

DNA responsible for specificity in the biosynthesis of paratose and tyvelose has been localised. The 1.4 kb BglII - EcoRI fragment extending from positions 10.0 to 8.6 on the *S. typhi* Ty21a map encodes enough information to make a *S. paratyphi A* strain produce tyvelose. Similarly, the1.2 kb EcoRI - HindIII fragment spanning positions 10.9 to 9.7 can make a *S. typhimurium* strain produce paratose. Presumably transfer and polymerization steps are carried out by the genes present in the host strain and do not discriminate between these sugars. The 2.3 kb EcoRI fragment of *S. paratyphi A* IMVS1316 shows complete restriction site homology with the corresponding fragment of *S. typhi* Ty21a but seems to have a functional gene only for the 0-2 epitope specificity (*rfbS*) bounded by the HindIII (15.3) and EcoRI (16.5) sites. This suggests that the gene for the 0-9 epitope specificity (*rfbE*) although present must be inactive in *S. paratyphi A* IMVS1316.

DNA sequence for the *rfbS* and *rfbE* region was determined using the dideoxy sequencing method (Sanger *et al* 1977, 1980). The sequenced region has two open reading frames. One open reading frame of 837 bp (*rfbS*) is equivalent to the *rfbJ* gene of *S*.

typhimurium (P. Wyk, Ph.D. thesis, 1988); the second open reading frame of 1044 bp (*rfbE*) has a protein product with an epimerase function. The presence of a ribosome binding site for *rfbS* could not be predicted with confidence, but the long open reading frame found in this region by Shepherds RNY preference method confirms that a functional gene is likely to exist.

The protein predicted from the coding region *rfbS* has a mol. wt. of 31501 daltons comprising 279 amino acids. It is a highly charged molecule and the amino acid sequence of rfbSp shares substantial homology with the amino acid sequence of rfbJp in *S. typhimurium* (Fig 4.8). A small region (14 amino acids) at the N-terminus and a longer portion nearer the C-terminus from position 169 to 250 of rfbSp and from 189 to 272 of rfbJp, show good homology. The function of the protein encoded by *rfbS* is similar to the protein encoded by *rfbJ* of *S. typhimurium*. Both these enzymes have a dehydrogenase function and use NADPH as a coenzyme. The first three beta strands (βA, βB and βC) of rfbSp show a very good fit with the protein encoded by *rfbJ* of *S. typhimurium* and with the NADPH binding site of a group of dehydrogenases whose secondary structures are known. This suggests that the region at the N-terminus may have a role in forming a NADPH binding domain. The C-terminus portions of rfbSp (169-279) and rfbJp (189-299), which show homology in their amino acid sequences, may be involved in binding and modifying the substrate since both these enzymes use CDP-4-keto-3,6-dideoxy glucose as a substrate for the reaction. Nucleotide sequences of *rfbJ* and *rfbS* show more variation than the amino acid sequences. The nucleotide sequence of *rfbS* (837 bp) and *rfbJ* (900 bp) genes were aligned and computer analysis found

482 matches, 341 replacements and 79 deletion/insertions; of the deletions/insertions, 17 were in *rfbJ* and 62 were in *rfbS*. The high level of deletions/insertions made precise alignment impossible. This variaton in nucleotide sequence indicates that the polymorphism has been present for a very long time and makes conclusions on the evolutionary relationship between these two genes of almost similar function very difficult.

The protein predicted from the open reading frame *rfbE* has a mol. wt. of 39062 daltons comprising 348 amino acids. It has an epimerase function converting CDP-3.6-dideoxy-D-glucose (paratose) to CDP-3,6-dideoxy-D-mannose (tyvelose) by acting on carbon 2. This protein shows reasonable homology at the N-terminus region with the amino acid sequence of a known epimerase (*galE*) of *E. coli* which acts on carbon 4. The enzyme, galactose epimerase, uses NAD$^+$ as a coenzyme. The first 130 amino acids of paratose epimerase (rfbEp) show some homology with galactose epimerase (galEp), suggesting that this region may be involved in forming a NAD$^+$ binding domain. The predicted secondary structure of both proteins also indicates homology in this region.

One striking feature of this sequence is the low G+C content; only 35% of all bases. There may be a cut off mark at position 10.7 to right of the EcoRI (10.9) where G+C content rises to about 40%. Low G+C content has also been found in the corresponding *rfb* region of *S. typhimurium* LT2 (P. Wyk, Ph.D. thesis, 1988). It is quite likely that this DNA originally came from some other bacterial genome of low G+C content and was integrated into the *Salmonella* chromosome (where the G+C content is generally 51% of all bases).

This possibility was also postulated by P. Wyk (Ph.D. thesis, 1988) in his analysis of the *rfb* genes in *S. typhimurium* LT2.

As shown in chapter 3, the region of homology between *S. typhi* Ty21a, *S. paratyphi A* IMVS1316 and *S. typhimurium* LT2 in the vicinity of position 9.57, extends from position 9.57 to about 8.17 on the *S. typhimurium* map (Fig 3.14). Sequence data from left of position 10.9 of *S. typhi* Ty21a presented in this chapter show regions of homology with the corresponding sequence of *S. typhimurium* ; a 330 bp sequence present before the start of *rfbS* was conserved while the *rfbS* showed only about 55% DNA homology with the corresponding sequence of the *rfbJ* gene of *S.typhimurium*. We also sequenced the 1.284 kb region in the vicinity of position 5.5 (*S. typhi* Ty21a unit) to localise precisely the boundary of DNA homologous with *S. typhimurium* LT2. When compared with the corresponding sequence in *S. typhimurium*, there seemed to be a clear cut off mark at position 5.71. This reaffirms the heteroduplex finding that homology extends from position 0 to 5.8 (Chapter 3). The open reading frame which lies in this sequenced region shows a 2% difference from the corresponding coding region ORF353 of *S. typhimurium* LT2 but this is not unusual; a structural gene (*trpB* of *E. coli*) was shown to have up to 4% variation in nucleotide sequence (Milkman and Crawford 1983). No deletion or insertion was found in the nucleotide sequence of these two coding regions when compared.

CHAPTER 5

Analysis of nonfunctional epimerase gene (*rfbE*) and triplicated region of *S. paratyphi A* IMVS1316

5.1 Introduction

The genes conferring specificity for paratose (*rfbS*) and tyvelose (*rfbE*) in *S. typhi* Ty21a have been identified and sequenced (Chapter 4). Both genes, *rfbS* and *rfbE* , lie on a 2.3 kb fragment bounded by Eco<u>RI</u> sites on either side (at positions 8.6 and 10.9 on *S. typhi* Ty21a map). This fragment exhibits restriction site homology with the corresponding EcoRI fragment in *S. paratyphi A* IMVS1316 extending from positions 14.2 to 16.5 (Chapter 3). It has been demonstrated that only the gene for O-2 specific epitope (*rfbS*), is functional on the 2.3 kb EcoRI fragment of *S. paratyphi A* IMVS1316 (pPR617), although it contains essentially all the DNA in the corresponding fragment of *S. typhi* Ty21a. This suggests that the gene for tyvelose specificity (*rfbE*) is present but nonfunctional. We sequenced the *rfbE* gene from *S. paratyphi A* IMVS1316 to look for possible mutational changes in the coding region.

Restriction maps of the *rfb* region of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 revealed that the region in *S. paratyphi A* IMVS1316 resembled that of *S. typhi* Ty21a but had triplication of an approximate 2.8 kb region (Fig 3.4). We analysed DNA in the vicinity of both ends of the triplicated segment at the sequence level, in order to locate the ends more precisely and to look for repetitive structures or evidence of structures which may have been involved in the triplication.

5.2 Nucleotide sequence of *rfbE* gene from *S. paratyphi A* IMVS1316

The start codon of *rfbE* in *S. typhi* Ty21a lies between the HindIII (9.70) and BglII (10.0) sites. It was assumed that the gene *rfbE* would start at the same position in the corresponding 2.3 kb EcoRI fragment from *S. paratyphi A* IMVS1316. An EcoRI (14.20) - BglII (15.6) fragment (which would contain the whole *rfbE* gene) was subcloned from pPR617 (Table 2.2) into the EcoRI-BamHI sites of M13mp18; designated as pPR816. Two oligonucleotides were synthesized and used as primers in the dideoxy sequencing method (Sanger *et al* 1977) to complete the sequence in one direction . The sequence and position of these primers are listed in Table 5.1. Clones used to sequence in reverse orientation are shown in Fig 5.1 and the nucleotide sequence obtained is presented in Fig 5.2.

5.3 Comparison of the *rfbE* genes from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316

The alignment of coding regions of the *rfbE* genes in *S. typhi* Ty21a and in *S. paratyphi A* IMVS1316 is shown in Fig 5.3. The coding regions of both genes were identical to each other except for one base replacement at 1104, and one base deletion at base 8 at the start of the gene sequence in *S. paratyphi A* IMVS1316; this deletion caused a translation frame shift. The frame for *rfbE* of *S. typhi* Ty21a begins with the start codon, ATG, at position 1198 (Fig

FIGURE 5.1

Clones used to sequence the *rfb E* gene from *S. paratyphi A* IMVS1316

A restriction endonuclease map of the 2.3 kb EcoRI fragment extending from positions 14.2 to 16.5 of *S. paratyphi A* IMVS1316 is shown. Lines represent M13 clones used and arrowheads denote the direction in which sequence was obtained. The position of oligonucleotides is shown by broken lines.

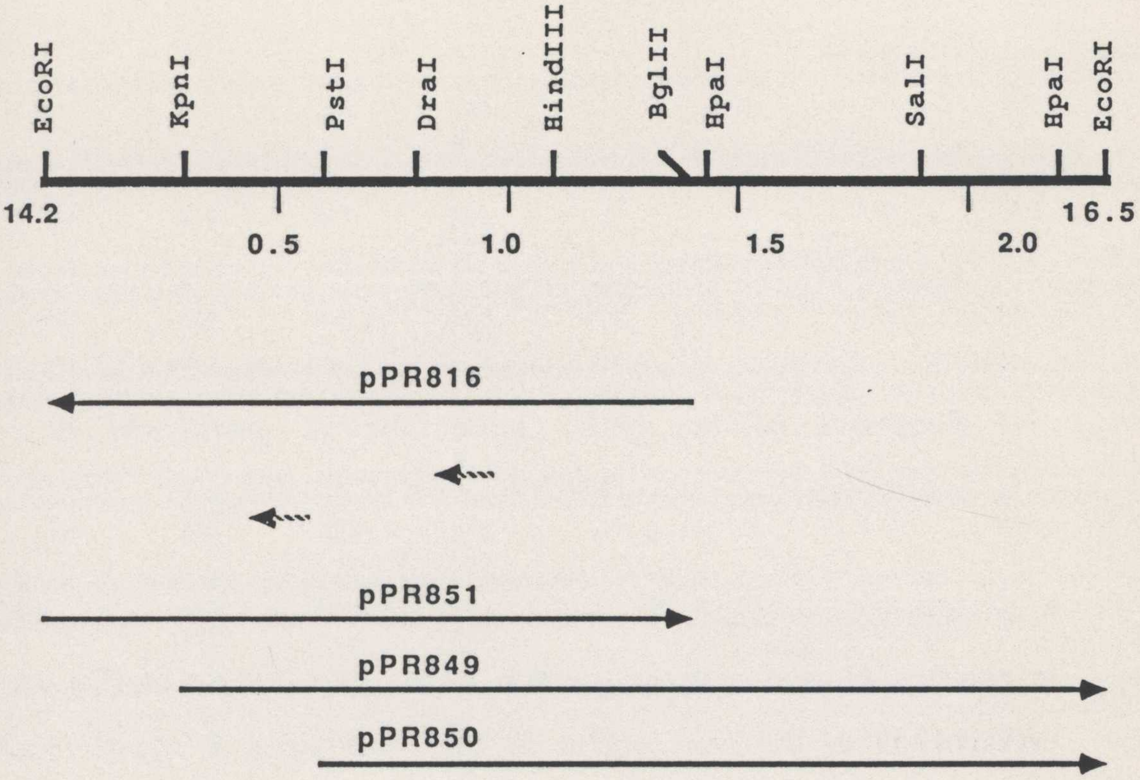


Table 5.1

Oligonucleotide	Position
CAC GTA AAG GTG CAA CAG	164-181 (Fig. 5.2)
AGG TGC TGC AGA CCA ATA	567-584 (Fig. 5.2)
GTT GGG AGT GTG GTC TAT	410-417 (Fig. 5.6)

FIGURE 5.2

Sequence of the *rfbE* gene from *S. paratyphi A* IMVS1316

The sequence obtained is presented here. The *rfbE* gene starts at position 65 and ends at position 1110. The codon (ATG) which corresponds to the start codon of the *rfbE* from *S. typhi* Ty21a is used as the first codon in the frame translated. The amino acids are denoted by their international three-letter code, stops are indicated by three asterisks.

```
                                                               MetLysLeu***LeuProValAspVal
TAGGATGGAAAAGAGAGTTCTCTCTTGTTGATGCATTAACTGAAATAATTGAAGAGGAAGAGAAATGAAGCTTTAATTACCGGTGGATGT
        10        20        30        40        50        60        70        80        90

AlaSerLeuGlyValIleLeuProProLeuLeu***ValLysGlyLeuIle***LeuTyrSerIleIleTyrHisValLysValGlnGln
GGCTTCCTTGGGAGTAATCTTGCCTCCTTTGCTTTAAGTCAAGGGATTGATTTAATTGTATTCGATAATCTATCACGTAAAGGTGCAACA
       100       110       120       130       140       150       160       170       180

IleIleTyrIleGlyTyrProPro***GluThrLeuSerLeuTyrMetValIlePheAlaThrLysMetMetLeuGlnAsp******Leu
GATAATTTACATTGGTTATCCTCCTTAGGAAACTTTGAGTTTGTACATGGTGATATTCGCAACAAAAATGATGTTACAAGATTAATAACT
       190       200       210       220       230       240       250       260       270

SerIleCysLeuIleAlaValPheIleLeuGlnValLysTrpGln***LeuHisLeuLeuThrIleLeuValTrpIleLeuLysLeuMet
AAGTATATGCCTGATAGCTGTTTTCATCTTGCAGGTCAAGTGGCAATGACTACATCTATTGACAATCCTTGTATGGATTTTGAAATTAAT
       280       290       300       310       320       330       340       350       360

***ValGluLeu***IleTyrLeuArgGlnTyrGlySerIleIleGlnIleValIle***PheIleHisGlnGlnIleLysTyrThrAla
GTAGGTGGAACTTTAAATTTACTTGAGGCAGTACGGCAGTATAATTCAAATTGTAATATAATTTATTCATCAACAAATAAAGTATACGGC
       370       380       390       400       410       420       430       440       450

IleLeuSerAsnIleAsnThrMetLysGlnLysLeuAspThrLeuVal***IleSerLeuMetAspMetMetArgAlaHisAsn***Ile
GATCTTGAGCAATATAAATACAATGAAACAGAAACTAGATACACTTGTGTAGATAAGCCTAATGGATATGATGAGAGCACACAATTAGAT
       460       470       480       490       500       510       520       530       540

SerThrHisHisMetValValGlnLysValLeuGlnThrAsnThrCysLeuIleMetGlnGlyPheLeuVal***IleGlnTrpCysSer
TTCCACTCACCATATGGTTGTTCAAAAGGTGCTGCAGACCAATACATGCTTGATTATGCAAGGATTTTTGGTTTGAATACAGTGGTGTTC
       550       560       570       580       590       600       610       620       630

GlyIleHisGlnCysMetValGlyAspSerLeuLeuLeuMetIleLysAlaGly***ValGlyPheValLysLysArgLeuLysLeuLys
AGGCATTCATCAATGTATGGTGGGAGACAGTTTGCTACTTATGATCAAGGCTGGGTAGGTTGGTTTTGTCAAAAAGCGGTTGAAATTAAA
       640       650       660       670       680       690       700       710       720

MetValLeuIleAsnProSerLeuPheLeuValMetValSerLysLeuGlyMetPheCysMetLeuLysIle***PheArgTyrIleSer
AATGGTATTAATAAACCCTTCACTATTTCTGGTAATGGTAAGCAAGTTAGGGATGTTTTGCATGCTGAAGATATGATTTCGTTATATTTC
       730       740       750       760       770       780       790       800       810

LeuProTrpGlnMetTyrGlnLysLeuGlyGlyThrLeuLeuIleLeuValValProLeuSerThrAlaTyrHisTyrLeuAsnTyrSer
ACTGCCTTGGCAAATGTATCAAAAATTAGGGGGAACGCTTTTAATATTGGTGGTACCATTGTCAACAGCCTATCATTACTTGAATTATTC
       820       830       840       850       860       870       880       890       900

AsnCysLeuLysIleIleAlaThr***Ile***GlySerLeuIleTyrLeu***GlyLys***SerAlaCysPheCysCysArgTyr***
AAATTGCTTGAAGATTATTGCAACATAGATATGAGGTTCACTAATTTACCTGTAAGGAAAGTGATCAGCGTGTTTTTGTTGCAGATATTA
       910       920       930       940       950       960       970       980       990

LysAsnHis***CysAsn***LeuGluProGluSerLeuGlyLysArgTrpCysProGluAsnVal***LeuAsp***PheTyrMetMet
AAAAAATCACTAATGCAATTGACTGGAGCCCGAAAGTCTCGGCAAAAGATGGTGTCCAGAAAATGTATGATTGGACTAGTTCTATATGAT
      1000      1010      1020      1030      1040      1050      1060      1070      1080

PheLeuProAlaLysAlaGlyArgIleLys
GTTTCTGCCCGCGAAAGCGGGCAGAATTAAAGAATGAAATATTTTTT
      1090      1100      1110      1120
```

FIGURE 5.3

Alignment of the *rfbE* from *S. typhi* Ty21a with the *rfbE* from *S. paratyphi A* IMVS1316

The alignment of nucleotide sequence of the *rfbE* from *S. typhi* Ty21a and the *rfbE* from *S. paratyphi A* IMVS1316 is shown. The sequence of both genes is identical except for a one base replacement at position 1104 and a one base deletion in coding region of the *rfbE* from *S. paratyphi A* IMVS1316. The one base deletion at base 8 at the start of gene is shown by a dash.

```
      1210      1220      1230      1240      1250      1260      1270
ATGAAGCTTTTAATTACCGGTGGATGTGGCTTCCTTGGGAGTAATCTTGCCTCCTTTGCTTTAAGTCAAGGGATTGATTT
::::::: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ATGAAGC-TTTAATTACCGGTGGATGTGGCTTCCTTGGGAGTAATCTTGCCTCCTTTGCTTTAAGTCAAGGGATTGATTT
   70        80        90       100       110       120       130       140
      1290      1300      1310      1320      1330      1340      1350
AATTGTATTCGATAATCTATCACGTAAAGGTGCAACAGATAATTTACATTGGTTATCCTCCTTAGGAAACTTTGAGTTTG
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
AATTGTATTCGATAATCTATCACGTAAAGGTGCAACAGATAATTTACATTGGTTATCCTCCTTAGGAAACTTTGAGTTTG
  150       160       170       180       190       200       210       220
      1370      1380      1390      1400      1410      1420      1430
TACATGGTGATATTCGCAACAAAAATGATGTTACAAGATTAATAACTAAGTATATGCCTGATAGCTGTTTTCATCTTGCA
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TACATGGTGATATTCGCAACAAAAATGATGTTACAAGATTAATAACTAAGTATATGCCTGATAGCTGTTTTCATCTTGCA
  230       240       250       260       270       280       290       300
      1450      1460      1470      1480      1490      1500      1510
GGTCAAGTGGCAATGACTACATCTATTGACAATCCTTGTATGGATTTTGAAATTAATGTAGGTGGAACTTTAAATTTACT
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
GGTCAAGTGGCAATGACTACATCTATTGACAATCCTTGTATGGATTTTGAAATTAATGTAGGTGGAACTTTAAATTTACT
  310       320       330       340       350       360       370       380
      1530      1540      1550      1560      1570      1580      1590
TGAGGCAGTACGGCAGTATAATTCAAATTGTAATATAATTTATTCATCAACAAATAAAGTATACGGCGATCTTGAGCAAT
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TGAGGCAGTACGGCAGTATAATTCAAATTGTAATATAATTTATTCATCAACAAATAAAGTATACGGCGATCTTGAGCAAT
  390       400       410       420       430       440       450       460
      1610      1620      1630      1640      1650      1660      1670
ATAAATACAATGAAACAGAAACTAGATACACTTGTGTAGATAAGCCTAATGGATATGATGAGAGCACACAATTAGATTTC
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ATAAATACAATGAAACAGAAACTAGATACACTTGTGTAGATAAGCCTAATGGATATGATGAGAGCACACAATTAGATTTC
  470       480       490       500       510       520       530       540
      1690      1700      1710      1720      1730      1740      1750
CACTCACCATATGGTTGTTCAAAAGGTGCTGCAGACCAATACATGCTTGATTATGCAAGGATTTTTGGTTTGAATACAGT
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
CACTCACCATATGGTTGTTCAAAAGGTGCTGCAGACCAATACATGCTTGATTATGCAAGGATTTTTGGTTTGAATACAGT
  550       560       570       580       590       600       610       620
      1770      1780      1790      1800      1810      1820      1830
GGTGTTCAGGCATTCATCAATGTATGGTGGGAGACAGTTTGCTACTTATGATCAAGGCTGGGTAGGTTGGTTTTGTCAAA
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
GGTGTTCAGGCATTCATCAATGTATGGTGGGAGACAGTTTGCTACTTATGATCAAGGCTGGGTAGGTTGGTTTTGTCAAA
  630       640       650       660       670       680       690       700
      1850      1860      1870      1880      1890      1900      1910
AAGCGGTTGAAATTAAAAATGGTATTAATAAACCCTTCACTATTTCTGGTAATGGTAAGCAAGTTAGGGATGTTTTGCAT
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
AAGCGGTTGAAATTAAAAATGGTATTAATAAACCCTTCACTATTTCTGGTAATGGTAAGCAAGTTAGGGATGTTTTGCAT
  710       720       730       740       750       760       770       780
      1930      1940      1950      1960      1970      1980      1990
GCTGAAGATATGATTTCGTTATATTTCACTGCCTTGGCAAATGTATCAAAAATTAGGGGGAACGCTTTTAATATTGGTGG
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
GCTGAAGATATGATTTCGTTATATTTCACTGCCTTGGCAAATGTATCAAAAATTAGGGGGAACGCTTTTAATATTGGTGG
  790       800       810       820       830       840       850       860
      2010      2020      2030      2040      2050      2060      2070
TACCATTGTCAACAGCCTATCATTACTTGAATTATTCAAATTGCTTGAAGATTATTGCAACATAGATATGAGGTTCACTA
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TACCATTGTCAACAGCCTATCATTACTTGAATTATTCAAATTGCTTGAAGATTATTGCAACATAGATATGAGGTTCACTA
  870       880       890       900       910       920       930       940
      2090      2100      2110      2120      2130      2140      2150
ATTTACCTGTAAGGAAAGTGATCAGCGTGTTTTTGTTGCAGATATTAAAAAAATCACTAATGCAATTGACTGGAGCCCGA
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
ATTTACCTGTAAGGAAAGTGATCAGCGTGTTTTTGTTGCAGATATTAAAAAAATCACTAATGCAATTGACTGGAGCCCGA
  950       960       970       980       990      1000      1010      1020
      2170      2180      2190      2200      2210      2220      2230
AAGTCTCGGCAAAAGATGGTGTCCAGAAAATGTATGATTGGACTAGTTCTATATGATGTTTCTGCCCGCGAAAGCGGGCA
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
AAGTCTCGGCAAAAGATGGTGTCCAGAAAATGTATGATTGGACTAGTTCTATATGATGTTTCTGCCCGCGAAAGCGGGCA
 1030      1040      1050      1060      1070      1080      1090      1100

TAATTAA
 ::::::
GAATTAA
 1110
```

4.4). This corresponds to the ATG at position 65 (Fig 5.2) in *rfbE* of *S. paratyphi A* IMVS1316 used as start codon of the frame in translating this region. One stop codon (UAA) was present after the first three codons; and presumably that would stop further translation from the mRNA. Several other stop codons were seen in this translation frame indicating that absence of protein encoded by the gene *rfbE* from *S. paratyphi A* IMVS1316 was due to a frame shift mutation in the coding region (Fig 5.3).

5.4 Proteins produced in minicells

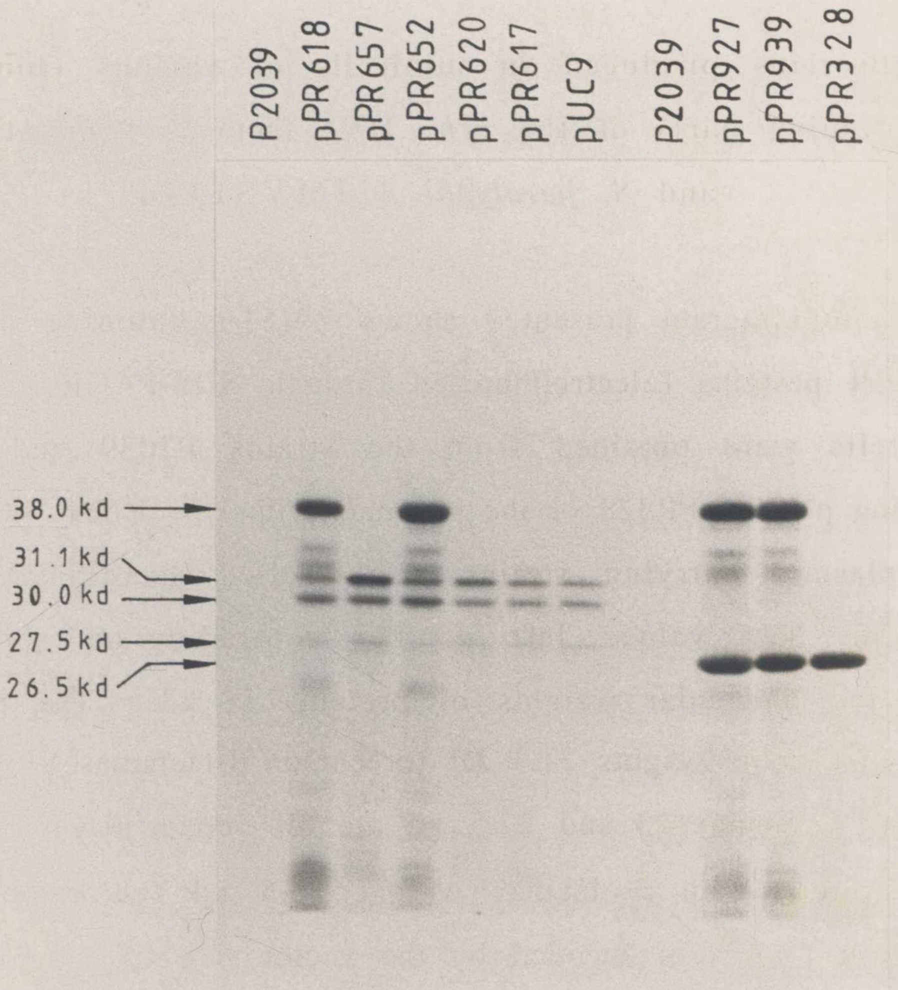
The nucleotide sequence presented in Fig 5.2 had shown a frame shift mutation in the coding region of the gene *rfbE* from *S. paratyphi A* IMVS1316. We then analysed the gene *rfbE* from *S. typhi* Ty21a as well as from *S. paratyphi A* IMVS1316, in minicells to identify any protein produced and to confirm the translation frame shift present in the *rfbE* gene of *S. paratyphi A* IMVS1316.

Appropriate plasmids were transformed into the minicell producing strain P2039. Minicells were labelled using ^{35}[S]-methionine and electrophoresed on 11% SDS polyacrylamide gel. The autoradiogram obtained is shown in Fig 5.4. The plasmid pPR618, harbouring both genes *rfbS* and *rfbE* from *S. typhi* Ty21a, produced only one protein in addition to the proteins encoded by the vector pUC9. The estimated molecular weight of that protein was 38,000 d, corresponding to the predicted molecular weight (39,062 d from sequence data) of the protein encoded by the *rfbE* of *S. typhi* Ty21a. Other faint bands present on the autoradiogram, which do not seem to correspond to vector polypeptides, may be

FIGURE 5.4

Proteins produced in minicells of various clones containing parts of the *rfb* DNA from *S. typhi* Ty21A and *S. paratyphi A* IMVS1316

The autoradiogram presented shows ^{35}[S]-methionine labelled minicell proteins (electrophoresed through SDS-PAGE gel). The minicells were obtained from the strains P2039 and P2039 carrying pUC9, pPR328 or the plasmids with *rfb* DNA. The tracks for plasmid carrying strains are labelled according to the plasmid. The values left of the autoradiogram show the estimated molecular weights of proteins in kd. The proteins with molecular weights 31.1 kd (precursor β-lactamase), 30.0 kd (mature β-lactamase) and 27.5 kd are all vector pUC9 encoded. The protein with molecular weight 26.5 kd (chloramphenicol acetyltransferase) is encoded by the vector pPR328.


P2039
pPR618
pPR657
pPR652
pPR620
pPR617
pUC9
P2039
pPR927
pPR939
pPR328
38.0 kd
31.1 kd
30.0 kd
27.5 kd
26.5 kd

breakdown products of the protein M_r 38,000 d. The M_r 31,501 d protein (predicted by sequence data) encoded by the gene *rfbS* was not seen. Plasmid pPR617, carrying the 2.3 kb *S. paratyphi A* IMVS1316 *rfb* DNA fragment from position 14.2 to 16.5, does not produce the M_r 38,000 d protein (*rfbE*); this confirms the translation frame shift.

Plasmid pPR652, a cutdown of pPR618 which does not contain *rfbS,* shows a similar protein profile to pPR618. Plasmid pPR657, with only *rfbS*, does not appear to produce a protein of M_r 31,500 d indicating either that rfbS protein produced is masked by the beta-lactamase band, which is of very similar size (M_r 31,100 d), or that it is not stable in minicells. We used another vector, pPR328 (Quigley and Reeves 1987), which has a chloramphenicol resistance gene, to check whether protein is masked by the beta-lactamase. Plasmids pPR927 and pPR939 are pPR328 with insert fragments spanning *S. typhi* Ty21a *rfb* DNA, from positions 8.6 to 10.9 and from positions 8.6 to 10.0, respectively, in correct orientation for expression. Both pPR927 and pPR939 produce a M_r 38,000 d protein (product of *rfbE*) apart from the M_r 26,500 d chloramphenicol transacetylase encoded by the vector pPR328. The protein (expected molecular weight of 31,501 d) encoded by *rfbS* could not be seen.

5.5 Proteins visualised by Coomassie Brilliant Blue staining on PAGE gel

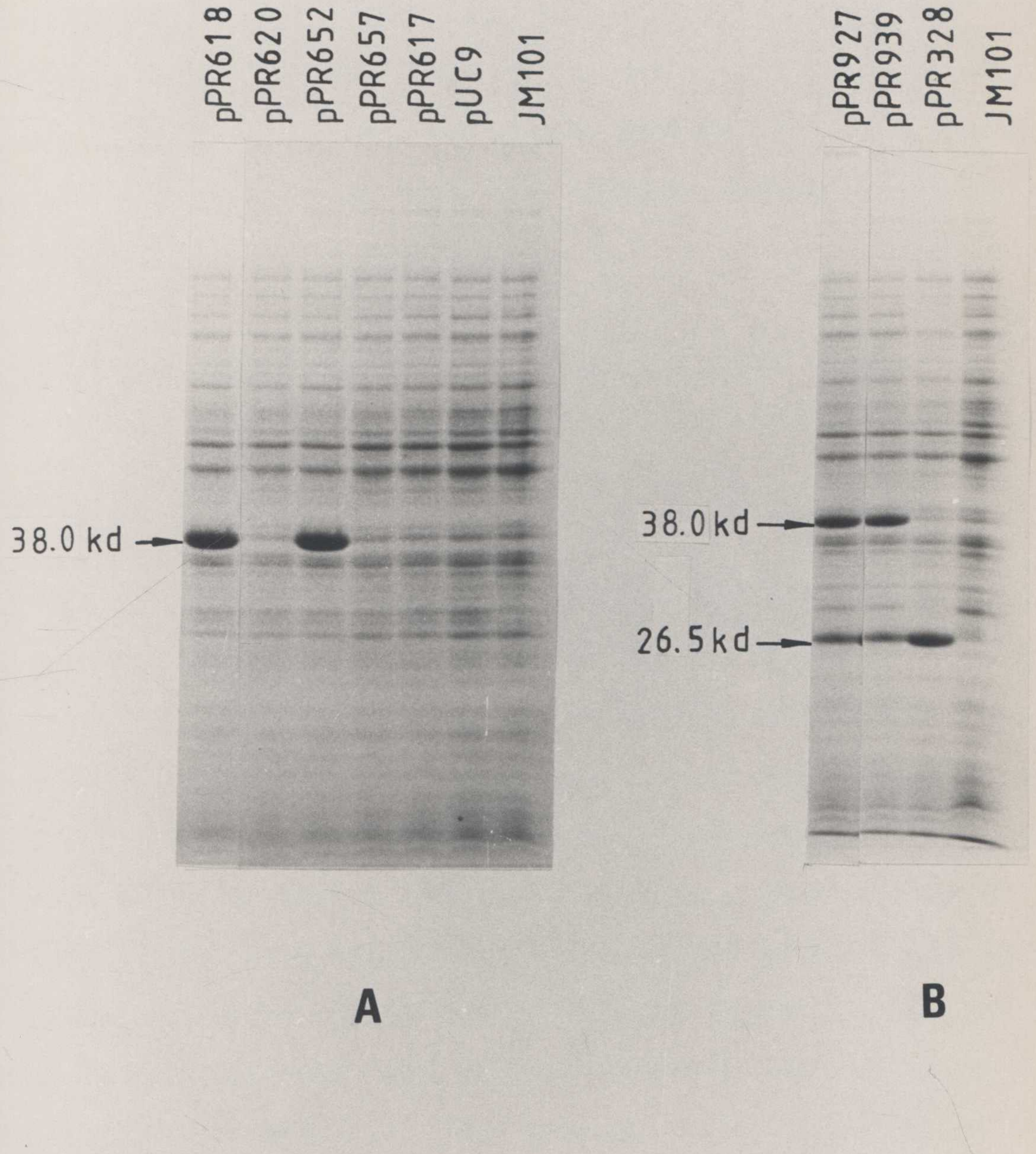
Proteins produced by the *rfb* region of *S. typhi* Ty21a (positions 8.6 to 10.9) and of *S. paratyphi A* IMVS1316 (position 14.2 to 16.5) in *E. coli* strain JM101 were visualised on PAGE gel.

FIGURE 5.5

Proteins produced in strains carrying clones of parts of the *rfb* DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316

The Coomassie Blue stained SDS-polyacrylamide gel shows protein profile obtained from electrophoresed whole cellular material of strain JM101 carrying the *rfb* clones.

The arrows left of the gel indicate the extra proteins produced. The protein with molecular weight 38.0 kd corresponds to the protein encoded by the gene *rfbE* . Other extra bands correspond to the vector encoded proteins.


pPR618
pPR620
pPR652
pPR657
pPR617
pUC9
JM101
38.0 kd
A
pPR927
pPR939
pPR328
JM101
38.0 kd
26.5 kd
B

Appropriate plasmids were transformed into JM101 and whole cell extracts from the resultant strains were electrophoresed on 11% polyacrylamide gel. The proteins visualised by coomassie Brilliant Blue staining are shown in Fig 5.5. A protein (estimated molecular weight 38,000 d) encoded by the *rfbE* gene, seen in these gels, corresponds to the protein seen in minicells (section 5.4). The expected protein of M_r 31,500 d, encoded by *rfbS* , could not be seen. *This confirms that the protein encoded by* *rfbS* is either unstable or produced in very low amounts, making its detection very difficult.

5.6 Localising ends of the triplicated DNA segment in the *rfb* region of *S. paratyphi A* IMVS1316 at base sequence level

Restriction analysis of *S. paratyphi A* IMVS1316 *rfb* DNA, presented in chapter 3, showed that a 2.8 kb region is triplicated (Fig 3.4); we attempted to locate precisely both ends of this region. One end of the triplicated region lies between positions 13.6 and 14.2 (Fig 3.4). A 1.8 kb EcoRI fragment extending from positions 12.4 to 14.2 was used to sequence this region. This fragment from pPR437 (Table 2.2), was cloned into M13mp19 with its orientation such that the primer would read through the EcoRI site at position 14.2. The sequence was further extended using a synthetic oligonucleotide (Table 5.1). The sequence (872 bp) was obtained in one direction and is presented in Fig 5.6.

Physical data presented in chapter 3 revealed that the other end of the triplicated region lies between positions 4.87 to 5.50. The

FIGURE 5.6

Nucleotide sequence present at the ends of the triplicated segment in the *rfb* region of *S. paratyphi A* IMVS1316

A) DNA sequence from position 14.2 toward the *his* locus is presented.

B) shows DNA sequence obtained from the NruI site (position 11.10) of pPR848 in order to locate the junction where one segment of the triplicated region ends and another segment starts (Fig 5.7). The ClaI restriction endonuclease site is indicated by italic letters and the overlapping sequence is underlined. The arrows indicate the ends of the triplicated region.

A

```
GAATTCCAAGACATCTTATTATTGCCGCTTCCTCTTGGCTTTCAAAGATAATAATTGCCGGTGTTCAGTTAGTAAGTGTTAAATTTCTTT
        10        20        30        40        50        60        70        80        90
TGGAAATTCTTGGCGAAGAATCATACCTGTATTTACTCTTTTAACTGGATTATTGGTCTGGTTTAGCATTGCAGATATTGGGATTGGTAG
       100       110       120       130       140       150       160       170       180
TAGTCTACAAAATTATATATCTGAGTTGAAAGCTGATAGAAAATCATATGATGCATATATCAAGGCCGCAGTTCATATTCTATTCGCATC
       190       200       210       220       230       240       250       260       270
CTTAATCACTTTTAAGCTCTACATTATTCTTCTTATCAGATAAATTATCGTCACTATATCTTACTTCATTTAGCGATGAATTGAAAAACA
       280       290       300       310       320       330       340       350       360
CTCAGGAAGTTATTTTTTTATAGCAAGTATATTATTTCATATTCACGGCGTTGGGAGTGTGGTCTATAAAATATTATTTGCGGAACTGTT
       370       380       390       400       410       420       430       440       450
AGGGTGGCAAAGCTAATATAATTAATGCATTATCTTATCTTTTAGGTTTTTAGATGTAGTTGCGATCATTATTTAATGTCAGATTGAGT
       460       470       480       490       500       510       520       530       540
ATTACTCGCTTTAGTAGCATTGTATGTCGGTAGCAATATGTCCATTATATATATATCGTTTCGGTATATATATGTTCTTAAGCGAAAGTA
       550       560       570       580        ↓        600       610       620       630
AACTTTAACACTATAAATTATTACTATCACGTTCATCAGGGTTTCTGATTTTTTCGTCCTTTATCGATAATAGTTTTACAAACTGATTAT
       640       650       660       670       680       690       700       710       720
ATTGTGATGTCTCAGAAATTATCTGCTGCAGATATTATAAAATATACTGTAACGATGAAAATATTTGGTTTAATGTTTTTTATTTATACT
       730       740       750       760       770       780       790       800       810
GCGGTATTACAAGCATTATGGCCAGTATGTGCTGAATTACGAGTGAAATGCAGTGGAGAAGC
       820       830       840       850       860       870
```

B

```
CGACTTAGACCATTAGTATGGTATGTATATAGTTACTGGCTTGCGCTGAGGGTTTTAATTAAGTTTGGTAATAAAAAATTGGTGTGTACT
        10        20        30        40        50        60        70        80        90
ACGCATCACACTATTCCCTTACTGAGAAACCAAACGATAACCATACATGATATAAGACCTTTTTATTATCCAGATAGTCTTATTCAGAAA
       100       110       120       130       140       150       160       170       180
GTGTATTTTCGCTTTTTATTAAAAATGTCTGTTAAGCGATGTAAGCATGTTTTAACGGTATCTTATACCGTTAAAGATAGCATTGCTAAA
       190       200       210       220       230       240       250       260       270
ACTTATAATGTAGATAGTGAGAAAATATCAGTAATTTATAACAGCGTTAGTAAATCTGATTTTATACAAAAAAAAGAAAAAGAGAATTAC
       280       290       300       310       320       330       340       350       360
TTTTTAGCTGTTGGAGCAAGTTGGCCACATAAAAATATTCATTCATTCATAAAAAATAAAAAAGTTTGGTCTGACTCTTATAATTTAATT
       370       380       390       400       410       420       430   ↓   440       450
ATTGTATGTGGTCGTACCGACTATGCAATGTCTCTCCAACAAATGGTCGTTGATCTGAAACTAACAGATAAATTTTCGTCCTTTATCGAT
       460       470       480       490       500       510       520       530       540
AATAGTTTTACAAACTGATTATATTGTGATGTCTCAGAAATTATCTGCTGCAGATATTATAAAATATACTGTAACGATGAAAATATTTGG
       550       560       570       580       590       600       610       620       630
TTTAATGTTTTTTATTTATACTGCGGTATTACAAGCATTATGGCCAGTATGTGCTGAATTACGAGTGAAATGCAGTGGAGAAGCTGCA
       640       650       660       670       680       690       700       710
```

corresponding region from *S. typhi* Ty21a had already been sequenced (Fig 4.13); it was assumed on the basis of restriction site homology that sequence of this region in *S. paratyphi A* IMVS1316 would be identical.

The NruI - EcoRI fragment from pPR848 (Table 2.2) was cloned into the SmaI and EcoRI sites of M13mp18 for use in sequencing to locate the junction where one segment of triplicated region ends and another segment starts. Two Bal31 deletion clones of this region, starting from NruI sites, were generated and used for sequencing by the dideoxy method of Sanger *et al* (1977). The nucleotide sequence of 718 bp, obtained in one direction, is presented in Fig 5.6. The junction where two segments of the triplicated region meet is shown in Fig 5.7.

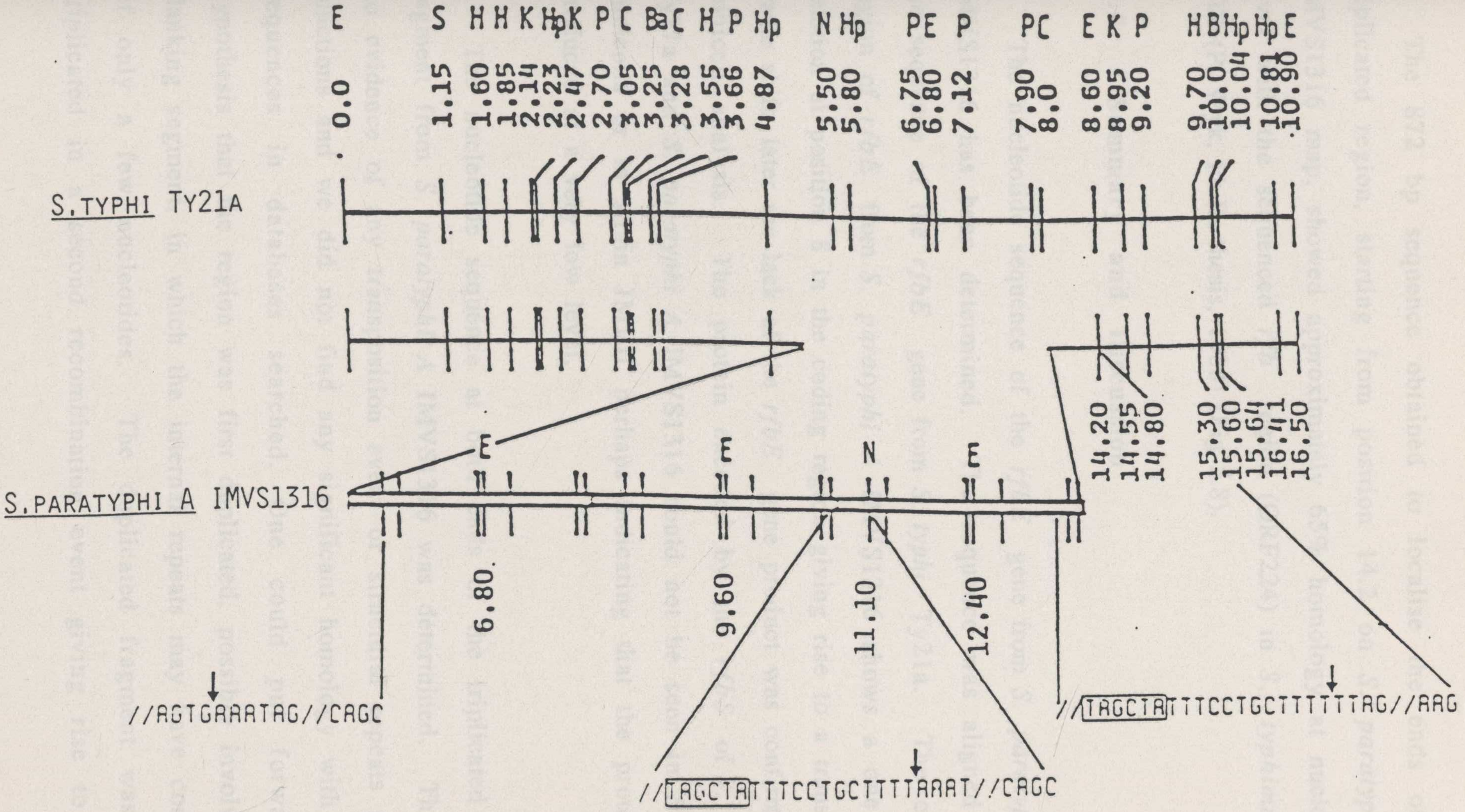
5.7 Homology with other sequences

GenBank Nucleic Acids Release 45.0 22 Sept 1986, NBRF (National Biomedical Research Foundation) Nucleic Acids Release 29, 18 Sept 1986 and EMBL (European Molecular Biology Laboratory) Nucleic Acid Database Release 9, Sept 1986 were searched for homology with the nucleotide sequence present at the junction of both ends of the triplicated region. However, none was found that had significant homology with our sequence. This search does not rule out the possibility of a transposition event, but either we did not find the relevant clues or some rearrangment has occurred and made it impossible to detect.

FIGURE 5.7

Localising ends of the triplicated region

Restriction endonuclease maps of *rfb* DNA from positions 0 to 10.9 of *S. typhi* Ty21a and from positions 0 to 16.5 of *S. paratyphi A* IMVS1316 are shown. The triplicated region is represented by double lines and arrows denote the junction point of the segment triplicated. The ClaI restriction endonuclease sites are boxed. Abbreviations for restriction endonucleases used are same as described in Fig 3.4. The junction points indicated are the same as shown in Fig 5.6, but note that the reading direction is reversed.

S.TYPHI TY21A
S.PARATYPHI A IMVS1316
E S H H K Hp K P C Ba C H P Hp N Hp P E P P C E K P H B Hp Hp E
0.0
1.15
1.60
1.85
2.14
2.23
2.47
2.70
3.05
3.25
3.28
3.55
3.66
4.87
5.50
5.80
6.75
6.80
7.12
7.90
8.0
8.60
8.95
9.20
9.70
10.0
10.04
10.81
10.90
14.20
14.55
14.80
15.30
15.60
15.64
16.41
16.50
E
E
N
E
6.80
9.60
11.10
12.40
//AGTGAAATAG//CAGC
//TAGCTATTTCCTGCTTTTAAAT//CAGC
//TAGCTATTTCCTGCTTTTTTAG//AAG

The 872 bp sequence obtained to localise the ends of the triplicated region, starting from position 14.2 on *S. paratyphi A* IMVS1316 map, showed approximately 65% homology at nucleotide level with the sequenced *rfb* region (ORF224) in *S. typhimurium* LT2 (P. Wyk, Ph.D. thesis, 1988) (Fig 5.8).

5.8 Summary and Discussion

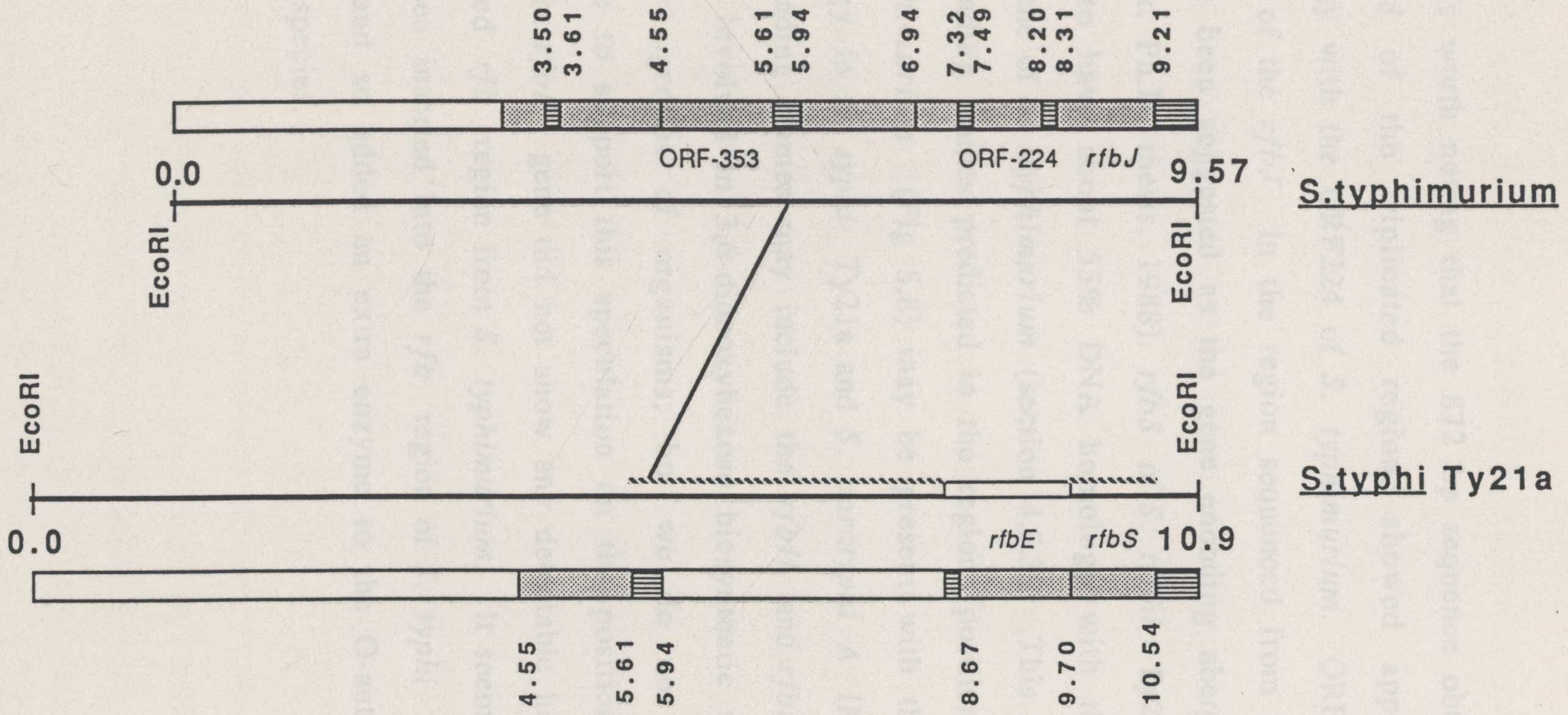
The nucleotide sequence of the *rfbE* gene from *S. paratyphi A* IMVS1316 has been determined. The sequence was aligned with the sequence of the *rfbE* gene from *S. typhi* Ty21a. The coding region of *rfbE* from *S. paratyphi A* IMVS1316 shows a one base deletion at position 8 in the coding region giving rise to a translation frame shift; later the lack of the *rfbE* gene product was confirmed in minicell analysis. The protein encoded by the *rfbS* of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 could not be seen in *E. coli* minicells or in strain JM101 perhaps indicating that the protein is produced at a very low level.

The nucleotide sequence at both ends of the triplicated DNA segment from *S. paratyphi A* IMVS1316 was determined. There is no evidence of any transposition event or structural repeats at the junctions and we did not find any significant homology with other sequences in databases searched. One could put forward a hypothesis that the region was first duplicated, possibly involving a flanking segment, in which the internal repeats may have consisted of only a few nucleotides. The duplicated fragment was then triplicated in a second recombination event giving rise to three

FIGURE 5.8

Comparative map of the *rfb* genetic structure present in *S. typhimurium* LT2 and *S. typhi* Ty21a

The map is drawn on the basis of sequencing data obtained in the *rfb* region of *S. typhi* Ty21a and *S. typhimurium* LT2. The information on the *S. typhimurium rfb* region is adapted from P. Wyk, Ph.D. thesis, 1988. Regions shown by sequencing data are indicated as follows: (▤), non-coding region; (▒), open reading frame; (☐), sequencing not done; (—), common in both; (▭), not present in *S. typhimurium* ; (⋯), scattered homology.


S.typhimurium
EcoRI
0.0
ORF-353
ORF-224
rfbJ
9.57
EcoRI
3.50
3.61
4.55
5.61
5.94
6.94
7.32
7.49
8.20
8.31
9.21
S.typhi Ty21a
EcoRI
0.0
rfbE
rfbS
10.9
EcoRI
4.55
5.61
5.94
8.67
9.70
10.54

copies of the same segment in the *S. paratyphi A* IMVS1316 *rfb* region.

It is worth noting that the 872 bp sequence obtained to localise one end of the triplicated region, showed approximately 65% homology with the ORF224 of *S. typhimurium*. ORF224 is located to the left of the *rfbJ* in the region sequenced from *S. typhimurium* and has been suggested as the gene encoding abequosyl transferase (P. Wyk, Ph.D. thesis, 1988). *rfbS* of *S. typhi* Ty21a has also been shown to have about 55% DNA homology with the corresponding *rfbJ* gene of *S. typhimurium* (section 4.5.3). This suggests that the open reading frames predicted in the region (positions 5.8 to 7.5) of *S. typhimurium* (Fig 5.8) may be present with the same level of homology in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. These open reading frames may include the *rfbH* and *rfbI* genes (section 1.3.2.4) involved in 3,6-dideoxyhexose biosynthetic steps common to all three groups of organisms; but we do not have sufficient evidence to support this speculation on the positions of *rfbH* and *rfbI*. The *rfbE* gene did not show any detectable homology with the sequenced *rfb* region from *S. typhimurium.* It seems that *rfbE* may have been inserted into the *rfb* region of *S. typhi* from some other source and so added an extra enzyme to the O-antigen biosynthesis in this species.

CHAPTER 6

Analysis of *rfb* genetic structure from natural isolates of *Salmonella* serogroups A, B and D

6.1 Introduction

O-antigen, encoded by the *rfb* gene cluster, is highly polymorphic among isolates of *Salmonella* from natural populations. The structure of the *rfb* region of strains *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 described in earlier chapters of this thesis has provided a basis for comparative genetic analysis. We analysed several strains of *Salmonella* serogroups A, B and D to see if the differences observed between *S. paratyphi A* IMVS1316, *S. typhimurium* LT2 and *S. typhi* Ty21a are representative of all strains of groups A, B and D. The *rfb* genetic structure was obtained by using colony blot and Southern hybridization techniques for the natural isolates studied; 11 strains of group A, 12 strains of group B and 6 strains of group D.

6.2 Analysing overall *rfb* genetic structure in strains of groups A, B and D using colony blots

A small bank of *Salmonella* strains of groups A, B and D was analysed; the strains studied are listed in Table 6.1. All strains including one positive control for each group and an *rfb* delete strain of *S. typhimurium* LT2 (P9029) were patched onto nitrocellulose filters (grid pattern shown in Fig 6.1D). Several plasmids spanning most of the *rfb* region of *S. typhimurium*

TABLE 6.1

Serogroup	Number	Strain
A	1	*Salmonella paratyphi A* (M107)
	2	*Salmonella paratyphi A* (M108)
	3	*Salmonella paratyphi A* (M109)
	4	*Salmonella paratyphi A* (M110)
	5	*Salmonella paratyphi A* (M111)
	6	*Salmonella paratyphi* (M217)
	7	*Salmonella paratyphi* (M218)
	8	*Salmonella paratyphi* (M219)
	9	*Salmonella nitra* (M354)
	10	*Salmonella kiel* (M355)
	11	*Salmonella paratyphi A* var *durazzo* (M 356)
	12	*Salmonella paratyphi A* IMVS1316 (M8)
D	13	*Salmonella enteritidis* (M125)
	14	*Salmonella panama* (M128)
	15	*Salmonella javiana* (M129)
	16	*Salmonella victoria* (M212)
	17	*Salmonella eastbourne* (M214)
	18	*Salmonella enteritidis* (M224)
	19	*Salmonella typhi* Ty2 (M18)
B	20	*Salmonella typhimurium* (M2)
	21	*Salmonella typhimurium* (M5)
	22	*Salmonella stanley* (M15)
	23	*Salmonella budapest* (M16)

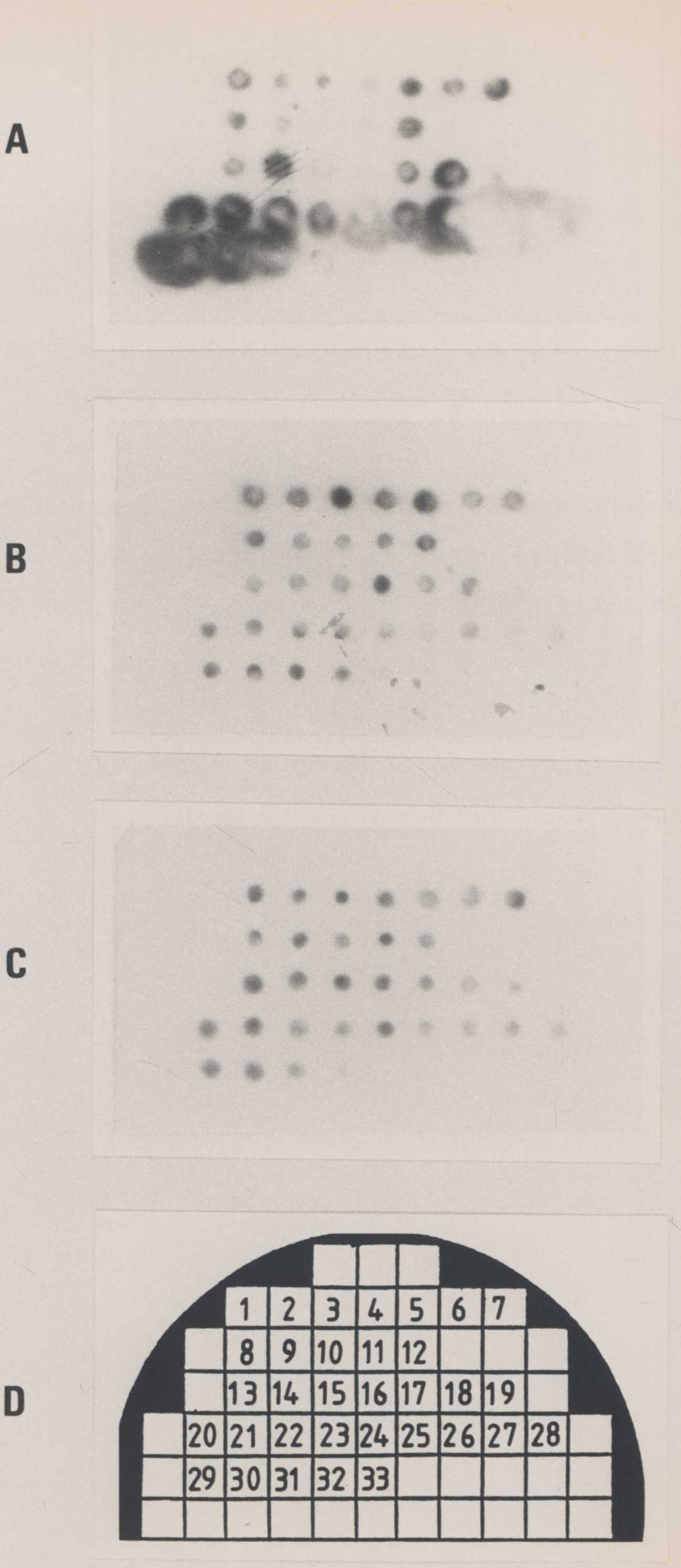
24	*Salmonella derby*	(M17)
25	*Salmonella agona*	(M117)
26	*Salmonella heidelberg*	(M118)
27	*Salmonella bredeney*	(M120)
28	*Salmonella hessarek*	(M121)
29	*Salmonella derby*	(M122)
30	*Salmonella saintpaul*	(M123)
31	*Salmonella typhimurium*	(M228)
32	*Salmonella typhimurium* LT2	(P9003)
33	*Salmonella typhimurium* LT2	(P9029)

The M and P numbers when in parentheses are laboratory numbers.

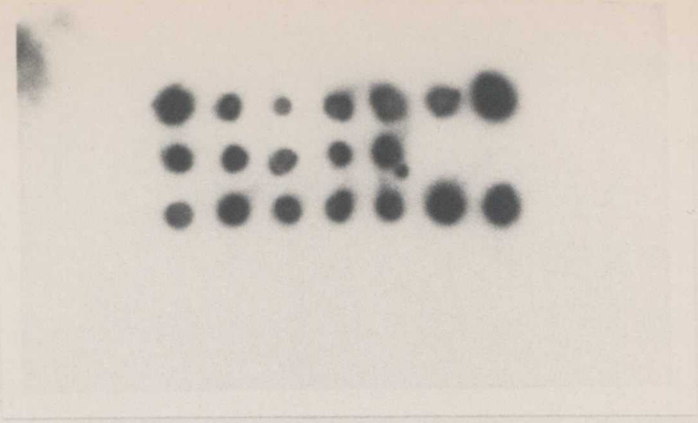
FIGURE 6.1

Analysing the *rfb* genetic structure in strains of groups A, B and D using colony blots

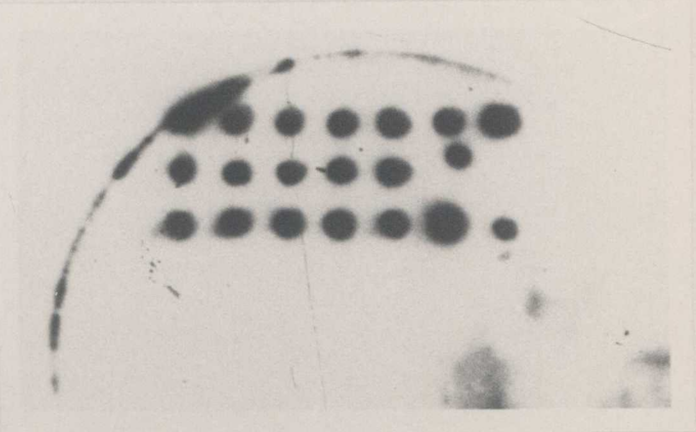
Autoradiograms of colony hybridization on nitrocellulose filters. The ^{32}P-labelled probes used for hybridization were **A**) pPR281, **B**) pPR303, **C**) pPR341, **E**) a 1.8 kb EcoRI purified fragment from pPR437, **F**) a 0.75 kb HindIII-KpnI DNA fragment from pPR620, and **G**) a 0.77 kb SalI-HindIII DNA fragment from pPR618. The grid pattern is shown in **D** and **H**. The numbers on the grid correspond to strains that are listed in Table 6.1.


A
B
C
D
1 2 3 4 5 6 7
8 9 10 11 12
13 14 15 16 17 18 19
20 21 22 23 24 25 26 27 28
29 30 31 32 33

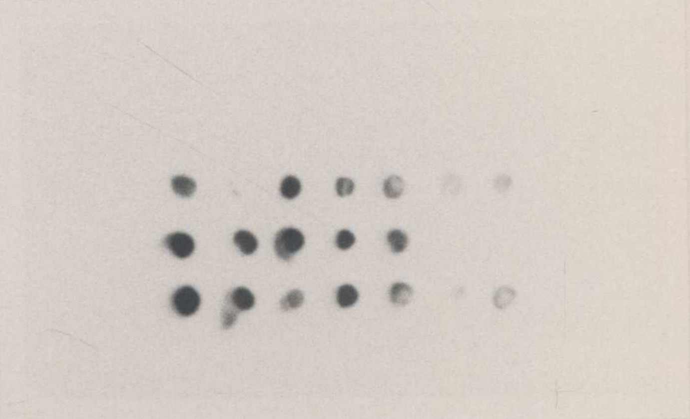
E



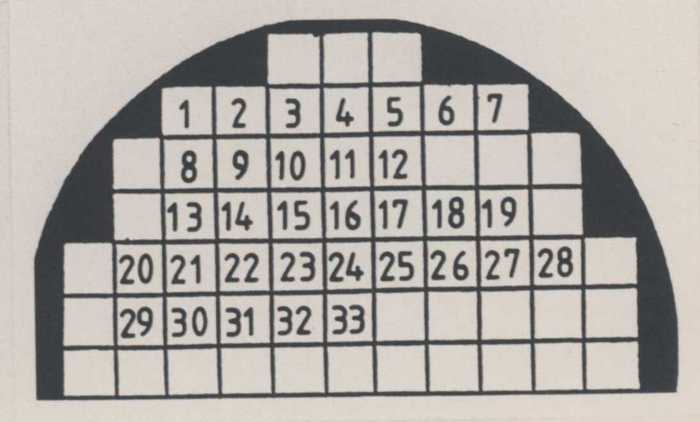
F



G



H


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

(Brahmbhatt *et al* 1988), were used as molecular probes against these strains in colony blots. The probes showed homology with all the strains of groups A, B and D (Fig 6.1A, B, C) except P9029 (deleted *rfb* region). A purified 1.8 kb fragment (pPR437) from *S. paratyphi A* IMVS1316, which did not show any detectable homology with *S. typhimurium* LT2 *rfb* DNA (section 3.5), was used as a probe to look for the region nonhomologous between these strains. This probe hybridized with all the strains of groups A and D, but did not show any homology with strains of group B (Fig 6.1E).

S. paratyphi A IMVS1316 contains the *rfbE* gene in an inactive form; due to a frame shift mutation (section 5.3). An *rfbE* specific probe, extending from position 8.95 to 9.70 (Fig 3.4), was derived from HindIII and KpnI digested plasmid pPR620 (Table 2.2) and used to see whether the *rfbE* gene was present in all strains of group A. The 0.75 kb fragment was purified from a gel and used to probe the whole bank. All group A strains seem to have the *rfbE* gene (Fig 6.1F) but none of the group B strains showed any homology.

An *rfbS* specific probe was also made to look for homology with *rfbJ* in strains of group B. The 0.77 kb SalI-HindIII fragment from plasmid pPR618 (Table 2.2) was purified from low melting agarose gel and used to probe the strain bank. DNA from all strains of groups A and D were hybridized with the probe used (Fig 6.1G). The probe did not hybridize with DNA from strains of group B. *rfbS* of *S. typhi* Ty21a was shown to have only 55% homolgy with the *rfbJ* gene of *S. typhimurium* LT2 at DNA level (Fig 4.9). This data of course does not show that *rfbJ* in all strains of group B is like *rfbJ* of *S. typhimurium* LT2.

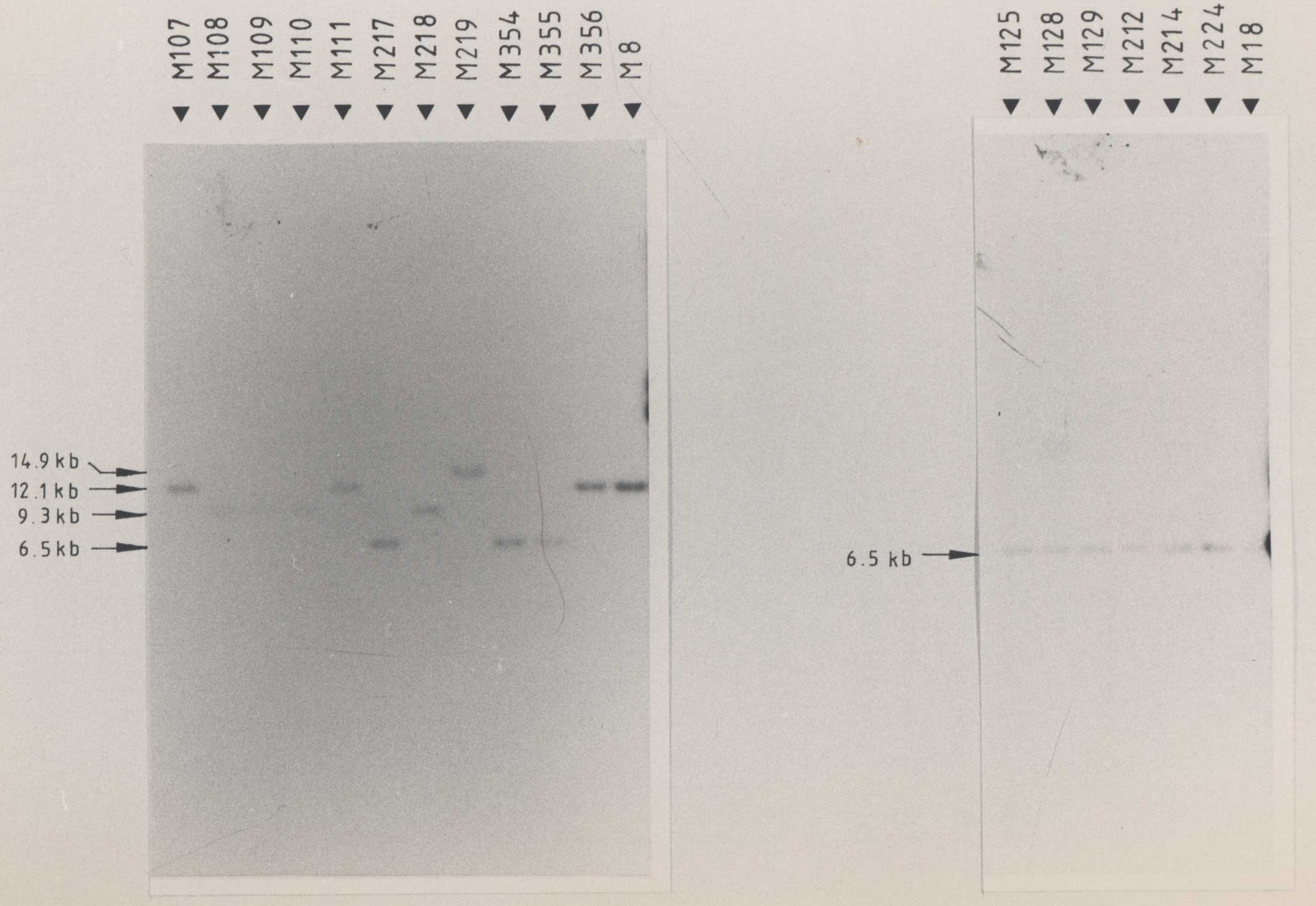
6.3 Occurrence of triplication in strains of group A

A 2.8 kb segment of DNA is triplicated in the *rfb* region of *S. paratyphi A* IMVS1316 (Fig 3.14). Both ends of the triplicated segment have been localised at base sequence level (Fig 5.7) but the cause of the segment being triplicated remains unclear. We looked for this triplication in the *rfb* region of the other strains of group A. Restriction analysis of the *rfb* region from *S. paratyphi A* IMVS1316 revealed that the triplicated region is flanked by two KpnI sites at positions 2.47 and 14.55 (Fig 3.4). A 1.8 kb EcoRI *S. paratyphi A* IMVS1316 DNA fragment that lies between positions 12.4 and 14.2 from pPR437 was used as a radioactive probe against KpnI digests of chromosomal DNA from all 11 strains of group A as well as from *S. paratyphi A* IMVS1316 (M8). The probe hybridized with a 14.9 kb fragment from M219, a 12.1 kb fragment from M107, M111 and M356, and a 9.3 kb fragment from M108, M109, M110 and M218. The KpnI fragment sizes of 14.9 kb, 12.1 kb and 9.3 kb correspond to the sizes expected if the 2.8 kb segment is repeated four, three or two times respectively. *S. paratyphi* (M217), *S. nitra* (M354) and *S. kiel* (M355) do not seem to have any repeat (Fig 6.2A). The same probe was also used to probe KpnI digests of chromosomal DNA from group D strains. In each case, the probe identified the expected (if only a single copy of the 2.8 kb segment were present) 6.5 kb KpnI fragment (Fig 6.2B). These results also indicate that the KpnI sites present at positions 2.47 and 8.95 (*S. typhi* Ty21a map)

FIGURE 6.2

Occurrence of triplication in strains of group A

Southern hybridization analysis of KpnI digests of chromosomal DNA from group A strains (**A**) and from group D strains (**B**). A 1.8 kb purified radioactive fragment of *S. paratyphi A* IMVS1316 *rfb* DNA (position 12.4 to 14.2) from pPR437 was used as a probe.


M107
M108
M109
M110
M111
M217
M218
M219
M354
M355
M356
M8
14.9 kb
12.1 kb
9.3 kb
6.5 kb
M125
M128
M129
M212
M214
M224
M18
6.5 kb
A
B

and 2.47 and 14.55 (*S. paratyphi A* IMVS1316) are conserved in all strains of group D and A, respectively.

6.4 Localising *rfbE* in strains of groups A and D

The *rfbE* gene has been identified and sequenced in *S. typhi* Ty21a (section 4.5.1) and *S. paratyphi A* IMVS1316 (section 5.3). Results of colony blots revealed that the *rfbE* gene was present in all strains of group A (section 6.2). Plasmid pPR620, carrying *rfbE* DNA, was used to probe EcoRI digested chromosomal DNA from these 11 strains of group A and 6 strains of group D to locate its position. The probe hybridized to the expected 2.3 kb fragment in strains of group A (M107, M108, M109, M110, M111, M217, M218, M219, M354 and M356) (Fig 6.3) and in strains of group D (M125 and M129) (Fig 6.4), indicating that in these strains, the *rfbE* gene is located on this conserved fragment (also in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316). A 1.9 kb fragment hybridized with the probe in strains M355, M128, M126 and M214, suggesting that while the same *rfbE* gene is present, either the EcoRI sites, corresponding to positions 8.6 and 10.9 in *S. typhi* Ty21a and 14.2 and 16.5 in *S. paratyphi A* IMVS1316, are not conserved in these strains or *rfbE* is present elsewhere on the chromosome. KpnI sites at positions 8.95 (*S. typhi* Ty21a) and 14.5 (*S. paratyphi A* IMVS1316) that lie within the *rfbE* coding region were conserved in all strains of groups A and D (section 6.3).

FIGURE 6.3

Localising the *rfbE* in strains of group A

The chromosomal DNA from group A strains digested with EcoRI and probed with ^{32}P-labelled pPR620 in a Southern hybridization experiment. The plasmids present in these strains seem to have homology with the vector pUC9. The extra bands visible in M107, M108, M109 and M218 correspond to the bands of plasmids present in these strains.

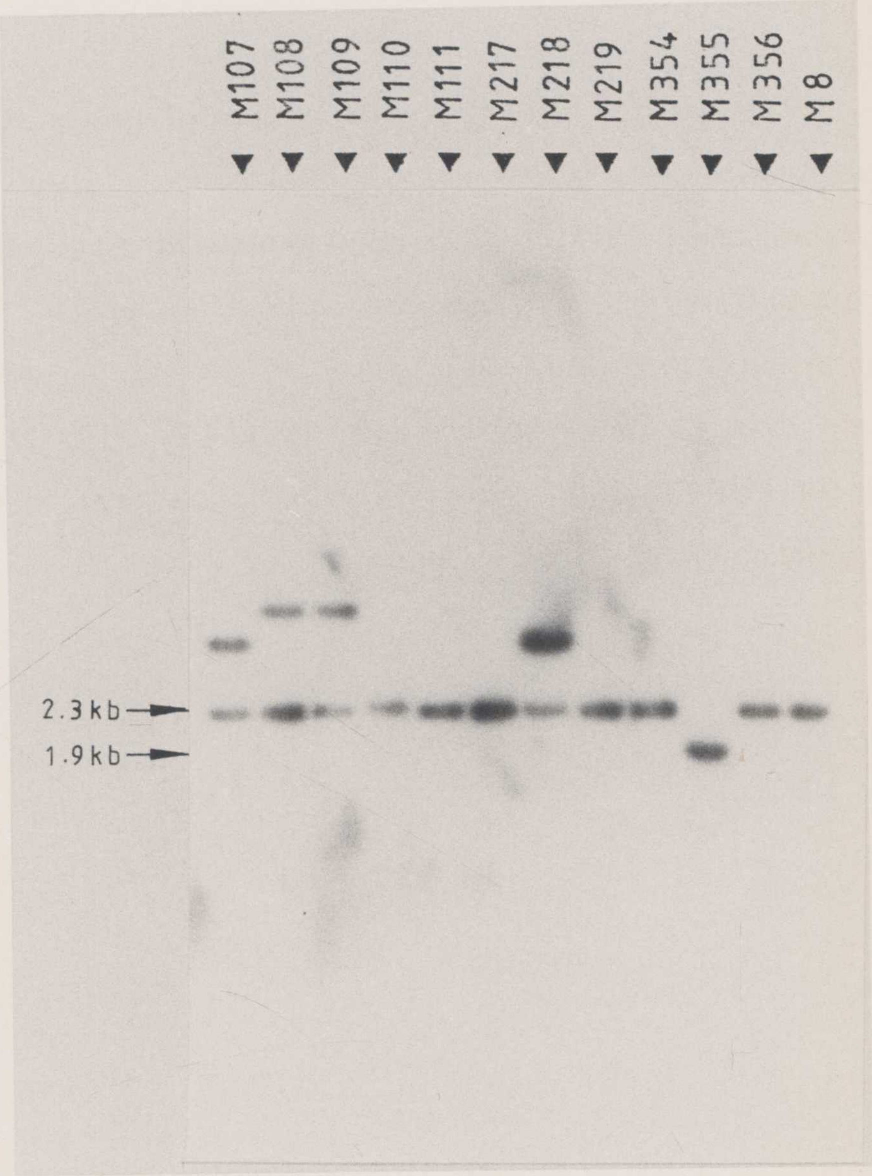
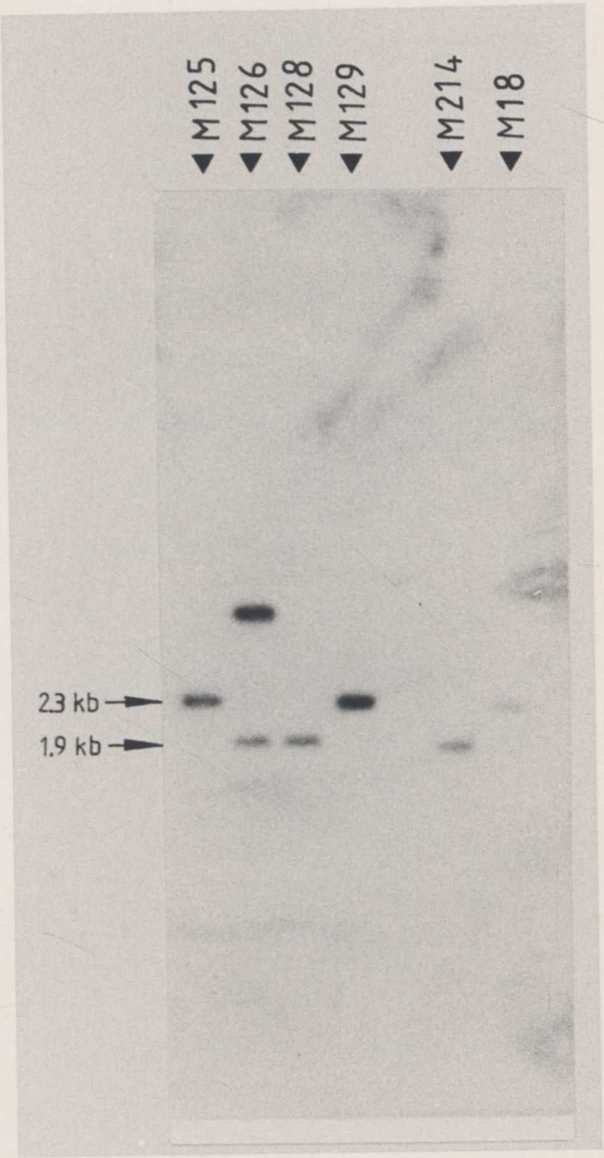

M107
M108
M109
M110
M111
M217
M218
M219
M354
M355
M356
M8
2.3 kb
1.9 kb

FIGURE 6.4

Localising the *rfbE* in strains of group D

Southern hybridization analysis of the chromosomal DNA from group D strains digested with EcoRI. The gel was blotted and probed with ^{32}P-labelled pPR620. An upper band visible in M126 corresponds to the plasmid present in this strain.



M125
M126
M128
M129
M214
M18
2.3 kb
1.9 kb

6.5 O-2 and O-9 specific LPS expression in other *Salmonellae*

The expression of O-2 and O-9 specific LPS in *S. typhimurium* and O-9 specific LPS in *S. paratyphi A* IMVS1316 has been demonstrated (sections 4.2 and 4.3). Other strains of group A: *S. kiel* (M355), *S. nitra* (M354), *S. paratyphi A* var *durazzo* (M356); and of group B: *S. derby* (M17) and *S. chester* (M124); were transformed with several plasmids carrying *rfb* DNA to see whether O-2 and O-9 specific LPS were expressed. O-antigens on the resultant hybrid strains were checked by agglutination tests using specific antisera. The results obtained are presented in Table 6.2. All group A strains carrying plasmid DNA conferring *rfbE* specificity (pPR652) showed a positive reaction with anti O-2 and with anti O-9 antisera indicating that both antigens (O-2 and O-9) are expressed in these hybrid strains. Group B strains harbouring plasmids pPR617 and pPR618 (with DNA specific for O-2 and O-9 epitope) also revealed the presence of mixed antigens on the surface with agglutination tests.

6.6 Summary and conclusions

Preliminary data obtained by using colony blots demonstrated that the overall *rfb* genetic structure with respect to homologous and nonhomologous regions in strains of group A, B and D is similar to the structure present in *S. paratyphi A* IMVS1316, *S. typhimurium* LT2 and *S. typhi* Ty21a, respectively. A 2.8 kb segment is repeated two, three or four times in the *rfb* region of *S. paratyphi A* strains. The function of the gene(s) present on this 2.8

TABLE 6.2

Expressing O-2 and O-9 specific LPS in other *Salmonellae*

The strains used to transform are listed here. The resultant hybrid strains were checked for their LPS by agglutination test. + or - indicate the positive and negative agglutination reaction, respectively. O-2, O-4 and O-9 represent the antisera used for O-2 (paratose), O-4 (abequose) and O-9 (tyvelose) epitope specificity.

Plasmid	Host	0-2	0-4	0-9
pPR618	*S. kiel* (M355)	+	-	+
	S. nitra (M354)	+	-	+
	S. paratyphi A var *durazzo* (M356)	+	-	+
	S. derby (M17)	-	+	+
	S. chester (M124)	-	+	+
pPR617	*S. derby* (M17)	+	+	-
	S. chester (M124)	+	+	-
pPR620	*S. kiel* (M355)	+	-	-
	S. nitra (M354)	+	-	-
	S. paratyphi A var *durazzo* (M356)	+	-	-
pPR652	*S. kiel* (M355)	+	-	+
	S. nitra (M354)	+	-	+
	S. paratyphi A var *durazzo* (M356)	+	-	+

kb EcoRI segment has not been determined and the cause of the repetition is unclear.

All the strains of groups A and D studied seem to carry the *rfbE* gene on a 2.3 kb EcoRI fragment (except M126, M128, M214 and M355 where the probe hybridized to a 1.9 kb fragment), as in *S. paratyphi A* IMVS1316 and *S. typhi* Ty21a. The *rfbE* gene in *S. paratyphi A* IMVS1316 was shown to contain a frame shift mutation rendering it inactive (Chapter 5); it would be interesting to look for any other mutational changes in the *rfbE* of these strains of group A. The strains of serogroup A, *S. kiel*, *S. nitra* and *S. paratyphi A* var *durazzo*, all show complementation with the *rfbE* from *S. typhi* Ty21a. It is likely that all naturally occurring group A strains arose from group D strains by losing the epimerase (*rfbE*) function.

CHAPTER 7

Final discussion and conclusions

7.1 Introduction

This thesis describes the molecular cloning and analysis of the *rfb* gene clusters of *S. paratyphi A* IMVS1316 (serogroup A) and *S. typhi* Ty21a (serogroup D). The *rfb* gene cluster from *S. typhimurium* had been studied in our laboratory (Brahmbhatt *et al* 1986, 1988; P. Wyk, Ph.D. thesis 1988). The genetic basis of O-antigen variation in serogroups A, B and D of *Salmonella*, and the evolutionary derivation of serogroup A from a serogroup D progenitor are also discussed.

7.2.1 Cloning and restriction enzyme analysis of *rfb* gene clusters from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316

The *rfb* gene clusters of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 were cloned in cosmid vectors. *rfb* DNA between positions 9.95 and 12.20 of *S. typhimurium* (Brahmbhatt *et al* 1988) was used to probe the cosmid banks. Eight cosmid clones of *S. typhi* Ty21a and two of *S. paratyphi A* IMVS1316 were identified using this probe. Partial EcoRI maps of these ten overlapping cosmids were then deduced (Fig 3.2). None of the cosmid clones covered the complete *rfb* gene cluster. A similar situation also occurred for *S. typhimurium* LT2. This suggests that clones of the whole *rfb* region may be lethal or unstable. Most cloned fragments from cosmids carrying part of the *rfb* DNA were stable in high copy number vectors.

Restriction analysis reveals that a common 11 kb *rfb* DNA fragment of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 is homologous with *S. typhimurium* LT2 *rfb* DNA between the EcoRI sites at positions 9.57 and 20.59. It was assumed that the same level of homology would be present in the region to the right of the EcoRI site (position 20.59) as the 5 EcoRI sites from position 20.59 to 40.09 were conserved and it is probable that the *rfb* cluster terminates within this region. The EcoRI site at position 0 of *S. typhimurium* LT2 is conserved and used as point 0 for this study. The other EcoRI site at position 9.57 of *S. typhimurium* LT2 corresponds to EcoRI sites at position 10.9 and 16.5 in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316, respectively. Restriction enzyme maps from position 0 to 10.9 in *S. typhi* Ty21a and 0 - 16.5 in *S. paratyphi A* IMVS1316 have been constructed. The regions of DNA homology and nonhomology in the *rfb* regions of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 with respect to *S. typhimurium* LT2 were then analysed in detail. The only gross difference found between *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 is that a 2.8 kb DNA segment is repeated three times in *S. paratyphi A* IMVS1316, giving rise to 5.6 kb additional DNA (Fig 3.14).

7.2.2 Defining the junction between the homologous and nonhomologous regions of *S. paratyphi A* IMVS1316 and *S. typhimurium* LT2

Southern hybridization demonstrated that the 6.8 (position 0 to 6.8, *S. typhi* Ty21a and *S. paratyphi A* IMVS1316) and 2.3 (position 8.6 to 10.9, *S. typhi* Ty21a; 14.2 to 16.5, *S. paratyphi A* IMVS1316) kb EcoRI fragments (Fig 3.4) from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 share homology with the 9.57 kb EcoRI (between positions 0 and 9.57) DNA fragment of *S. typhimurium*. The junctions between homologous and nonhomologous regions at both ends were localised approximately by restriction mapping and then defined by heteroduplex technique. It should be noted that although heteroduplex analysis indicates homology in the region from position 8.17 to 9.57, little restriction site homology was detected. The results obtained are summarised in Fig 3.14.

7.3.1 Localising the genes responsible for specificity in the biosynthesis of paratose (*rfbS*) and tyvelose (*rfbE*)

We have shown that the 1.2 kb EcoRI - HindIII *rfb* DNA fragment (position 10.9 to 9.7) on the *S. typhi* Ty21a map encodes a gene equivalent to the *rfbJ* gene of *S. typhimurium* LT2; this gene has been called *rfbS*.

The *rfbS* gene has also been localised on the corresponding EcoRI - HindIII *rfb* DNA fragment (16.5 to 15.3) of *S. paratyphi A* IMVS1316. An extra gene, which has an epimerase function, is present on the 1.4 kb BglII - EcoRI fragment (position 10.0 to 8.6; *S. typhi* Ty21a unit) of pPR657 and can make a *S. paratyphi A* strain produce tyvelose. Similarly, the product of the *rfbS* gene makes a *S. typhimurium* strain produce a mixed type of LPS (both O-2 and O-4 specific).

It has been reported that the transferase in group B *Salmonella*, is specific for CDP-abequose, although it does have a low affinity for CDP-tyvelose (Osborn and Weiner 1968). It was also thought that a specific enzyme might be involved in catalysing the transfer of tyvelose in group D organisms. Sasaki and Uchida (1974) showed that tyvelose transferase has an affinity for CDP-paratose; it is possible that the same enzyme is capable of transferring paratose in group A organisms. Our results from complementation studies suggest that the transferases in strains of groups A, B and D can transfer all these three sugars, from CDP-abequose, CDP-paratose and CDP-tyvelose; note that assays were not quantitative and the ratio of the products present is not known. It seems likely that tyvelose transferase has an affinity both for CDP-paratose and CDP-tyvelose. CDP-paratose-2-epimerase has been partially purified from *Y. pseudotuberculosis* Type IV (Matsuhashi 1966). Matsuhashi (1966) found that, at equilibrium, 55 to 58% of the compound present was CDP-tyvelose and 42-45% was CDP-paratose. Since group D LPS contains only tyvelose, it would be interesting to determine why tyvelose is selectively transferred. We expect that the enzyme may possess a much higher affinity for CDP-tyvelose than for CDP-paratose, but possibly some other mechanism could be involved in regulating CDP-paratose concentration. Presumably the polymerization of O-2 and O-9 like O-antigen in *S. typhimurium* occurs in the same way as for O-4 LPS. Shibaev (1978) and Shibaev *et al* (1982) have found that processing is not very specific in the polymerase system.

7.3.2 Homology between predicted *rfbS* and *rfbJ* of *S. typhimurium*

The products of genes *rfbS* and *rfbJ* both have dehydrogenase function and use the same substrate (CDP-4-keto-3,6-dideoxy glucose) for their reactions. We sequenced the *rfbS* gene from *S. typhi* Ty21a. A comparison of the predicted amino acid sequence of the protein encoded by *rfbS* with the amino acid sequence of rfbJp of *S. typhimurium* LT2 (P. Wyk, Ph.D. thesis 1988) showed substantial homology. Good homology is found in a smaller region at the N-terminus and a longer stretch nearer the C-terminus (Fig 4.8). It appears that the latter portion of the protein may be involved in binding and modifying the substrate. Both enzymes use NADPH as coenzyme; presumably the N-terminus region of the protein is involved in forming the NADPH binding domain (this is discussed in section 7.3.4).

There was a high degree of variation (approximately 55% homology only) between these genes at the nucleotide level. We estimated the number of substitutions per site for the nucleotide sequences of these two genes (Nei 1987). Alignment of these genes showed 420 differences including 79 deletions/insertions; these made it difficult to determine the best fit. The estimated number of nucleotide substitutions per site between the *rfbJ* and *rfbS* genes is 0.71. We could not estimate the rate of synonymous substitution because of the large number of deletions/insertions present in the nucleotide sequences. It is therefore very difficult to estimate the divergence time between these two apparently closely related genes (of almost similar function and mapping at the same position on the chromosomes). The polymorphism has probably been present for a very long time. It seems likely that both genes were derived from a common ancestor before the divergence between *E. coli* and *S. typhimurium* (120-160 Myr; the time estimated for the divergence between *E. coli* and *S. typhimurium* by Ochman and Wilson 1987). The C-terminus portion of the protein has extensive homology with the protein encoded by *rfbJ* while the N-terminus portion has a higher degree of variation in the predicted amino acid sequence. This may have been due to selection pressure on the C-terminus region resulting from a structural requirement for binding and modifying of the substrate, since both enzymes use the same substrate for their reactions. It is also possible that a recombination event occurred at some stage to give both genes the same C-terminus end after much divergence.

7.3.3 Comparison of *rfbE* (*S. typhi* Ty21a) with *galE* of *E. coli*

The reversible NAD-dependent 2-epimerisation of CDP-3,6-dideoxy glucose (paratose) and CDP-3,6-dideoxy mannose (tyvelose) has been studied in a strain of group D *Salmonella* (*S. enteritidis*) and in *Y. pseudotuberculosis*, Type IV (Matsuhashi 1966). Matsuhashi (1966) showed that CDP-paratose is reversibly converted to CDP-tyvelose by an epimerisation at C-2. This epimerisation is NAD-dependent and so resembles the UDP-galactose-4-epimerase sequenced recently in *E. coli* (Lemaire and Müller-Hill 1986). The nucleotide sequence of the *rfbE* (CDP-paratose-2-epimerase) in *S. typhi* Ty21a has been determined (section 4.5.1). The amino acid sequence of proteins encoded by the *rfbE* and *galE* show a good homology at the N-terminus region (Fig 4.10). The predicted secondary structure of these two proteins also indicated a region of homology in the N-terminus portion.

7.3.4 NAD$^+$ dehydrogenase binding domain in rfbSp, rfbEp and galEp

The amino acids of proteins encoded by *rfbJ* of *S. typhimurium*, *rfbS* (*S. typhi* Ty21a), *rfbE* (*S. typhi* Ty21a) and *galE* of *E. coli* could be aligned with the amino acids from the NAD$^+$ binding domain in structurally known dehydrogenases; the first three predicted beta strands (βA, βB and βC) of rfbSp and rfbJp show a very good fit, it seems likely that the first 105 amino acids of rfbSp are involved in NADPH binding. The latter portions toward the C-terminus are probably involved with substrate binding. Similarly, the first 130 amino acids in the N-terminus portion of rfbEp and galEp seem to have a role in forming an NAD$^+$ binding domain; the first two predicted beta strands (βA and βB) show a reasonable fit with corresponding strands of known dehydrogenases as well as with dehydrogenases encoded by *rfbS* and *rfbJ*.

7.3.5 Low G+C content

One interesting feature of the *rfb* region sequenced from *S. typhi* Ty21a was the low G+C content. This comprises only 35% of all bases, whereas in *S. typhimurium* and *E. coli* the G+C content is generally 51%. There may be a cut off point at position 10.7, right of the EcoRI (10.9), where the G+C content rises to about 40%. Low G+C content has also been found in the corresponding *rfb* region of *S. typhimurium* LT2 (P. Wyk, Ph.D. thesis 1988). Sueoka (1962) presented a theory to account for diversification of the G+C content of the bacterial genome; that is, the G+C content of DNA of a given bacterium will be determined by the effective base conversion rate u (G·C to A·T) and v (A·T to G·C). The G+C content of equilibrium (p) is v/(v+u). The biased mutation pressure acts uniformly on the entire genome and the nature of this biased A·T/G·C pressure is unknown. This suggests that the genes for 3,6-dideoxyhexose (paratose or abequose) biosynthesis may have been transferred from some other bacterial genome of low G+C content and that the *Salmonellae* may have evolved with the polymorphism. The concept of species to species transfer of genes in evolution has been described as a cross-species gene transfer (Syvanen 1986) or a horizontal gene transfer (Busslinger *et al* 1982). A species to species transfer event has been found in the evolution of several bacterial pathogenic determinants (Yamamoto *et al* 1987). However, the *rfb* regions in *S. typhimurium* LT2 and *S. typhi* Ty21a are the first examples of a gene transfer, of low G+C content, for a basic function of the cell.

7.3.6 Boundaries of homologous and nonhomologous DNA in the vicinity of positions 9.57 (*S. typhimurium* LT2) and 5.5 (*S. typhi* Ty21a)

Heteroduplex results presented in chapter 3 show that the region of homology between the three strains studied extends about

1.4 kb to the left of position 9.57 (*S. typhimurium* LT2). Sequence analysis of this region from *S. typhi* Ty21a and *S. typhimurium* LT2 indicates a sharp cut off point at position 9.24 (*S. typhimurium* LT2). Scattered homology was seen in the rest of this region. There appears to be a clear cut off mark at position 5.71 (*S. typhi* Ty21a map) when the region sequenced in the vicinity of 5.5 is compared to the corresponding sequence in *S. typhimurium* LT2. The open reading frame which lies in this sequenced region shows only 2% differences compared with the homologous corresponding coding region ORF353 of *S. typhimurium* LT2. This level of variation is in agreement with the finding of Milkman and Crawford (1983) who found that 2 - 4% differences in the nucleotide sequences of translated regions of the *trp* operon in several strains of *E. coli*.

7.4.1 Frame shift mutation in the *rfbE* of *S. paratyphi A* IMVS1316

Physical data, discussed in sections 4.2 and 4.3, indicated that the gene for tyvelose specificity (*rfbE*) is present but nonfunctional in *S. paratyphi A* IMVS1316. We sequenced this *rfbE* gene and found a one base deletion at position 8. This one base deletion in the coding region causes a translation frame shift and a stop codon occurs after just three amino acids. Minicell data confirmed the absence of protein encoded by *rfbE* in *S. paratyphi A* IMVS1316. It had been suggested that group A organisms might be mutants of group D strains (Böhlck 1965); Sasaki and Uchida (1974) have demonstrated that mutants of *S. durban* (group D), carrying the O-antigen of group A *Salmonella*, were lacking in CDP-paratose-2-epimerase activity. These mutants could not form CDP-tyvelose. We have shown that, in at least one instance, a naturally occurring group A strain arose in this way.

7.4.2 Localising the ends of the triplicated DNA segment in *S. paratyphi A* IMVS1316

The nucleotide sequence at both ends of the triplicated DNA segment in *S. paratyphi A* IMVS1316 was determined in order to look for transposition events or repetitive structures in this region. We did not find any evidence of a transpositional event or structural repeat in the sequence present at the junctions. The origin of this triplication remains unclear. It is interesting to consider whether the genes present on this segment, repeated three times, could have any effect on the expression of neighbouring genes in the *rfb* gene cluster. One could hypothesize that the DNA segment triplicated is larger than a gene and hence the genes which include the ends are likely to be present once in functional form, with intervening genes replicated as complete genes. The triplication may thus affect only gene dosage and have negligible effect on phenotype. The function of the gene(s) present on the triplicated segment has not yet been determined.

7.5 *rfb* genetic structure in natural isolates of *Salmonella* serogroups A, B and D

We have studied *rfb* genetic structure in several strains of groups A, B and D independently isolated from natural populations. Southern hybridization analysis reveals that a 2.8 kb DNA segment is repeated two, three or four times in the *rfb* region of most *S*.

paratyphi A strains tested (section 6.3). *S. nitra*, *S. kiel* (rare isolates of group A) and *S. paratyphi* (M217) do not have any repeat. *S. paratyphi A* IMVS1316 has been shown to be a derivative of group D *Salmonella* (section 5.8) as it contains a full *rfbE* gene with a single mutation in it. All other group A strains tested also seem to contain the *rfbE* gene, but in inactive form, indicating their derivation from group D progenitors. We have not yet defined the mutations present in the *rfbE* gene for each of these strains.

APPENDIX

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

Verma, N.K., Quigley, N.B. and Reeves, P.R. 1988. O-antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. *J. Bacteriol.* 170 : 103-107.

Verma, N.K. and Reeves, P.R. 1988. Analysis and nucleotide sequence of genes encoding paratose and tyvelose specificity in a *Salmonella* strain of serogroup D. Manuscript in preparation.

JOURNAL OF BACTERIOLOGY, Jan. 1988, p. 103–107
0021-9193/88/010103-05$02.00/0

Vol. 170, No. 1

O-Antigen Variation in *Salmonella* spp.: *rfb* Gene Clusters of Three Strains

NARESH K. VERMA,* NEIL B. QUIGLEY, AND PETER R. REEVES

Department of Microbiology, the University of Sydney, Sydney, New South Wales 2006, Australia

Received 20 April 1987/Accepted 23 September 1987

The O antigens of *Salmonella* serogroups A, B, and D differ structurally in their side-chain sugar residue. The genes encoding O-antigen biosynthesis are clustered in the *rfb* operon. We report here the molecular cloning and analysis of the *rfb* operons of *Salmonella paratyphi A* (serogroup A) and *S. typhi* (serogroup D). The regions of DNA nonhomology between the *rfb* operons of these serogroup A, B, and D representatives are identified, and the evolutionary derivation of serogroup A from a serogroup D progenitor is discussed.

Lipopolysaccharides are integral components of the outer membranes of gram-negative bacteria and generally consist of three structural regions: the lipid A moiety, which is hydrophobic and is embedded in the outer membrane, an oligosaccharide core, and a polysaccharide chain commonly known as the O antigen. The O-antigen chains of *Salmonella typhi*, *S. paratyphi A*, and *S. typhimurium*, representative of *Salmonella* serogroups D, A, and B respectively, have identical trisaccharide subunit backbones (mannosyl-rhamnosyl-galactosyl). These chains can be distinguished, since each has a different 3,6-dideoxyhexose attached to the mannosyl residues as a side branch. The side branch is tyvelose in *S. typhi*, paratose in *S. paratyphi A*, and abequose in *S. typhimurium* (11, 13). Much of the antigenic variation among *Salmonella* species is the result of genetic variation in the *rfb* gene cluster, which maps at 42 min on the chromosome of *S. typhimurium* LT2 (17). We previously reported the cloning and restriction analysis of the *rfb* cluster from *S. typhimurium* LT2 (1, 2). We report here the cloning and analysis of the *rfb* loci from *S. typhi* and *S. paratyphi A* strains and compare them with the *S. typhimurium rfb* locus.

We have used the traditional taxonomy for *Salmonella* species, although as Brenner (3) points out, the statement in the 1974 (8th) edition of *Bergey's Manual* that "Scientifically none of the present methods of nomenclature of *Salmonella* is satisfactory" is as true today as more than a decade ago. It is clear that all *Salmonella* strains are in one species (3, 7), and Ewing (7) has used a single species name, although the International Subcommittee has still not made a recommendation. It is important to note that the strains of *S. typhimurium*, *S. paratyphi A*, and *S. typhi* used in this study are in reality not only members of the same species, but also of the same subspecies, and the homology and nonhomology we describe for different regions of the *rfb* gene cluster must be seen in this light.

MATERIALS AND METHODS

Bacteria and plasmids. Strains used are listed in Table 1. A series of plasmids which carry DNA spanning the entire *rfb* operon of *S. typhimurium* LT2 (2) were used in this study.

Enzymes and radiochemicals. All enzymes and radiochemicals used in this study were obtained as described previously (2).

DNA techniques. The methods used for DNA preparation, agarose gel electrophoresis, radioactive labeling of DNA, in situ DNA hybridization, autoradiography, ligation, and bacterial transformation were those described by Maniatis et al. (14). Molecular size standards were used as described in the companion paper (2).

Cosmid cloning. *Sau*3A partial digests of chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 were prepared by the method of Maniatis et al. (14). The cosmid used in this experiment was pHC79 (10). It was digested with *Bam*HI and ligated with the partial *Sau*3A fragments in a final volume of 20 µl by mixing 2 µg of insert DNA with *Bam*HI-digested and phosphatase-treated vector DNA in a molar ratio of 1:3. A 3-µl portion of this ligation mixture was packaged in vitro by the method of Sternberg et al. (19), with a commercially available packaging kit (Promega Biotec, Madison, Wis.). DH1 was used as the host strain for infection and was plated out on ampicillin agar (containing 100 µg of ampicillin per ml). 2×10^4 recombinants per µg of ligated *S. typhi* DNA and 10^3 recombinants per µg of ligated *S. paratyphi* DNA were recovered.

Preparation of heteroduplex DNA and electron microscopy. Heteroduplex formation and DNA spreading were performed by the method of Davis and Parkinson (6) with minor modifications. Grids were prepared by the basic protein film technique (12) and stained with uranyl acetate or rotary shadowed with platinum-palladium.

RESULTS

A 2.25-kilobase (kb) *Kpn*I fragment, which lies approximately in the middle of the *S. typhimurium rfb* operon (positions 9.95 to 12.20), was isolated from pPR300 (2) and used as a radioactive probe to screen cosmid libraries for the presence of homologous DNA. Eight clones in the *S. typhi* Ty21a library and two clones in the *S. paratyphi A* IMVS1316 library were identified with this probe. Cosmid DNAs from these 10 clones were digested with *Eco*RI and run on agarose gels; they were found to have a number of common bands, many of which also occur in *S. typhimurium* (2). These fragments were assumed to be homologous to those of *S. typhimurium*, and on this basis partial maps of the cosmids were deduced (Fig. 1).

Hybridization of the same 2.25-kb radioactive probe with blotted *Eco*RI digests of all cosmids and with chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 identified a single fragment in each case which, with the exception of pPR429, had the same mobility as the 11.02-kb

---

* Corresponding author.

TABLE 1. Bacterial strains

Strain	Characteristics	Source or reference
E. coli K-12 DH1	*supE44 gyrA96 recA endA1 thi-1 hsdR17 relA1*	9
S. typhimurium LT2 SL1654	*hsdL6 trpC2 nml H1*b *fla-66 H2*-enx *rpsL120 xyl-404 ilv-452 metE551 metA22 hsdSA29*	I. Beacham
S. paratyphi A IMVS1316		S. Dixon
S. typhi Ty2 Ty21a	*galE*	8

*Eco*RI fragment of *S. typhimurium* DNA (positions 9.57 to 20.59) from which the probe was derived (Fig. 2).

All *Eco*RI fragments from the *S. typhi* and *S. paratyphi A* cosmid clones were subcloned into plasmid vectors pBR325 and pUC9 (Fig. 1). However, only the relevant subclones are discussed here. The 11.0-kb fragments corresponding to the *S. typhimurium* fragment from positions 9.57 to 20.59, which covers most of the *rfb* gene cluster, were examined in detail. Restriction enzyme digests with *Hin*dIII, *Kpn*I, *Cla*I, *Hpa*I, *Pst*I, *Xba*I, and *Bgl*II gave identical patterns to those given with the corresponding *S. typhimurium* fragment. A *Sac*I site which was absent in *S. typhimurium* was found in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 and mapped at a position corresponding to coordinate 12.6 of *S. typhimurium*, and a *Sac*I site at coordinate 9.97 in *S. typhimurium* LT2 was absent in both *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. We conclude that DNA between the *Eco*RI sites at positions 9.57 and 20.59 in *S. typhimurium* LT2 is essentially homologous with the corresponding regions in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 DNA. This conclusion was supported by hybridizing radioactive *S. typhimurium* DNA from this region with blotted DNA fragments from double enzyme digests of *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 DNA from this region.

Restriction site mapping was carried out on the 6.8-, 2.3-, and 1.8-kb *Eco*RI fragments from *S. typhi* Ty21a and the 6.8-, 2.8-, 2.3-, and 1.8-kb *Eco*RI fragments from *S. paratyphi A* IMVS1316. We determined the order of these fragments by using them separately as probes to hybridize with double enzyme digests of chromosomal DNA; we then constructed a restriction map of this region, from positions 0 to 10.9 of *S. typhi* Ty21a and positions 0 to 16.5 of *S. paratyphi A* IMVS1316 (Fig. 3). All restriction sites on the 6.8-, 2.3-, and 1.8-kb fragments from *S. paratyphi A* IMVS1316 were found to map at the same positions on the corresponding fragments of *S. typhi* Ty21a. The 2.8-kb *Eco*RI fragment, which stains with an intensity approximately twice that expected by comparison with adjacent bands, appears from its restriction map to duplicate the adjacent ends of the 6.8- and 1.8-kb *S. typhi* Ty21a *Eco*RI fragments (Fig. 3). We conclude that a 2.8-kb region is triplicated (Fig. 4). *Nru*I, *Hpa*I, *Eco*RI, and *Cla*I digests all gave a 2.8-kb fragment, which was shown by densitometry of gel photographs to be present at twice the molarity of other fragments. This confirmed the 2.8-kb triplication and showed that one end lies between positions 4.87 and 5.50. Further confirmation was given by the existence of the expected 12.1-kb *Kpn*I fragment in the chromosome (pPR615) which extends from position 2.47 to position 14.55 (Fig. 3).

To identify the region of nonhomology between these three *Salmonella* species, we purified a 9.57-kb *Eco*RI *S. typhimurium* DNA fragment from pPR301 (2) for use as a radioactive probe against *Eco*RI digests of chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. The probe, which contained *S. typhimurium* DNA between map positions 0 and 9.57, hybridized strongly with the 6.8-kb and weakly with the 2.3-kb *Eco*RI fragments from these strains. In addition to these fragments, the probe hybridized weakly with the 2.8-kb *Eco*RI fragment unique to the *S. paratyphi A* IMVS1316 digests (Fig. 5). The common 1.8-kb fragment did not exhibit any detectable homology with the 9.57-kb *S. typhimurium* DNA probe.

To determine whether the homology extended beyond *rfb* toward the *his* operon, we used plasmid pPR294 (1), which covers 3.35 kb directly to the left of the *Eco*RI site at

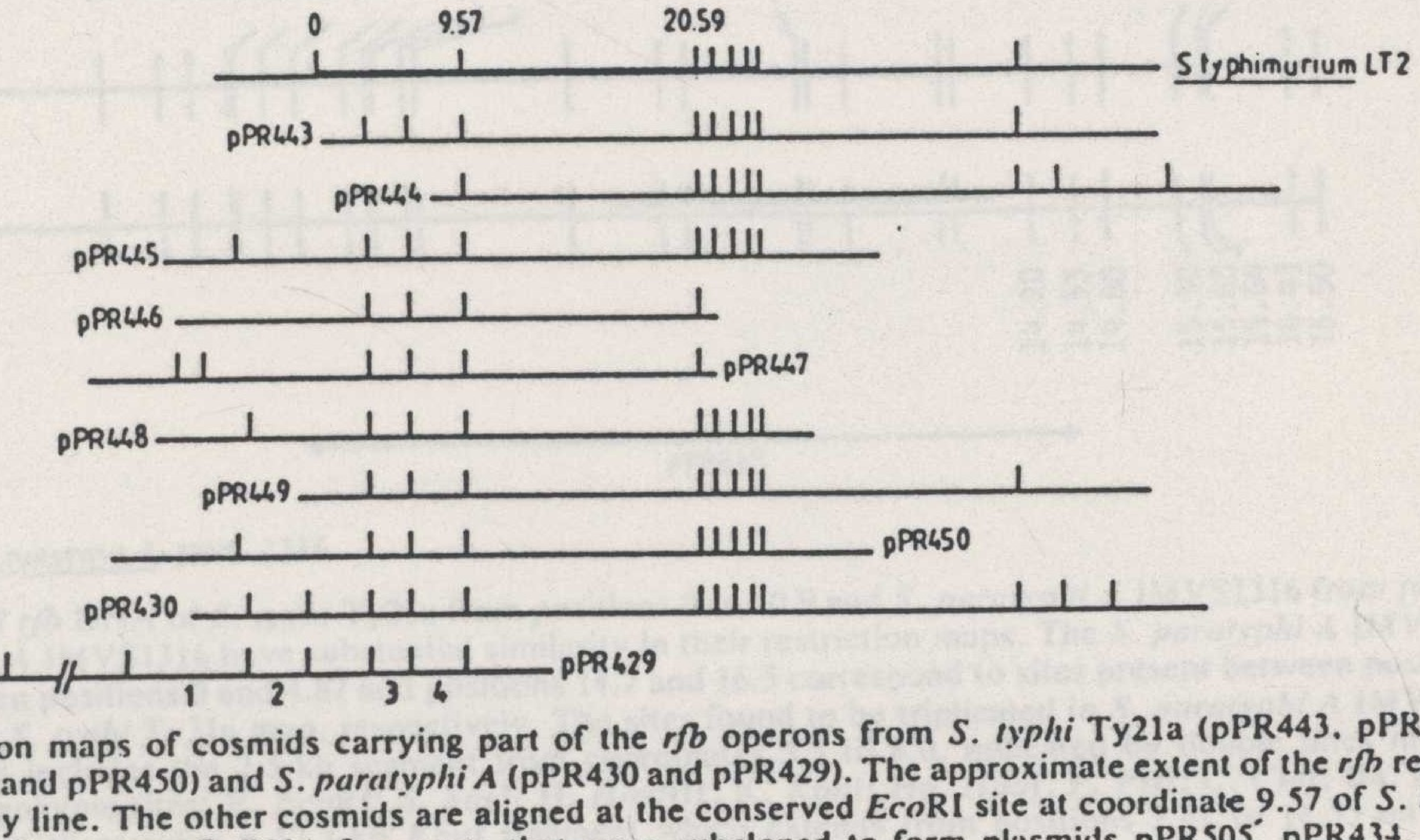


FIG. 1. *Eco*RI restriction maps of cosmids carrying part of the *rfb* operons from *S. typhi* Ty21a (pPR443, pPR444, pPR445, pPR446, pPR447, pPR448, pPR449, and pPR450) and *S. paratyphi A* (pPR430 and pPR429). The approximate extent of the *rfb* region of *S. typhimurium* LT2 is indicated by a heavy line. The other cosmids are aligned at the conserved *Eco*RI site at coordinate 9.57 of *S. typhimurium* LT2. The numbers 1, 2, 3, and 4 indicate the *rfb* DNA fragments that were subcloned to form plasmids pPR505, pPR434, pPR436, and pPR437, respectively.

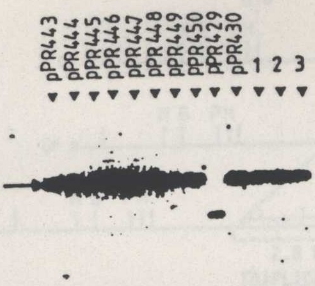


FIG. 2. *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 cosmid clones digested with *Eco*RI and probed with the 2.25-kb *Kpn*I radioactive fragment of *S. typhimurium* LT2 *rfb* DNA (positions 9.95 to 12.20) from pPR300. Lanes 1, 2, and 3 contain *S. typhimurium* LT2, *S. typhi* Ty21a, and *S. paratyphi A* IMVS1316 chromosomal *Eco*RI digests, respectively.

position 0, as a molecular probe to hybridize with several restriction enzyme digests of chromosomal DNA from *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. The restriction enzyme sites toward the *his* operon for *Eco*RI, *Hpa*I, *Bgl*II, and *Pst*I were mapped previously in *S. typhimurium* (1, 4) and found to be conserved in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316.

Thus, the restriction analyses presented above showed that most of the sites from positions 0 to 5.5 and from position 9.57 rightward on the *S. typhimurium rfb* map were conserved, with the 9.57-kb *Eco*RI site of *S. typhimurium* corresponding to the *Eco*RI sites at position 10.9 of *S. typhi* Ty21a and position 16.5 of *S. paratyphi A* IMVS1316. The junction between the homologous and nonhomologous regions was localized by heteroduplex analysis as follows.

Plasmid pPR505, which consists of a 6.8-kb *Eco*RI DNA fragment from *S. paratyphi A* IMVS1316 (positions 0 to 6.8) cloned in pcos2EMBL (16) was digested with *Eco*RI. Plasmid pPR508, which is pGB2 (5) carrying the 9.57-kb *Eco*RI fragment (positions 0 to 9.57) from *S. typhimurium*, was linearized with *Sal*I. Vectors pGB2 and pcos2EMBL were used because they exhibit no homology with each other (data not shown). The digestion products were allowed to hybridize in a heteroduplex reaction. Denaturation and renaturation of these plasmid DNA fragments gave rise to double-stranded linear molecules which branched into two single strands at one end (Fig. 6). By measuring the lengths of double-stranded segments of these molecules, we found that the length of homologous DNA was about 5.8 ± 0.06 kb (average of 10 measurements), showing that the region of homology extends from positions 0 to 5.8 of this strain.

To localize the other end of the nonhomologous region, plasmid pPR617, which contains the 2.3-kb *Eco*RI fragment of *S. paratyphi A* IMVS1316 cloned in pUC9, and plasmid pPR507, which is pGB2 carrying the 9.57-kb fragment of *S. typhimurium* in the opposite orientation to that in pPR508, were used in a similar heteroduplex reaction, since pUC9 also has no homology with pGB2. Plasmids pPR617 and pPR507 were linearized with *Bam*HI and *Sal*I, respectively before being mixed in a heteroduplex reaction. The double-stranded region was measured and found to be about 1.4 ± 0.05 kb (average of eight measurements), showing that the region of homology extends from position 9.57 to about 8.17 on the *S. typhimurium* map (Fig. 7).

DISCUSSION

Southern hybridization and heteroduplex analysis demonstrated that the *rfb* gene clusters of *S. typhimurium* LT2, *S. typhi* Ty21a, and *S. paratyphi A* IMVS1316 share substantial homology. The LT2 regions from positions 0 to 5.8 and from 8.17 to 9.57 were shown to be homologous by heteroduplex analysis, and with the exception of one *Sac*I site, all restriction sites examined in the region from 9.57 (the first *Eco*RI

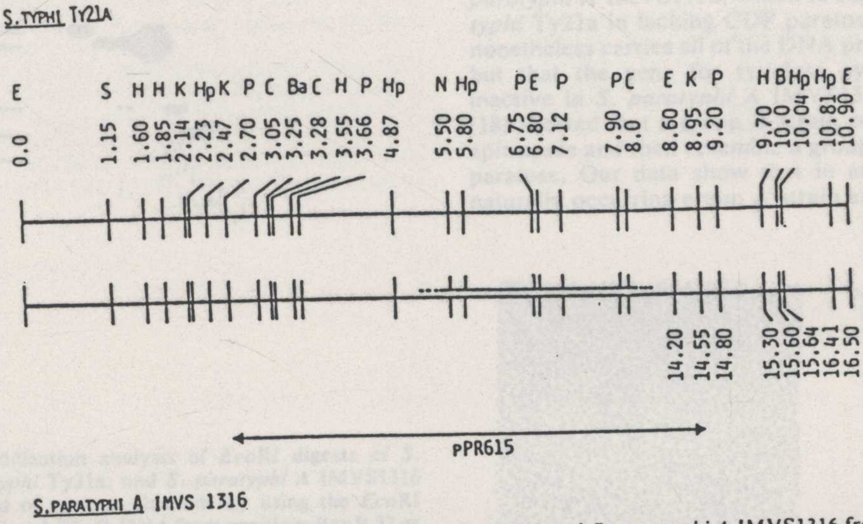


FIG. 3. Restriction map of *rfb* DNA of *S. typhi* Ty21a from positions 0 to 10.9 and *S. paratyphi A* IMVS1316 from positions 0 to 16.5. *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 have substantial similarity in their restriction maps. The *S. paratyphi A* IMVS1316 map is drawn such that sites present between positions 0 and 4.87 and positions 14.2 and 16.5 correspond to sites present between positions 0 and 4.87 and positions 8.6 and 10.9 on the *S. typhi* Ty21a map, respectively. The sites found to be triplicated in *S. paratyphi A* IMVS1316 (see text) are within a 2.8-kb region which includes the 2.5-kb segment from coordinates 5.5 to 8.0, indicated by double lines in the *S. paratyphi A* IMVS1316 map. Restriction enzyme sites: E, *Eco*RI; S, *Sac*I; H, *Hin*dIII; K, *Kpn*I; Hp, *Hpa*I; P, *Pst*I; C, *Cla*I; Ba, *Bam*HI; N, *Nru*I; B, *Bgl*II. Plasmid pPR615 is pUC19 carrying a 12.1-kb *Kpn*I fragment which extends from positions 2.47 to 14.55 on the *S. paratyphi A* IMVS1316 map.

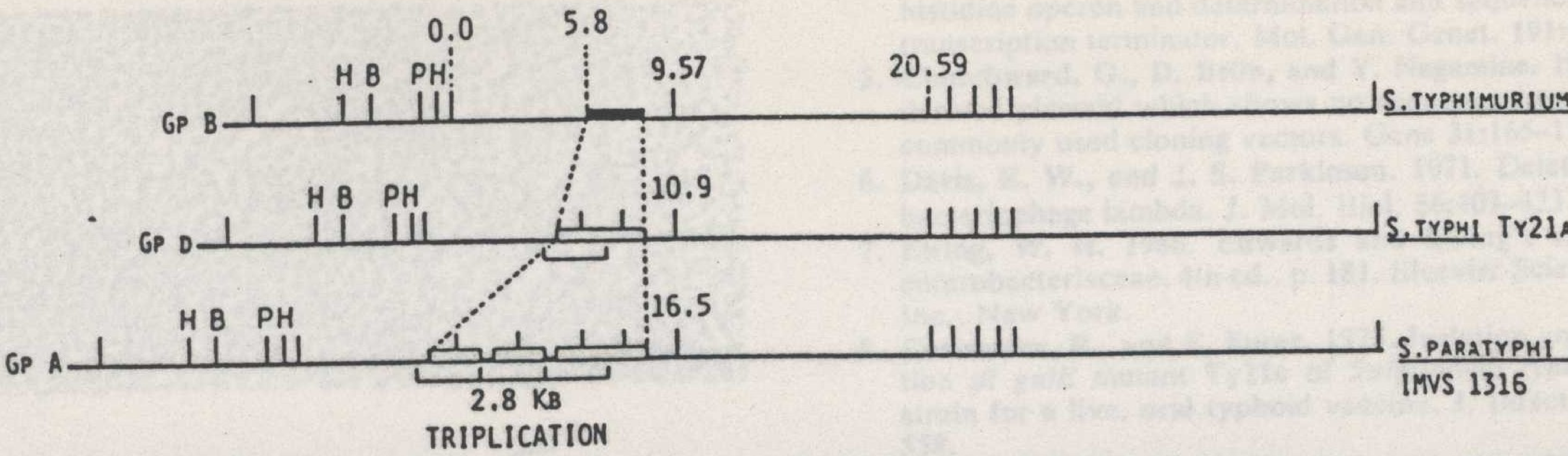


FIG. 4. Comparative map of *rfb* operons from *Salmonella* groups B, D, and A. *Eco*RI sites are indicated by vertical bars. H, B, and P are *Hpa*I, *Bgl*II, and *Pst*I recognition sites, respectively. Regions shown by restriction site homology and heteroduplex analysis to be homologous or unique are indicated as follows: —, common to all; ▬, *S. typhimurium* LT2 specific; ▭, common to *S. typhi* Ty21a and *S. paratyphi A* IMVS1316. Note that digestion at any site present only once in a triplicated interval produces a fragment of the same length as this interval and that the fragment is present at twice the molarity of other fragments generated by this enzyme.

site known after position 8.17) to 20.59 of *S. typhimurium* were conserved in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 DNA. However, it should be noted that although heteroduplex analysis has indicated homology in the region from positions 8.17 to 9.57, some restriction site nonhomology was detected. The remaining region of homology rightward of position 20.59 has not been examined in the same detail, but presumably the same high level of homology exists throughout this conserved region.

The DNA between positions 5.8 and 8.17 in *S. typhimurium* is replaced by a nonhomologous segment in *S. typhi* Ty21a and *S. paratyphi A* IMVS1316 (Fig. 4). The segment in *S. paratyphi A* IMSV1316 differs from that of *S. typhi* Ty21a only in having a 2.8-kb fragment triplicated, giving rise to 5.6 kb of additional DNA. The extent of the triplication was established by the presence of a 2.8-kb fragment at twice the molarity of other fragments in separate digests with four different enzymes. One end of the triplicated region must lie between the *Cla*I and *Nru*I sites at positions 8.0 and 8.6 in *S. paratyphi A* IMVS1316, and hence the other end must lie between positions 4.87 and 5.5. Thus, the 2.8-kb repeated region includes a short piece of DNA homologous with *S. typhimurium* LT2, accounting for the homology of the 2.8-kb *Eco*RI fragment with pPR301. Abequose is replaced by paratose and tyvelose in the O antigens of *S. paratyphi* and *S. typhi*, respectively. Other studies in our laboratory on *S. typhimurium* LT2 have shown that the genes for abequose synthesis (*rfbF*, *rfbG*, and *rfbH*) map in the general region of the structural difference between the three strains studied (2), and we are continuing our study of this variable region to precisely locate the genes involved in the synthesis of these three sugars.

At this stage, it seems reasonable to conclude that *S. paratyphi A* IMVS1316, which is expected to differ from *S. typhi* Ty21a in lacking CDP paratose-2-epimerase (15, 20), nonetheless carries all of the DNA present in *S. typhi* Ty21a, but that the gene for tyvelose synthesis is presumably inactive in *S. paratyphi A* IMVS1316. Sasaki and Uchida (18) showed that a group D strain could mutate to lose the epimerase and then resemble a group A strain by producing paratose. Our data show that in at least one instance, a naturally occurring group A strain arose in this way.

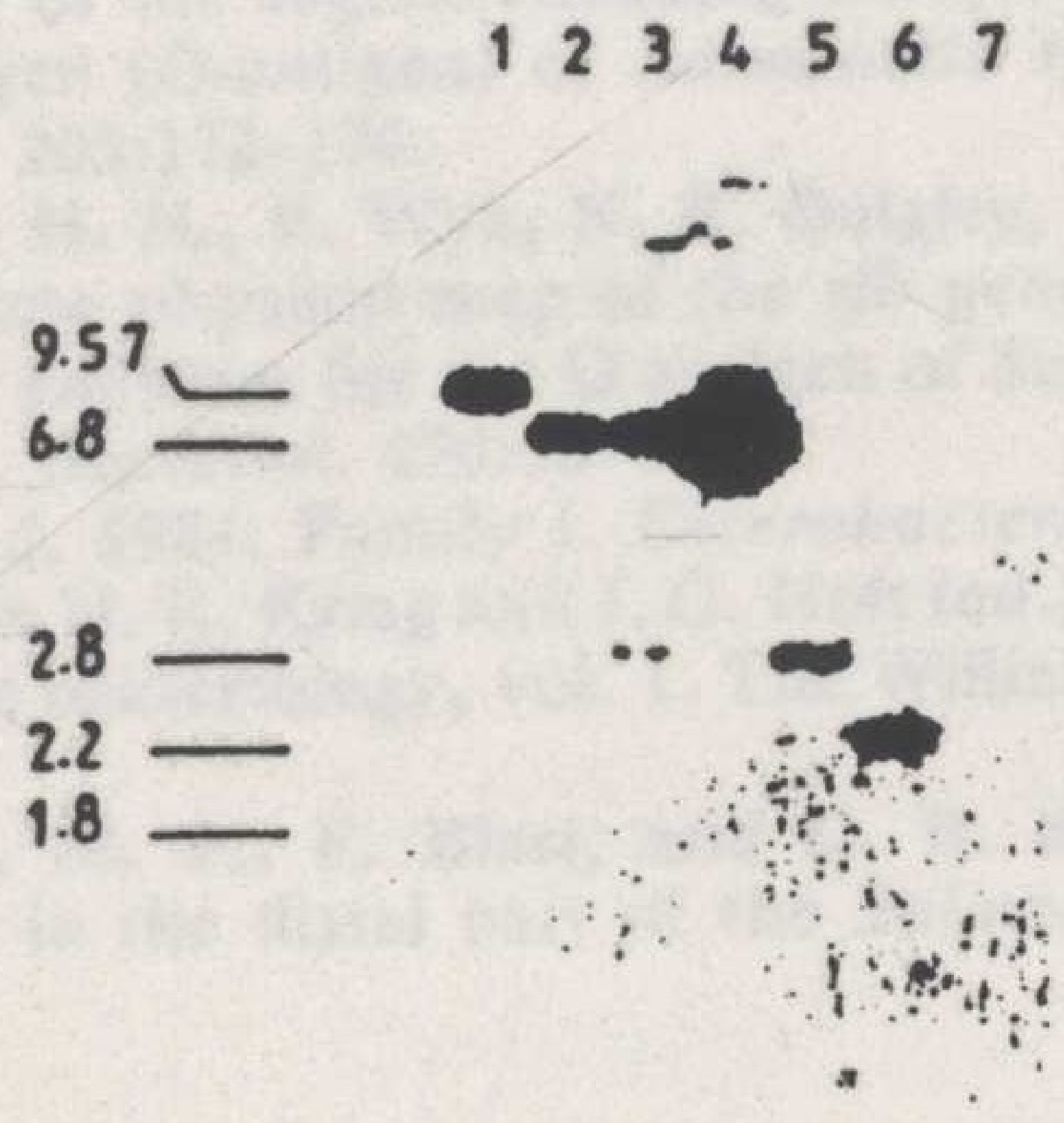


FIG. 5. Southern hybridization analysis of *Eco*RI digests of *S. typhimurium* LT2, *S. typhi* Ty21a, and *S. paratyphi A* IMVS1316 chromosomal DNA and of various plasmids by using the *Eco*RI fragment of *S. typhimurium* LT2 *rfb* DNA from positions 0 to 9.57 as a radioactive probe. Lanes: 1, *S. typhimurium* LT2 chromosomal DNA; 2, *S. typhi* Ty21a chromosomal DNA; 3, *S. paratyphi A* IMVS1316 chromosomal DNA; 4, pPR505; 5, pPR434; 6, pPR436; 7, pPR437. The *Eco*RI subclones in lanes 4 to 7 carry the 6.8-, 2.8-, 2.3-, and 1.8-kb *rfb* DNA fragments, respectively, from *S. paratyphi A* IMVS1316. The two bands of least intensity visible in lane 4 correspond to partial digestion products of pPR505.

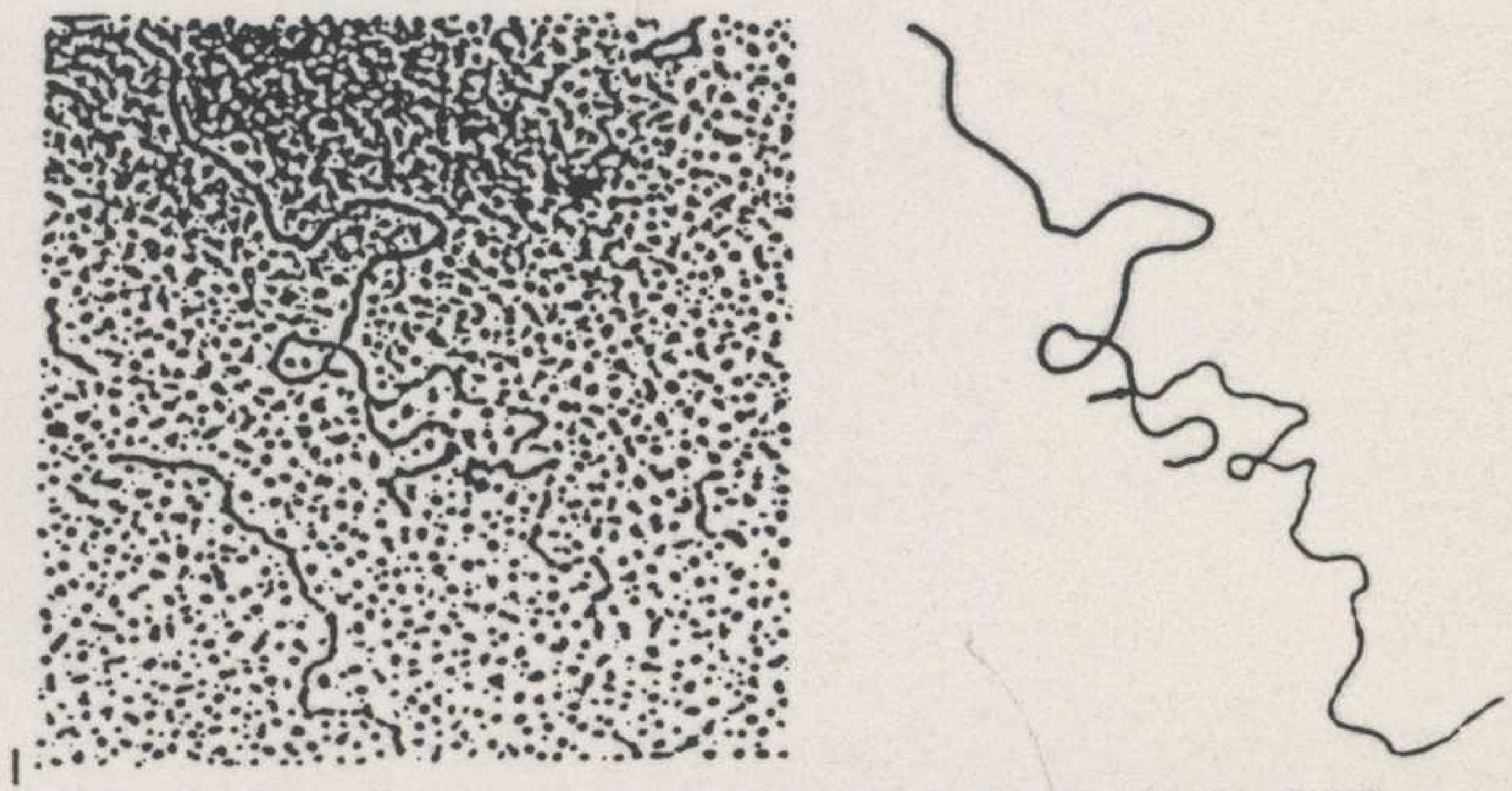


FIG. 6. Heteroduplex formed between a 6.8-kb *Eco*RI fragment from pPR505 (*S. paratyphi A* IMVS1316) and plasmid pPR508 (*S. typhimurium* LT2) linearized with *Sal*I. The arrow indicates the end of the region of homology.

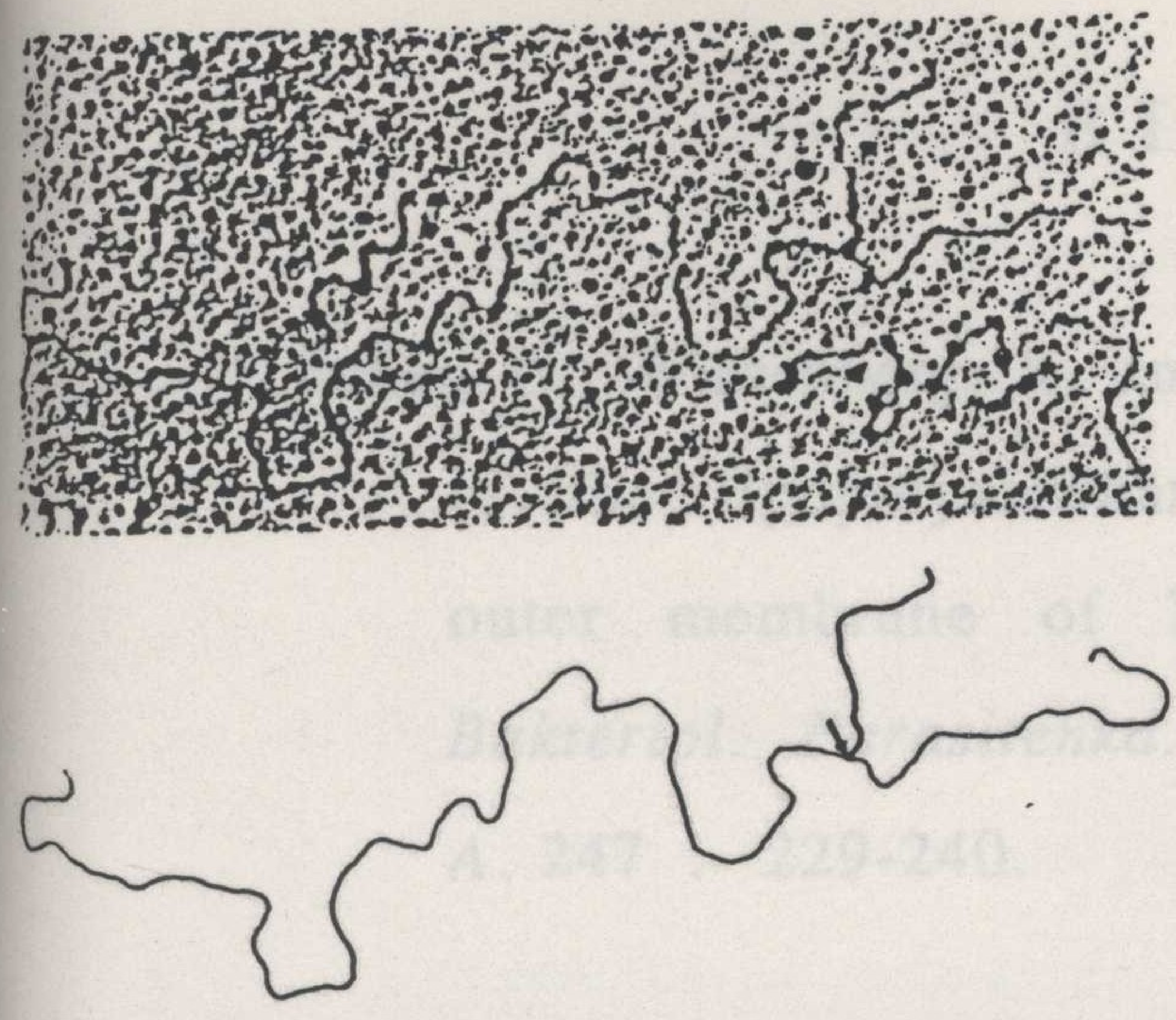


FIG. 7. Heteroduplex formed between plasmids pPR617 and pPR507 linearized with *Bam*HI and *Sal*I, respectively. The arrow indicates the end of the region of homology.

ACKNOWLEDGMENTS

We acknowledge the help of S. Dixon and C. Murray of the Institute of Medical and Veterinary Science, Adelaide, Australia, for providing *S. paratyphi* A strain IMVS1316 and antisera for strain confirmation. We are thankful to R. Czolij for his technical assistance in electron microscopy work.

LITERATURE CITED

1. **Brahmbhatt, H. N., N. B. Quigley, and P. R. Reeves.** 1986. Cloning part of the region encoding biosynthetic enzymes for surface antigen (O-antigen) of *Salmonella typhimurium*. Mol. Gen. Genet. 203:172–176.
2. **Brahmbhatt, H. N., P. Wyk, N. B. Quigley, and P. R. Reeves.** 1988. Complete physical map of the *rfb* gene cluster encoding biosynthesic enzymes for the O antigen of *Salmonella typhimurium* LT2. J. Bacteriol. 170:98–102.
3. **Brenner, D. J.** 1984. Family I. *Enterobacteriaceae* Rahn 1937, p. 408–516. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.
4. **Carlomagno, M. S., F. Blasi, and C. B. Bruni.** 1983. Gene organization in the distal part of the *Salmonella typhimurium* histidine operon and determination and sequence of the operon transcription terminator. Mol. Gen. Genet. 191:413–420.
5. **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.
6. **Davis, R. W., and J. S. Parkinson.** 1971. Deletion mutants of bacteriophage lambda. J. Mol. Biol. 56:403–423.
7. **Ewing, W. H.** 1986. Edwards and Ewing's identification of enterobacteriaceae, 4th ed., p. 181. Elsevier Science Publishing. Inc., New York.
8. **Germanier, R., and E. Furer.** 1975. Isolation and characterization of *galE* mutant Ty21a of *Salmonella typhi*: a candidate strain for a live, oral typhoid vaccine. J. Infect. Dis. 131:553–558.
9. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557–580.
10. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene 11:291–298.
11. **Jann, K., and B. Jann.** 1984. Structure and biosynthesis of O-antigens, p. 138–186. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Elsevier Science Publishers, Amsterdam.
12. **Kleinschmidt, A. K., and R. K. Zahn.** 1959. Uber Deoxyribonukleinsaure-Molekeln in Protein-Mischfilmen. Z. Naturforsch. 14b:770–779.
13. **Makela, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Elsevier Science Publishers, Amsterdam.
14. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Matsuhashi, S., and J. L. Strominger.** 1965. Reversible 2-epimerization of CDP-paratose and CDP-tyvelose. Biochem. Biophys. Res. Commun. 20:169–175.
16. **Poustka, A., H. Rackwitz, A. Frischauf, B. Hohn, and H. Lehrach.** 1984. Selective isolation of cosmid clones by homologous recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81:4129–4133.
17. **Sanderson, K. E., and J. R. Roth.** 1983. Linkage map of *Salmonella typhimurium*, edition VI. Microbiol. Rev. 47:410–453.
18. **Sasaki, T., and T. Uchida.** 1974. Mutants of group D1 *Salmonella* carrying the somatic antigen of group A organisms: evidence for the lack of cytidine diphosphate paratose-2-epimerase. J. Bacteriol. 117:13–18.
19. **Sternberg, N., D. Tiemeier, and L. Enquist.** 1977. In vitro packaging of a λ *dam* vector containing *Eco*R1 DNA fragments of *Escherichia coli* and phage P1. Gene 1:255–280.
20. **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.

BIBLIOGRAPHY

Acker, G., Wartenberg, K. and Knapp, W. 1980. Sugar composition of the lipopolysaccharide and ultrastructural study of the outer membrane of *Yersinia enterocolitica*. *Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt.* 1: *Orig. Reihe A*. **247** : 229-240.

Almon, L. and Stovall, W.D. 1939. The lack of one of the somatic antigens of typhoid cultures. *J. Bacteriol.* **38** : 419-430.

Ames, B.N., Lee, F.D. and Durston, W.E. 1973. An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc. Natl. Acad. Sci. U.S.A.* **70** : 782-786.

Ames, G.F., Spudick, E.N. and Nikaido, H. 1974. Protein composition of the outer membrane of *Salmonella typhimurium* : effect of lipopolysaccharide mutations. *J. Bacteriol.* **117** : 406-416.

Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. 1988. Codon usage tabulated from GenBank Genetic Sequence Data. *Nucl. Acids Res.* **16** (Supp) : r315- r391.

Barr, K. and Rick, P.D. 1987. Biosynthesis of enterobacterial common antigen in *Escherichia coli* : *in vitro* synthesis of lipid-linked intermediates. *J. Biol. Chem.* **262** : 7142-7150.

Batley, M., Packer, N.H. and Redmond, J.W. 1985a. Analytical studies of lipopolysaccharide and its derivatives from *Salmonella minnesota* R595. I. Phosphorus magnetic resonance spectra. *Biochem. Biophys. Acta* **821** : 179-194.

Batley, M., Packer, N.H. and Redmond, J.W. 1985b. Analytical studies of lipopolysaccharide and its derivatives from *Salmonella minnesota* R595. II. Proton and carbon magnetic resonance spectra. *Biochem. Biophys. Acta* **821** : 195-204.

Batley, M., McNicholas, P.A. and Redmond, J.W. 1985c. Analytical studies of lipopolysaccharide and its derivatives from *Salmonella minnesota* R595. III. Reappraisal of established methods. *Biochem. Biophys. Acta* **821** : 205-216.

Bayer, M.E. 1975. Role of adhesion zones in bacterial cell surface function and biogenesis, pp 393-427. *In* A. Tzagoloff (ed.), *Membrane Biogenesis* . Plenum Publ. Co., New York.

Bayer, M.E. 1979. The fusion sites between the outer membrane and cytoplasmic membrane of bacteria: their role in membrane assembly and virus infection. pp 167-202. *In* M. Inouye (ed.), *Bacterial Outer Membrane,* Biogenesis and Functions. John Wiley and Sons, New York.

Bayer, M.H., Costello, G.P. and Bayer, M.E. 1982. Isolation and partial characterization of membrane vesicles carrying markers of the membrane adhesion sites. *J. Bacteriol.* **149** : 758-767.

Beher, M., Pugsley, A. and Schnaitman, C. 1980. Correlation between the expression of an *Escherichia coli* cell surface protein and the ability of the protein to bind to the lipopolysaccharide. *J. Bacteriol.* **143** : 403-410.

Berst, M., Hellerqvist, C.G., Lindberg, B., Lüderitz, O., Svensson, S. and Westphal, O. 1969. Structural investigations on T1 lipopolysaccharides. *Eur. J. Biochem.* **11** : 353-359.

Berst, M., Lüderitz, O. and Westphal, O. 1971. Studies on the structure of T1 lipopolysaccharides. *Eur. J. Biochem.* **18** : 361-368.

Blache, D., Bruneteau, M. and Michel, G. 1980. Lipid A of lipopolysaccharaide from *Escherichia coli* K-12 CR34. *Biochemie* **62** : 191-194.

Böhlck, I. 1965. Ein Beitrag uber das Aufreten von Verlust Varianten bei Salmonellen. *Arch. Hyg. Bakteriol.* **149** : 185-187.

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

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

Bray, D. and Robbins, P.W. 1967a. The direction of chain growth in *Salmonella anatum* O-antigen biosynthesis. *Biochem. Biophys. Res. Commun.* **28** : 334-339.

Bray, D. and Robbins, P.W. 1967b. Mechanism of ε^{15} conversion studies with bacteriophage mutants. *J. Mol. Biol.* **30** : 457-475.

Bruneteau, M., Volk, W.A., Sing, P.P. and Lüderitz, O. 1974. Structural investigations on the *Salmonella* T2 polysaccharide. *Eur. J. Biochem.* **43** : 501-508.

Bucholtz, C.A. and Reisner, A.H. 1986. MBIS, an integrated system for the retrieval and analysis of sequence data for nucleic acids and proteins. *Nucl. Acids Res.* **14** : 233.

Busslinger, M., Rusconi, S. and Birustiel, M.L. 1982. An unusual evolutionary behaviour of a sea urchin histone gene cluster. *EMBO J.* **1** : 27-33.

Carlomagno MS, Blasi F, Bruni CB (1983) Gene organization in the distal part of the *Salmonella typhimurium* histidine operon and determination and sequence of the operon transcription terminator. Mol Gen Genet 191:413–420

Chatterjee, A.K., Ross, H. and Sanderson, K.E. 1976. Leakage of periplasmic enzymes from lipopolysaccharide defective mutants of *Salmonella typhimurium*. *Can. J. Microbiol.* **22** : 1549-1560.

Chou, P.Y. and Fasman, G.D. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47** : 45-148.

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

Clark, W.R., McLaughlin, J. and Webster, M.E. 1958. An aminohexuronic acid as the principal hydrolytic component of the Vi antigen. *J. Biol. Chem.* **230** : 81-90.

Clark-Curtiss, J.E. and Curtiss III, R. 1983. Analysis of recombinant DNA using *Escherichia coli* minicells. *Methods Enzymol.* **101** : 347-362.

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

Crosa, J.H., Brenner, D.J., Ewing, W.H. and Falkow, S. 1973. Molecular relationship among the *Salmonellae*. *J. Bacteriol.* **115** : 307-315.

Crowell, D.N., Anderson, M.S. and Raetz, C.R.H. 1986. Molecular cloning of the genes for lipid A disaccharide synthase and UDP-N-acetylglucosamine in *Escherichia coli*. *J. Bacteriol.* **168** : 152-159.

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

Danilov, L.L., Troitzky, M.F., Utkina, N.S., Shibaev, V.N. and Kochetkov, N.K. 1980. Synthesis of oligosaccharide phosphates: fragments of the biosynthesis intermediates of *Salmonella* O-specific polysaccharides. *Carbohydr. Res.* **87** : 141-146.

Danilov, L.L., Maltsev, S.G., Shibaev, V.M. and Kochetkov, N.K. 1981. Synthesis of polyprenyl pyrophosphate sugars from unprotected mono- and oligosaccharide phosphates. *Carbohydr. Res.* **88** : 203-211.

Davis, R.W. and Parkinson, J.W. 1971. Deletion mutants of bacteriophage lambda. *J. Mol. Biol.* **56** : 403-423.

Dayhoff, M.D., Schwartz, R.M. and Orcutt, B.C. 1978. A model of evolutionary changes in proteins. *Atlas of protein sequence and structure*. 5, Supp. 3, 345-352.

Dell, A., Oates, J., Lugowski, C., Romanowska, E., Kenne, L. and Lindberg, B. 1984. The enterobacterial common-antigen, a cyclic polysaccharide. *Carbohydr. Res.* **133** : 95-104.

Dmitriev, B.A., Backinowsky, L.V., L'vov, V.L., Kocketkov, N.K. and Hofman, I.L. 1975. Somatic antigen of *Shigella dysenteriae* type 3. Structural features of specific polysaccharide chain. *Eur. J. Biochem.* **59** : 539-547.

Dmitriev, B.A., Backinowsky, L.V., Knirel, Y.A., Kochetkov, N.K. and Hofman, I.L. 1977. Somatic antigens of *Shigella*. The structure of the specific polysaccharide chain of *Shigella dysenteriae* type 5 lipopolysaccharide. *Eur. J. Biochem.* **78** : 381-387.

Dmitriev, B.A., Knirel, Y.A., Kocharova, N.A. and Kochetkov, N.K. 1980. Somatic antigens of *Pseudomonas aeruginosa*. The structure of the polysaccharide chain of *Ps. aeruginosa* O-serogroup 7 (Lanyi) liposaccharide. *Eur. J. Biochem.* **106** : 643-651.

Drewry, D.T., Symes, K.C., Gray, G.W. and Wilkinson, S.G. 1975. Studies of polysaccharide fractions from the lipopolysaccharide of *Ps. aeruginosa* N.C.T.C. 1999. *Biochem. J.* **149** : 93-106.

Dugaiczyk, A., Boyer, H.W. and Goodman, H.M. 1975. Ligation of EcoRI endonuclease generated DNA fragments into linear and circular structures. *J. Mol. Biol.* **96** : 171-184.

Edwards, P.R. 1945. Form variation in group C *Salmonella* strains. *Proc. Soc. Exp. Biol. Med.* **59** : 49-52.

Edwards, P.R. and Ewing, W.H. 1966. *Identification of Enterobacteriaceae*. Burgess Publ. Co., Minneapolis, Minnesota.

Eklund, K., Garegg, P.J., Kenne, L., Lindberg, A.A. and Lindberg, B. 1978. Structural studies on the *Escherichia coli* 0111 lipopolysaccharide. *In* Abstracts of the IXth International Symposium of Carbohydrate Chemistry, London, No 493.

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

Ellwood, D.C. and Kirk, G.R.A. 1971. The occurrence of 6-deoxy-L-altrose in the lipopolysaccharide of *Yersinia enterocolitica*. *Biochem. J.* **122**: 14P.

Elsbach, P. and Weiss, J. 1983. A re-evalution of the roles of the O_2-dependent and O_2-independent microbial systems of phagocytosis. *Rev. Infect. Dis.* **5** : 843-853.

Ewing, W.H. 1986. Edwards and Ewing's identification of Enterobacteriaceae, 4th ed. Elsevier Science Publishers, New York.

Feber, D.M. and Brubaker, R.R. 1981. Plasmids in *Yersinia pestis*. *Infect. Immun.* **31** : 839-841.

Feige, U. and Stirm, S. 1976. On the structure of the *Escherichia coli* cell wall lipopolysaccharide core and on its ØfX174 receptor region. *Biochem. Biophys. Res. Commun.* 71 : 566-573.

Feige U., Jann, B., Jann, K., Schmidt, G. and Stirm, S. 1977. On the primary structure of the *Escherichia coli* R4 cell wall lipopolysaccharide core. *Biochem. Biophys. Res. Commun.* 79 : 88-95.

Formal, S.B., Gemski, Jr. P., Baron, L.S. and LaBrec, E.H. 1970. Genetic transfer of *Shigella flexneri* antigens to *Escherichia coli* K12. *Infect. Immun.* 1 : 279-287.

Fuller, N.A. and Staub, A.M. 1968. Immunochemical studies on *Salmonella*. 13. Chemical changes appearing on the specific polysaccharide of *S. cholerasius* (6_2,7) after its conversion by phage 14 (6,7). *Eur. J. Biochem.* 6 : 525-533.

Fuller, N.A., Etievant, M. and Staub, A.M. 1968. Immunochemical studies on *Salmonella*. 14. Oligosaccharides which determine the specificities 6_2, and 14 in *Salmonella cholerasius*, after conversion by bacteriophage 14 (6,7). *Eur. J. Biochem.* 6 : 525-533.

Funahara, Y. and Nikaido, H. 1980. Asymmetric localization of lipopolysaccharides on the outer membrane of *Salmonella typhimurium*. *J. Bacteriol.* 141 : 1463-1465.

Galanos, C., and Lüderitz, O. 1975. Electrodialysis of lipopolysaccharides and their conversion to uniform salt forms. *Eur. J. Biochem.* **54** : 603-610.

Galanos, C., Lüderitz, O., Rietschel, Eth. and Westphal, O. 1977. Newer aspects of the chemistry and biology of bacterial lipopolysaccharides with special reference to their lipid A component, pp 239-335. *In* T. Goodwin (ed.), *International Review of Biochemistry,* Biochemistry of Lipids II, vol. 14, University Park Press, Baltimore.

Gamian, A. and Romanowska, E. 1982. The core structure of *Shigella sonnei* lipopolysaccharide and the linkage between O-specific polysaccharide and the core region. *Eur. J. Biochem.* **129** : 105-109.

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

Gemski, P. Jr. and Stocker, B.A.D. 1967. Transduction by bacteriophage P22 in nonsmooth mutants of *Salmonella typhimurium. J. Bacteriol.* **93** : 1588-1597.

Gemski, P., Lazere, J.R. and Casey, T. 1980a. Plasmid associated with pathogenicity and calcium dependency of *Yersinia enterocolitica. Infect. Immun.* **27** : 682-685.

Gemski, P., Lazere, J.R., Casey, T. and Wohlhieter, J.A. 1980b. Presence of a virulence-associated plasmid in *Yersinia pseudotuberculosis*. *Infect. Immun.* **28** : 1044-1047.

Germanier, R. and Furer, E. 1975. Isolation and characterization of *galE* mutant Ty21a of *Salmonella typhi* : a candidate strain for a live, oral typhoid vaccine. *J. Infect. Dis.* **131** : 553-558.

Ghalambor, M.A. and Heath, E.C. 1966. The biosynthesis of cell wall lipopolysacchaide in *Escherichia coli*. IV. Purification and proportion of cytidine monophosphate 3-deoxy-D-manno-octulosonate synthetase. *J. Biol. Chem.* **241** : 3216-3221.

Ghuysen, J.B. and Shockman, G.D. 1973. Biosynthesis of peptidoglycan. *In* L. Leive (Ed.), *Bacterial membranes and walls*, pp. 37-130. Microbiology Series I. Dekker, New York.

Gmeiner, S., Lüderitz, O. and Westphal, O. 1969. Biochemical studies on lipopolysaccharides of *Salmonella* R mutants. 6. Investigation on the structure of the lipid A component. *Eur. J. Biochem.* **7** : 370-379.

Goad, W.B. and Kanehisa, M. 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. *Nucl. Acids Res.* **10** : 247-263.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. 1981. Translation initiation in prokaryotes. *Annu. Rev. Microbiol.* **35** : 365-403.

Goldman, R.C. and Strominger, J.L. 1972. Purification and properties of C55-isoprenyl-pyrophosphate phosphatase from *Micrococcus lysodeikticus. J. Biol. Chem.* **247** : 5116-5122.

Goldman, R.C. and Leive, L. 1980. Heterogeneity of antigenic-side-chain length in lipopolysaccharide from *Escherichia coli* 0111 and *Salmonella typhimurium* LT2. *Eur. J. Biochem.* **107** : 145-153.

Gonzalez-Porque, P. and Strominger, J.L. 1972. The biosynthesis of 3,6-dideoxy sugars : formation of CDP-4-keto-3,6-dideoxyglucose from CDP-glucose. *Methods Enzymol.* **18** : 461-470.

Gromska, W. and Mayer, H. 1976. The linkage of lysine in the O-specific chains of *Proteus mirabilis. Eur. J. Biochem.* **62** : 391-399.

Grunstein, M. and Hogness, D. 1975. Colony hybridisation: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. U.S.A.* **72** : 3961-3965.

Hagiwara, A., Uetake, H. and Takida, K. 1966. Lysogenic conversion in the structure of cellular receptor for *Salmonella* phage g341. *Annu. Rep. Inst. Virus Res.* (Kyoto Univ.) **9** : 126-129.

Hämmerling, G., Lüderitz, O. and Westphal, O. 1970. Structural investigations on the core polysaccharide of *Salmonella typhimurium* and the mode of attachment of the O-specific chains. *Eur. J. Biochem.* **15** : 48-56.

Hartman, P.E., Loper, J.C. and Serman, D. 1960. Fine structure mapping by complete transduction between histidine requiring *Salmonella* mutants. *J. Gen. Microbiol.* **22** : 323-353.

Hase, S. and Rietschel, E.Th. 1976. Isolation and analysis of the lipid A backbone. Lipid A structure of lipopolysaccharides from various bacterial groups. *Eur. J. Biochem.* **63** : 101-107.

Hawley, D.K. and McClure, W.R. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl. Acids Res.* **11** : 2237-2255.

Hayes, W. 1947. The nature of somatic phase variation and its importance in the serological standardization of O-suspensions of Salmonellas for use in the widal reaction. *J. Hyg.* **45** : 111-117.

Hellerqvist, C.G., Lindberg, B., Svensson, S., Holme, T. and Lindberg, A.A. 1969a. Structural studies on the O-specific side chains of the cell wall lipopolysaccharides from *Salmonella typhimurium* LT2. *Carbohydr. Res.* **9** : 237-241.

Hellerqvist, C.G., Lindberg, B., Svensson, S., Holme, T. and Lindberg, A.A. 1969b. Structural studies on the O-specific side-chains of the cell-wall lipopolysaccharide from *Salmonella typhi* and *S. enteritidis*. *Acta Chem. Scand.* **23** : 1588-1596.

Hellerqvist, C.G., Larm, O., Lindberg, B., Holme, T. and Lindberg, A.A. 1969c. Structural studies on the O-specific chain on the cell wall lipopolysaccharide from *Salmonella bredeney*. *Acta Chem. Scand.* **23** : 2217-2222.

Hellerqvist, C.G., Lindberg, B., Svensson, S., Holme, T. and Lindberg, A.A. 1970. Structural studies of the O-specific side chains of the cell wall lipopolysaccharides from *Salmonella newport* and *Salmonella kentuckey*. *Carbohydr. Res.* **14** : 17-26.

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

Hellerqvist, C.G., Lindberg, B., Lonngren, J. and Lindberg, A.A. 1971a. Structural studies on the O-specific side chain of the cell wall lipopolysaccharide from *Salmonella muenster*. *Carbohydr. Res.* **16** : 289-296.

Hellerqvist, C.G., Lindberg, B. and Lonngren, J. 1971b. Fragmentation analysis of *Salmonella newport* lipopolysaccharide. *Acta Chem. Scand.* **25** : 601-602.

Hellerqvist, C.G., Lindberg, B., Samuelson, K. and Lindberg, A.A. 1971c. Structural studies on the O-specific side chain of *Salmonella paratyphi A* var Durazzo. *Acta Chem. Scand.* **25** : 955-961.

Hellerqvist, C.G., Lindberg, B., Pilotti, A. and Lindberg, A.A. 1971d. Structural studies on the O-specific side chain of the cell wall lipopolysaccharide from *Salmonella senftenberg*. *Carbohydr. Res.* **16**: 297-302.

Hellerqvist, C.G., Ruden, U. and Mäkelä, P.H. 1972. The group C_2-type modification of the B-type lipopolysaccharide in a hybrid between *Salmonella* groups B and C_2. *Eur. J. Biochem.* **25** : 96-101.

Hemming, F.W. 1970. Polyprenols, pp 105-117. *In* T.W. Goodwin (ed.), *Biochemistry Society Symposia,* No. 29. Academic Press, London.

Hoffman, J. Lindberg, B. and Brubaker, R.R. 1980. Structural studies of the O-specific side-chains of the lipopolysaccharide from *Yersinia enterocolitica* Ye128. *Carbohydr. Res.* **78** : 212-214.

Honn, B. and Collins, J. 1980. A small cosmid for efficient cloning of large DNA fragments. *Gene* **11** : 291-298.

Hukari, R., Helander, I.M. and Vaara, M. 1986. Chain length heterogeneity of lipopolysaccharide released from *Salmonella typhimurium* by ethylenediaminetetraacetic acid or polycations. *Eur. J. Biochem.* **154** : 673-676.

Iseki, S. and Sakai, T. 1952. On a variant strain of *Salmonella* B group which lost V factor and Forssan antigen, caused by the action of anti-V factor serum. *Proc. Jpn. Acad.* **28** : 374-379.

Iseki, S. and Kashiwagi, K. 1955. Induction of somatic antigen I by bacteriophage in *Salmonella* B group. *Proc. Jpn. Acad.* **31** : 558-563.

Iseki, S. and Kashiwagi, K. 1957. Lysogenic conversion and transduction of genetic characters by temperate phage Iota in *Salmonella*. I. Lysogenic conversion with regard to somatic antigen 1 in *Salmonella* groups A, B, and D. *Proc. Jpn. Acad.* **33** : 481-485.

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

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

Iwashita, S. and Kanegasaki, S. 1973. Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. *Biochem. Biophys. Res. Commun.* **55** : 403-409.

Jann, B., Jann, K. and Beyaert, G.O. 1973. 2-amino-2,6-dideoxy-D-glucose (D-quinovosamine): A constituent of the lipopolysaccharides of *Vibrio cholerae*. *Eur. J. Bioche.* **37** : 531-534.

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

Jann, K. and Westphal, O. 1975. Microbial polysaccharides. *In The Antigens,* vol. 3, pp. 1-125. Academic Press, New York-San Francisco-London.

Jann, B., Prehm, P. and Jann, K. 1978. *Citrobacter* O antigens: structure of the O antigenic polysaccharide from *Citrobacter* sp. 396. *J. Bacteriol.* **134** : 462-469.

Jann, K. and Jann, B. 1983. The K antigens of *Escherichia coli*, pp 53-79. *In* K. Ishizeka, P. Kallos, B.H. Waksman and A.L. De Weck (eds.), *Progress Allergy,* vol. 33. Karger, Basel.

Jann, K. and Jann, B. 1984. Structure and biosynthesis of O-antigens, pp 138-186. *In* E.T. Rietschel (ed.), *Handbook of Endotoxin,* vol. 1. Elselvier Science Publishers, Amsterdam.

Jansson, P.E., Lindberg, B., Lindberg, A.A. and Wollin, R. 1979. Structural studies on the hexose region of the *Enterobacteriaceae* Type R3 core polysaccharide. *Carbohydr. Res.* **68** : 385-389.

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

Johnson, E.M., Krauskopf, B. and Baron, L.S. 1965. Genetic mapping of Vi and somatic antigenic determinants in *Salmonella. J. Bacteriol.* **40** : 302-308.

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

Jones, R.T., Koeltzow, D.E., Stocker, B.A.D. 1972. Genetic transfer of *Salmonella typhimurium* and *Escherichia coli* lipopolysaccharide antigens to *Escherichia coli* K12. *J. Bacteriol.* **111** : 758-770.

Jousimies, H. and Mäkelä, P.H. 1974. Genetic analysis of *Salmonella minnesota* R mutants with defects in the biosynthesis of the lipopolysaccharide core. *J. Bacteriol.* **119** : 753-759.

Kadam, S.K., Rehemtulla, A. and Sanderson, K.E. 1985. Cloning of *rfaG, B, I,* and *J* genes for glycosyltransferase enzymes for synthesis of the lipopolysaccharide core of *Salmonella typhimurium*. *J. Bacteriol.* **161** : 277-284.

Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remault, E. and Helskinini, D.R. 1979. Plasmid cloning vehicles derived from plasmids ColE1, R6K and RK2. *Methods Enzymol.* **68** : 268.

Kamio, Y. and Nikaido, H. 1976. Outer membrane of *Salmonella typhimurium* : accessibility of phospholipid head groups to phospholipase C and cyanogen bromide activated dextran in the external medium. *Biochemistry.* **15** : 2561-2570.

Kanegasaki, S. and Wright, A. 1970. Mechanism of polymerization of the *Salmonella* O antigen: utilization of lipid-linked intermediates. *Proc. Natl. Acad. Sci. U.S.A.* **67** : 951-958.

Kanehisa, M.I. 1982. Los Alamos sequence analysis package for nucleic acids and proteins. *Nucl. Acids Res.* **10** : 183-196.

Kanehisa, M.I., Klein, P., Greif, P. and DeLisi, C. 1984. Computer analysis and structure prediction of nucleic acids and proteins. *Nucl. Acids Res.* **12** : 417-428.

Katzenellenbogen, E. and Romanowska, E. 1980. Structural studies on *Shigella flexneri* serotype 6 core region. *Eur. J. Biochem.* **113** : 205-211.

Kauffmann, F. 1935. Uber serologische und kulturelle Varianten der Paratyphus and Mausetphys-Bacillen. *Z. Hyg. Infektionskr.* **116** : 368-384.

Kauffmann, F. 1937. *Salmonella* -Probeleme. 2. Hyg. 120 : 177-197.

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

Kauffmann, F. 1956. A new antigen of *Salmonella paratyphi B* and *Salmonella typhimurium. Acta. Pathol. Microbiol. Scand.* **39** : 299-304.

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

Kauffmann, F. 1966. *The Bacteriology of Enterobacteriaceae.* Ejnar Munksgaard, Copenhagen: Williams and Wilkins, Baltimore.

Kenne, L., Lindberg, B. and Unger, P. 1978. Structural studies of the *Vibrio cholerae* O-antigen. *Carbohydr. Res.* **68** : 614-616.

Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E., Romanowska, E. 1980. Structural studies of the O-specific side-chains of *Shigella sonnei* phase I lipopolysaccharide. *Carbohydr. Res.* **78** : 119-126.

Kenne, L. and Lindberg, B. 1983. Bacterial polysaccharides. *In* G.O. Aspinall (ed.), *The Polysaccharides,* vol. 2. Academic press, London-New York.

Kenne, L., Lindberg, B., Soderholm, E., Bundle, D.R. and Griffith, D.W. 1983. Structural studies of the O antigens from *Salmonella greenside* and *Salmonella adelaide. Carbohydr. Res.* **111** : 289-296.

Kiefer, W., Schmidt, G., Jann, B., and Jann, K. 1976. Genetic transfer of Salmonella O antigens to *Escherichia coli* 08. *J. Gen. Microbiol.* **92** : 311-324.

Kishi, K. and Iseki, S. 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 Iseki, S. 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.

Kishi, K. and Iseki, S. 1974. Genetic analysis of the O antigens in *Salmonella*. IV. Inheritance of O antigens 1, 3, 19 and 34 of *Salmonella* group E. *Jpn. J. Genet.* **49** : 125-129.

Kishi, K. and Iseki, S. 1975. Genetic recombination between *Escherichia coli* K12 and *E. coli* strains with blood group-active antigens. *J. Immunogenet.* **2** : 205-210.

Kleinschmidt, A.K. and Zahn, R.K. 1959. Uber deoxyribonukleinsaure-molekeln in protein-mischfilmen. *Z. Naturfosrch.* **14b** : 770-779.

Kochibe, N. 1970. Immunochemical studies on antigenic substances produced by intergroup-hybrids of *Salmonella*. *J. Biochem.* (Tokyo). **68** : 81-89.

Kopecko, D.J., Washington, O. and Formal, S.B. 1980. Genetic and physical evidence for plasmid control of *Shigella sonnei* form I cell surface antigen. *Infect. Immun.* **29** : 207-214.

Koplow, J. and Goldfine, H. 1974. Alterations in the outer membrane of the cell envelope of heptose-deficient mutants of *Escherichia coli*. *J. Bacteriol.* **117** : 527-543.

Kotelko, K., Gromska, W., Papierz, M., Szer, K., Krajewska, D. and Sidorczyk, Z. 1974. The constitution of 'core' in *Proteus* lipopolysaccharides. *J. Hyg. Epidemiol. Microbiol. Immunol.* **18** : 405-410.

Kozak, M. 1983. Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles. *Microbiol. Rev.* **47** : 1-45.

Kuhn, H.M., Neter, E. and Mayer, H. 1983. Modification of the lipid moiety of the enterobacterial common antigen by the "Pseudomonas factor". *Infect. Immun.* **40** : 696-700.

Kuhn, H.M., Meier, U. and Mayer, H. 1984. ECA, das gemeinsame Antigen der Enterobacteriaceae-Stiefkind der Mikrobiologie. *Forum Microbiol.* **7** : 274-285.

Kutsukake, K. and Iino, T. 1980. A trans-acting factor mediates inversion of a specific DNA segment in flagellar phase variation of *Salmonella*. *Nature* (London) **284** : 479-481.

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

Labischinski, H., Barnickel, G., Bradaczek, H., Naumann, D., Rietschel, E.T. and Giesbrecht, P. 1985. High state of order of isolated bacterial lipopolysaccharide and its possible contribution to the permeation barrier property of the outer membrane. *J. Bacteriol.* **162** : 9-20.

Lederberg. E.M. and Cohen, S.N. 1974. Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. *J. Bacteriol.* **119**:1072-1074.

Leinonen, M. 1985. Serological methods for the study of bacterial surface antigens, pp179-206. *In* Korhonen, T.K., Dawes, E.A. and Mäkelä, P.H. (eds), *Enterobacterial surface antigens: methods for molecular characterization.* Elsevier Publishing Company, Amsterdam-New York-Oxford.

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.

Lemaire, H. and Müller-Hill, B. 1986. Nucleotide sequences of the *galE* gene and the *galT* gene of *E. coli*. *Nucl. Acids Res.* **14** : 7705-7711.

Le Minor, L., Popoff, M.Y., Laurent, B. and Hermant, D. 1986. Individualisation d'une septieme sous-espece de *Salmonella* : *S. cholerasuis* subsp. *indica* subsp. nov. *Ann. Microbiol.* (Paris) **137b** : 211-217.

Le Minor, L. and Popoff, M.Y. 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.

Levine, P. and Frisch, A.W. 1936. On differentiation in the suipestifer group and on resistance to phage. *J. Immunol.* **30** : 63-88.

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

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

Lieberman, M.M., Shaparis, A. and Markovitz, A. 1970. Control of uridine diphosphate-glucose dehydrogenase synthesis and uridine diphosphate-glucuronic acid accumulation by a regulator gene mutation in *Escherichia coli* K12. *J. Bacteriol.* **101** : 959-964.

Lindberg, A.A. 1973. Bacteriophage receptors. *Ann. Rev. Microbiol.* **27** : 205-237.

Lindberg, A.A. 1977. Bacterial surface carbohydrates and bacteriophage adsorption, pp 289-356. *In* I. Sutherland (ed.), *Surface Carbohydrates of the Prokaryotic Cell*. Academic Press, London-New York.

Lindberg, A.A., Hellerqvist, C.G., Bagdian-Motta, G. and Mäkelä, P.H. 1978. Lipopolysaccharide modification accompanying antigenic conversion by phage P27. *J. Gen. Microbiol.* **107** : 279-287.

Lindberg, B., Lindh, F., Lonngren, J., Lindberg, A.A. and Svenson, S.B. 1981. Structural studies of the O-specific side chain of the lipopolysaccharide from *Escherichia coli* O55. *Carbohydr. Res.* **97** : 105-112.

Lindsay, S.S., Wheeler, B., Sanderson, K.E. and Costerton, J.W. 1972. The release of alkaline phosphatase and of lipopolysaccharide during the growth of rough and smooth strains of *Salmonella typhimurium*. *Can. J. Microbiol.* **19** : 335-343.

Lipman, D.J. and Pearson, W.R. 1985. Rapid and sensitive protein similarity searches. *Science* **227** : 1435-1441.

Losick, R. and Robbins, P.W. 1967. Mechanism of 15 conversion studies with a bacterial mutant. *J. Mol. Biol.* **30** : 445-456.

Lüderitz, O., Galanos, C., Risse, H.J., Ruschmann, E., Schlecht, S., Schmidt, G., Schulte-Holthausen, H., Wheat, R., Westphal, O. and Schlosshardt, J. 1966a. Structural relationships of *Salmonella* O and R antigens. *Ann. N.Y. Acad. Sci.* **133** : 349-374.

Lüderitz, O., Staub, A.M. and Westphal, O. 1966b. Immunochemistry of O and R antigens of *Salmonella* and related *Enterobacteriaceae*. *Bacteriol. Rev.* **30** : 192-255.

Lüderitz, O., Jann, K. and Wheat, R. 1968. Somatic and capsular antigens of Gram-negative bacteria, pp 105-227. *In* M. Florkin and E.H. Stotz (eds.), *Comprehensive Biochemistry, Extracellular and Supporting Structures*, Vol. 26a. Elselvier Publishing Company, Amsterdam.

Lüderitz, O., Westphal, O., Staub, A.M. and Nikaido, H. 1971. Isolation and chemical and immunological characterization of bacterial lipopolysaccharides, pp 145-233. *In* G. Weinbaum, S. Kadis and S.J. ajl (eds.), *Microbial Toxins*, vol. 4. Academic Press, New York.

Lüderitz, O., Galanos, C., Lehmann, V., Mayer, H., Rietschel, Eth. and Weckesser, J. 1978. Chemical structure and biological activities of lipid A's from various bacterial families. *Naturwissenschaften* **65** : 578-585.

Lüderitz, O., Freudenberg, M.A., Galanos, C., Lehmann, V., Rietschel, Eth. and Shaw, D.H. 1982. Lipopolysaccharides of gram-negative bacteria, pp 79-151. *In* S. Razin and S. Rottem (eds.), *Current Topics in Membranes and Transport, Membrane Lipids of Procaryotes,* vol. 17. Academic Press, New York.

Lugowski, C., Romanowska, E., Kenne, L. and Lindberg, B. 1983. Identification of a trisaccharide repeating-unit in the enterobacterial common antigen. *Carbohydr. Res.* **118** : 173-181.

Lugtenberg, B., Meijers, J., Peters, R., van der Hoek, P. and van Alphen, L. 1975. Electrophoretic resolution of the major outer membrane protein of *Escherichia coli* K12 into four bands. *FEBS Lett.* **58** : 254-258.

Lugtenberg, B., Peters, R., Bernheimer, H. and Berendsen, W. 1976. Influence of cultural conditions and mutations on the compositions of the outer membrane proteins of *Escherichia coli. Mol. Gen. Genet.* **147** : 251-262.

Lugtenberg, B. and van Alphen, L. 1983. Molecular architecture and funtioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria. *Biochem. Biophys. Acta* **737** : 51-115.

Mäkelä, P.H. 1965. Inheritance of the O antigens of *Salmonella* groups B and D. *J. Gen. Microbiol.* **41** : 57-65.

Mäkelä, P.H. 1966. Genetic determination of the O antigens of *Salmonella* groups B (4,5,12) and C_1 (6,7). *J. Bacteriol.* **91** : 1115-1125.

Mäkelä, P.H. and Mäkelä, O. 1966. *Salmonella* antigen 12_2: genetics of form variation. *Ann. Med. Exp. Biol. Fenn.* **44** : 310-317.

Mäkelä, P.H., Jahkola, M. and Lüderitz, O. 1970. A new gene cluster *rfe* concerned with the biosynthesis of *Salmonella* lipopolysaccharide. *J. Gen. Microbiol.* **60** : 91-106.

Mäkelä, P.H. 1973. Glucosylation of lipopolysaccharide in *Salmonella*: mutants negative for O antigen factor 12_2. *J. Bacteriol.* **116** : 847-856.

Mäkelä, P.H. and Mayer, H. 1974. Participation of lipopolysaccharide genes in the determination of the enterobacterial common antigen: analysis in *Salmonella* groups B and C. *J. Bacteriol.* **119** : 765-770.

Mäkelä, P.H. and Mayer, H. 1976. Enterobacterial common antigen. *Bacteriol. Rev.* **40** : 591-632.

Mäkelä, P.H., Schmidt, G., Mayer, H., Nikaido, H., Whang, H.Y. and Neter, E. 1976. Enterobacterial common antigen in *rfb* deletion mutants of *Salmonella typhimurium*. *J. Bacteriol.* **127** : 1141-1149.

Mäkelä, P.H. and Stocker, B.A.D. 1984. Genetics of lipopolysaccharide, pp 59-137. *In* E.T. Rietschel (ed.), *Handbook of endotoxin,* vol. 1. Elselvier Science Publishers, Amsterdam.

Mäkelä, P.H. and Rhen, M. 1985. Antigenic variation in enteric bacteria, pp 75-92. *In* T.K. Korhonen, E.A. Dawes and P.H. Mäkelä (eds.), *Enterobacterial surface antigens: methods for molecular characterization.* Elsevier Publishing Company, Amsterdam-New York-Oxford.

Maniatis, R., Fritsch, E.F. and Sambrook, J. 1982. *Molecular cloning. A laboratory manual.* Cold Spring Harbour Laboratory, New York.

Männel, D. and Mayer, H. 1978. Isolation and chemical characterization of the enterobacterial common antigen. *Eur. J. Biochem.* **86** : 361-370.

Markovitz, A., Lieberman, M.M. and Rosenbaum, N. 1967a. Derepression of phosphomannose isomerase by regulator gene mutations involved in capsular polysaccharide synthesis in *Escherichia coli* K-12. *J. Bacteriol.* **94** : 1497-1501.

Markovitz, A., Sydiskis, R.J. and Lieberman, M.M. 1967b. Genetic and biochemical studies on mannose-negative mutants that are deficient in phosphomannose isomerase in *Escherichia coli* K12. *J. Bacteriol.* **94** : 1492-1496.

Markovitz, A. 1977. Genetics and regulation of bacterial capsular polysaccharide synthesis and radiation sensitivity, pp 415-462. *In* I. Sutherland (ed.), *Surface carbohydrates of the procaryotic cell.* Academic Press. Inc., London. pp. 415-462.

Matsuhashi, S. and Strominger, J.L. 1964. Thymidine diphosphate 4-acetamido-4,6-dideoxyhexsoses. I. Enzymatic synthesis by strains of Escherichia coli. *J. Biol. Chem.* **239** : 2454-2463.

Matsuhashi, S. and Strominger, J.L. 1965. Reversible 2-epimerization of CDP-paratose and CDP-tyvelose. *Biochem. Biophys. Res. Commun.* **20** : 169-175.

Matsuhashi, S. 1966. Enzymatic synthesis of cytidine diphophate 3,6-dideoxyhexoses. *J. Biol. Chem.* **241** : 4275-4287.

Matsuhashi, S. and Strominger, J.L. 1967. Enzymatic synthesis of cytidine diphosphate 3,6-dideoxyhexoses. IV. Partial resolution of the enzymes required for reduction of cytidine diphosphate 4-keto-6-deoxy-D-glucose and isolation of a new intermediate. *J. Biol. Chem.* **242** : 3494-3500.

Mayer, H. and Schmidt, G. 1979. Chemistry and biology of the enterobacterial common antigen (ECA). *Curr. Top. Microbiol. Immunol.* **85** : 99-153.

Meier, U. and Mayer, H. 1985. Genetic location of genes encoding Enterobacterial common antigen. *J. Bacteriol.* **163** : 756-762.

Milkman, R. and Crawford, I.P. 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*. *Science* **221** : 378-380.

Mühlradt, P.F., Menzel, J., Golecki, J.R. and Speth, V. 1973. Outer membrane of *Salmonella* sites of export of newly synthesised lipopolysaccharide on the bacterial surface. *Eur. J. Biochem.* **35** : 471-481.

Mühlradt, P.F. and Golecki, J.R. 1975. Asymmetrical distribution and artificial reorientation of lipopolysaccharide in the outer membrane bilayer of *Salmonella typhimurium*. *Eur. J. Biochem.* **51** : 343-352.

Mulford, C.A. and Osborn, M.J. 1983. An intermediate step in translocation of lipopolysaccharide to the outer membrane of *Salmonella typhimurium*. *Proc. Natl. Acad. Sci. U.S.A.* **80** : 1159-1163.

Muto, A. and Osawa, S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. U.S.A.* **84** : 166-169.

Naide, Y., Nikaido, H., Mäkelä, P.H., Wilkinson, R.G. and Stocker, B.A.D. 1965. Semirough strains of *Salmonella*. *Proc. Natl. Acad. Sci. U.S.A.* **53** : 147-153.

Nakae, T. 1971. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from *Salmonella typhimurium*. III. Interconversion between various forms. *J. Biol. Chem.* **246** : 4404-4411.

Nakae, T. and Nikaido, H. 1971a. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from *Salmonella typhimurium*. I. Catalytic properties of various forms. *J. Biol. Chem.* **246** : 4386-4396.

Nakae, T. and Nikaido, H. 1971b. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from *Salmonella typhimurium*. II. Genetic determination of multiple forms. *J. Biol. Chem.* **246** : 4397-4403.

Nakamura, K., Pirtle, R.M. and Inouye, M. 1979. Homology of the gene coding for outer membrane lipoprotein within various Gram-negative bacteria. *J. Bacteriol.* **137** : 595-604.

Needleman, S.B. and Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48** : 443-453.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nikaido, H. 1965. Biosynthesis of cell wall polysaccharides in mutant strains of *Salmonella*. Transfer of L-rhamnose and D-galactose. *Biochemistry* **4** : 1550-1562.

Nikaido, H., Nikaido, K. and Rapin, A.M. 1965. Biosynthesis of thymidine diphosphate L-rhamnose in *Escherichia coli* K12. *Biochim. Biophys. Acta* **111** : 548-551.

Nikaido, H., Nikaido, K. and Mäkelä, P.H. 1966. Genetic determination of enzymes synthesising O-specific sugars of *Salmonella* lipopolysaccharides. *J. Bacteriol.* **91** : 1126-1135.

Nikaido, H., Levinthal, M., Nikaido, K. and Nakane, K. 1967. Extended deletions in the histidine-rough-B region of the *Salmonella* chromosome. *Proc. Natl. Acad. Sci. U.S.A.* **57** : 1825-1832.

Nikaido, H. 1970. Structure of cell wall lipopolysaccharide from *Salmonella typhimurium*. Further studies on the linkage between O side chains and R core. *Eur. J. Biochem.* **15** : 57-62.

Nikaido, K. and Nikaido, H. 1971. Glucosylation of lipopolysaccharide in *Salmonella*: biosynthesis of O antigen factor 12_2. II. *J. Biol. Chem.* **246** : 3912-3919.

Nikaido, H., Nikaido, K., Nakae, T. and Mäkelä, P.H. 1971. Glucosylation of lipopolysaccharide in *Salmonella* : biosynthesis of O antigen factor 12_2. I. *J. Biol. Chem.* **246** : 3902-3911.

Nikaido, H. 1973. Biosynthesis and assembly of lipopolysaccharides and the outer membrane layer of gram-negative cell wall, pp131-208. *In* Leive, L. (ed), Bacterial membranes and walls. Marcel Dekker, Inc., New York.

Nikaido, H. 1979 Die Permeabilitat der auberen Bakterienmembran. *Angew. Chem.* **91** : 394-407.

Nikaido, H. and Nakae, T. 1979. The outer membrane of gram-negative bacteria. *Adv. Microb. Physiol.* **20** : 163-250.

Nikaido, H. and Vaara, M. 1985. Molecular basis of bacterial outer membrane permability. *Microbiol. Rev.* **49** : 1-32.

Nimmich, W. and Korten, G. 1970. Die chemische zusammensetzund der *Klebsiella* lipopolysaccharide (O-antigens). *Pathol. Microbiol.* **36** : 179-190.

Nomore, W.M. 1973. Guanine-plus-cytosine (GC) composition of the DNA of bacteria, fungi, algae and protozoa, pp585-740. *In* Laskin, A.I. and Lechevalier, H.A. (eds.), *Microbial Composition* (Handbook of Microbiology Vol II). CRC Press Cleveland Ohio USA.

Norrander, J., Kempe, T. and Messing, J. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* **26** : 101-106.

Nyman, K., Plosila, M., Howden, L., Mäkelä, P.H. 1979. Genetic determination of lipopolysaccharide: locus of O-specific unit polymerase in group E of *Salmonella*. *Zbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. I: Orig. Reihe A.* **243** : 355-362.

Ochman, H. and Wilson, A.C. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26** : 74-86.

Ørskov, F. and Ørskov, I. 1962. Behaviour of *E. coli* antigens in sexual recombination. *Acta Pathol. Microbiol. Scand.* **55** : 99-109.

Ørskov, F., Ørskov, I., Jann, B. and Jann, K. 1971. Immunoelectrophoretic patterns of extracts from all *Escherichia coli* O and K antigen test strains. Correlation with pathogenicity. *Acta Pathol. Microbiol. Scand. Sect. B* **79** : 142-152.

Ørskov, F. and Ørskov, I. 1972. Immunoelectrophoretic patterns of extracts from *Escherichia coli* O antigens test strains 01 to 0157. Examinations in homologous O and OK sera. *Acta. Pathol. Microbiol. Scand. Sect. B* **80** : 905-910.

Ørskov, I., Ørskov, F., Jann, B. and Jann, K. 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol. Rev.* **41** : 667-710.

Osborn, M.J. 1966. Biosynthesis and structure of the core region of the lipopolysaccharide in *Salmonella typhimurium*. *Ann. N.Y. Acad. Sci.* **133** : 375-383.

Osborn, M.J. and Tze-Yuen, R.Y. 1968. Biosynthesis of bacterial lipolysaccharide. 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 Weiner, I.M. 1968. Biosynthesis of a bacterial lipopolysaccharide. I. Enzymatic incorporation of galactose in a mutant strain of *Salmonella*. *Proc. Natl. Acad. Sci. U.S.A.* **48** : 1831-1838.

Osborn, M.J. and Rothfield, L.I. 1971. Biosynthesis of the core region of lipopolysaccharides. *In* Weinbaum, G., Kadis, S. and Ajl, S.J. (eds), *Microbiol. Toxins*, vol 4, pp311-350. Academic Press, New York.

Osborn, M.J. 1979. Biosynthesis and assembly of lipopolysaccharide of the outer membrane, pp 15-34. *In* M. Inouye (ed.), *Bacterial Outer Membranes: Biogenesis and Functions*. John Wiley and Sons, New York.

Palva, E.T. and Mäkelä, P.H. 1980. Lipopolysaccharide heterogeneity in *Salmonella typhimurium* analysed by sodium dodecyl sulphate/polyacrylamide gel electrophoresis. *Eur. J. Biochem.* **107** : 137-143.

Pape, H. and Strominger, J.L. 1969. Enzymatic synthesis of cytidine phosphate 3,6-dideoxyhexoses. *J. Biol. Chem.* **244** : 3598-3604.

Petrovskaya, V.G. and Licheva, T.A. 1982. A provisional chromosome map of *Shigella* and the regions related to pathogenicity. *Acta Microbiol. Acad. Sci. Hung.* **29** : 41-53.

Plosila, M. and Mäkelä, P.H. 1972. Mapping of gene *oafC* determining antigen 1 in *Salmonella* of group E. *Scand. J. Clin. Lab. Invest.* **29** (suppl. 122) : 55.

Poustka, A., Rackwitz, H., Frischauf, A., Hohn, B. and Lehrach, H. 1984. Selective isolation of cosmid clones by homologous recombination in *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* **81** : 4129-4133.

Prehm, P., Stirm, S., Jann, B., Jann, K. and Boman, H.G. 1976a. Cell-wall lipopolysaccharides of ampicillin-resistant mutants of *Escherichia coli* K12. *Eur. J. Biochem.* **66** : 369-377.

Prehm, P., Schmidt, G., Jann, B. and Jann, K. 1976b. The cell-wall lipopolysaccharide of *Escherichia coli* K12. Structure and acceptor site for O-antigen and outer substituents. *Eur. J. Biochem.* **70** : 171-177.

Prentki, P., Karch, F., Ida, S. and Meyer, J. 1981. The plasmid cloning vector pBR325 contains a 482 base-pair-long inverted duplication. *Gene* **14** : 289-299.

Quigley, N.B. and Reeves, P.R. 1987. Chloramphenicol resistance vector based on pUC9. *Plasmid* **17** : 54-57.

Radke, K.L. and Siegel, E.C. 1971. Mutation preventing capsular polysaccharide synthesis in *Escherichia coli* K12 and its effect on bacteriophage resistance. *J. Bacteriol.* **106** : 432-437.

Raetz, C.R.H. 1975. Isolation of *Escherichia coli* mutants defective in enzymes of membrane lipid synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **72** : 2274-2278.

Raetz, C.R.H. 1982. Genetic control of phospholipid bilayer assembly, pp 435-477. *In* J.N. Hawthorne and G.B. ansell (eds.), *Phospholipids.* Elselvier Biomedical Press, Amsterdam.

Ramia, S., Neter, E. and Brenner, D.J. 1982. Production of enterobacterial common antigen as an aid to classification of newly identified species of the families *Enterobacteriaceae* and *Vibrionaceae.* *Int. J. Syst. Bacteriol.* **32** : 395-398.

Rapin, A.M.C. and Mayer, H. 1966. Complex polysaccharides in different strains of *Escherichia coli* K23. *Ann. N.Y. Acad. Sci.* **133** : 425-437.

Ratcliff, S.W., Luh, J., Ganesan, A.T., Behrens, B., Thompson, R., Montenegro, M.A., Morelli, G. and Trautner, T.A. 1979. The genome of *Bacillus subtilis* SPP1: The arrangement of restriction endonuclease generated fragments. *Mol. Gen. Genet.* **168** : 165-172.

Raziuddin, S. 1980. Immunochemical studies of the lipopolysaccharides of *Vibrio cholerae* : Constitution of the O-specific side chain and core polysaccharide. *Infect. Immun.* **27** : 211-215.

Redmond, J.W., Korsch, M.J. and Jackson, G.D.F. 1973. Immunochemical studies of the O-antigens of *Vibrio cholerae*. Partial characterization of an acid-labile antigenic determinant. *Aust. J. Exp. Biol. Med. Sci.* **51** : 229-235.

Rick, P.D., Mayer, H., Neumeyer, B.A., Wolski, S. and Bitter-suermann, D. 1985. Biosynthesis of Enterobacterial Common Antigen. *J. Bacteriol.* **162** : 494-503.

Rietschel, E.Th., Galanos, C., Lüderitz, O. and Westphal, O. 1982. Chemical structure, physiological function and biological activity of lipopolysaccharides and lipid A, pp 183-229. *In* D.R. Webb (ed.), *Immunopharmacology and the Regulation of Leukocyte Function.* Marcel Dekker, New York.

Rietschel, E.Th., Wollenweber, H-W., Sidorczyk, Z., Zahringer, U. and Lüderitz, O. 1983. Analysis of the primary structure of lipid A, pp 195-218. *In* L. Anderson and F. Unger (eds.), *Bacterial Lipopolysaccharides: Structure, Synthesis, Biological Activities, American Chemical Society Symposium Series,* Vol. 231. Washington D.C.

Rietschel, E.Th., Wollenweber, H., Russa, R., Brade, H. and Zahringer, U. 1984a. Concepts of the chemical structure of lipid A. *Rev. Infect. Dis.* 6 : 432-438.

Rietschel, E.Th., Zahringer, U., Wollenweber, H-W., Tanamoto, K., Galanos, C., Lüdertiz, O., Kusumoto, S. and Shiba, T. 1984b. The chemistry of bacterial lipopolysaccharides with emphasis on the structure and synthesis of their lipid A component. *In* A. Tu., W.H. Habig and M.C. Hardegree (eds.), *Handbook of Natural Toxins,* vol. II. Bacterial Toxins. Marcel Dekker, New York, in press.

Rigby, P.W.J., Dieckmann, M., Rodes, C. and Berg, P. 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. MOl. Biol.* **113** : 237-251.

Roantree, R.J., Kuo, T-T. and MacPhee, D.G. 1977. The effect of defined lipopolysaccharide core defects upon antibiotic resistance of *Salmonella typhimurium. J. Gen. Microbiol.* **103** : 223-234.

Robbins, P.W. and Uchida, T. 1962. Studies on the chemical basis of the phage conversion in the E group *Salmonella*. *Biochemistry* **1** : 323.

Robbins, P.W. and Uchida, T. 1965. Chemical and macromolecular structure of O-antigens from *Salmonella anatum* strains carrying mutants of bacteriophage ε_{15}. *J. Biol. Chem.* **240** : 375-383.

Robbins, P.W., Wright, A. and Dankert, M. 1966. Polysaccharide biosynthesis. *J. Gen. Physiol.* **49** : 331-346.

Robbins, P.W. Bray, D., Dankert, M. and Wright, A. 1967. Direction of chain growth in polysaccharide synthesis. *Science* **158** : 1536-1542.

Robbins, P.W. and Wright, A. 1971. Biosynthesis of O antigens, pp 351-368. *In* G.Weinbaum, S. Kadis and S.J. Ajl (eds.), *Microbial Toxins,* vol. 4. Academic Press, New York.

Romanowska, E. and Reinhold, V. 1973. 2-Amino-2-deoxyhexuronic acid: a constituent of *Shigella sonnei* phase I lipopolysaccharide. *Eur. J. Biochem.* **36** : 160-166.

Rose, G.D. and Roy, S. 1980. Hydrophobic basis of packing globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **77** : 4643-4647.

Rosenberg, M. and Court, D. 1979. Regulatory sequences involved in the promotion and termination of DNA transcription. *Ann. Rev. Genet.* **13** : 319-353.

Rosenthal, K. and Storm, D. 1977. Disruption of the *Escherichia coli* outer membrane permeability barrier by immobilized polymyxin B. *J. Antibiot.* **30** : 1087-1092.

Rosner, M.R., Tsang, J.Y., Brazilay, I. and Khorana, H.G. 1979. Structure of the lipopolysaccharide from an *Escherichia coli* heptose-less mutant. I. Chemical degradations and identification of products. *J. Biol. Chem.* **254** : 5906-5917.

Rossman, M.G., Liljas, A., Branden, C.I. and Banazak, L.J. 1975. Evolutionary and structural relationships among dehydrogenases, pp. 61-102. *In* Boyer, P.D. (ed.), *The Enzymes*. Academic Press, New York.

Rothfield, L. and Romeo, D. 1971. Role of lipids in the biosynthesis of the bacterial cell envelope. *Bacteriol. Rev.* **35** : 14-38.

Rubenstein, P.A. and Strominger, J.L. 1974. Enzymatic synthesis of cytidine diphosphate 3,6-dideoxyhexoses. VIII. Studies on the properties of E3 and its role in the formation of cytidine diphosphate-4-keto 3,6-dideoxyglucose. *J. Biol. Chem.* **249** : 3782-3788.

Ruden, U. and Mäkelä, P.H. 1974. O-acetylation and glucosylation of lipopolysaccharide in hybrids between *Salmonella* groups B and C_2. *Eur. J. Biochem.* **48** : 11-20.

Samuelson, K., Lindberg, K. and Brubaker, R.R. 1974. Structure of O-specific side chains of lipopolysaccharides from *Yersinia pseudotuberculosis.* *J. Bacteriol.* **117** : 1010-1016.

Sanderson, K.E., MacAlister, T. and Costerton, J.W. 1974. Permeability of lipopolysaccharide-deficient (rough) mutants of *Salmonella typhimurium* to antibiotics, lysozyme, and other agents. *Can. J. Microbiol.* **20** : 1135-1145.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74** : 5463-5467.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143** : 161-178.

Sansonetti, P.J., David, M. and Toucase, M. 1980. Correlation entre la perte d'ADN plasmidique et le passage de la phase I virulente a la phase II avirulente chez *Shigella sonnei. A. R. Acad. Sci.* **290** : 879-882.

Sansonetti, P.J., Kopecko, D.J. and Formal, S.B. 1981. *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. *Infect. Immun.* **34** : 75-83.

Sarvas, M. 1967. Inheritance of *Salmonella* T_1 antigen. *Ann. Med. Exp. Finn.* **45** : 447-471.

Sarvas, M., Malinen, M., Nurnsinen, M. and Mäkelä, P.H. 1976. T_2 lipopolysaccharide antigen of *Salmonella* : comparison of the properties of T2 and mucoid forms. *Infect. Immun.* **14** : 839-842.

Sasaki, T. and Uchida, T. 1974. Mutant of group D_1 *Salmonella* carrying the somatic antigen of group A organisms : evidence for the lack of cytidine diphosphate paratose-2-epimerase. *J. Bacteriol.* **117** : 13-18.

Scher, M., Lennarz, W.J. and Sweeley, C.C. 1968. The biosynthesis of mannosyl-1-phosphoryl-polyisoprenol in *Micrococcus lysodeikticus* and its role in mannan synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **59** : 1313-1320.

Schindler, P.R.G. and Teuber, M. 1978. Ultrastructural study of *Salmonella typhimurium* treated with membrane-active agents: specific reaction of dansylchloride with cell envelope components. *J. Bacteriol.* **135** : 198-206.

Schlecht, S. and Schmidt, G. 1969. Moglichkeiten zur Differenzierung von *Salmonella* - R - Formen mittels Antibiotica und antibakterieller Farbstoffe. *Zentralbl. Bakteriol Parasitenk. Infektionskr. Hyg. Abt. 1, Orig.* **212** : 505-511.

Schlecht, S. and Westphal, O. 1970. Untersuchungen zur Typisierung von *Salmonella* - R - Formen. *Zentr. Bakteriol. Parasitenk. Infektionskr. Hyg. Abt. 1, Orig.* **213** : 356-381.

Schmidt, G., Jann, B. and Jann, K. 1969. Immunochemistry of R lipopolysaccharides of *Escherichia coli.* Different core regions in the lipopolysaccharides of O group 8. *Eur. J. Biochem.* **10** : 501-510.

Schmidt, G., Mayer, H. and Mäkelä, P.H. 1976. Presence of *rfe* genes in *Escherichia coli,* their participation in the biosynthesis of O antigens and enterobacterial common antigen. *J. Bacteriol.* **127** : 755-762.

Sedlak, J. and Rische, H. 1961. Enterobacteriaceae Infektionen. VEB Georg Thieme, Leipzig.

Seid, R.C.Jr., Kopecko, D.J., Sadoff, J.C., Schneider, H., Bann, L.S. and Formal, S.B. 1984. Unusual lipopolysaccharide antigens of a *Salmonella typhi* oral vaccine strain expressing the *Shigella sonnei* form I antigen. *J. Biol. Chem.* **259** : 9028-9034.

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.

Sheu, C.W. and Freese, E. 1973. Lipopolysaccharide layer protection of gram-negative bacteria against inhibition by long-chain fatty acids. *J. Bacteriol.* **115** : 868-875.

Shibaev, V.N. 1978. Biosynthesis of *Salmonella* O antigenic polysaccharides: specificity of glycosyl transferases. *Pure Appl. Chem.* **50** : 1421-1436.

Shibaev, V.N., Danilov, L.L., Druzhinina, T.N., Gogolashwili, L.M., Maltsev, S.D. and Kochetkov, N.K. 1982. Enzymatic synthesis of *Salmonella* O-specific polysaccharide analogs from modified polyprenol sugar acceptors. *FEBS Lett.* **139** : 177-180.

Shibaev, V.N. 1986. Biosynthesis of bacterial polysaccharide chains composed of repeating units. *Adv. Carbohydr. Chem. Biochem.* **44** : 277-339.

Shine, J. and Dalgarno, L. 1974. The 3'-terminal sequence of *E. coli* 16s ribosomal RNA : complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.* **71** : 1342-1346.

Sidorczyk, Z., Zahringer, U. and Rietschel, E.Th. 1983. Chemical structure of the lipid A component of the lipopolysaccharide from a *Proteus mirabilis* Re-mutant. *Eur. J. Biochem.* **137** : 15-22.

Siewert, G. and Strominger, J.L. 1967. Bacitracin: an inhibitor of the dephosphorylation of lipid pyrophosphate, and intermediate in biosynthesis of the peptidoglycan of bacterial cell walls. *Proc. Natl. Acad. Sci. U.S.A.* **57** : 767-773.

Silverman, M. and Simon, M. 1980. Phase variation: genetic analysis of switching mutants. *Cell* **19** : 845-854.

Simmons, D.A.R. 1971. Immunochemistry of *Shigella flexneri* O-antigens: a study of structural and genetic aspects of the biosynthesis of cell-surface antigens. *Bacteriol. Rev.* **35** : 117-148.

Smit, J., Kamio, Y. and Nikaido, H. 1975. Outer membrane of *Salmonella typhimurium* chemical analysis and freeze-fracture studies with lipopolysaccharide mutants. *J. Bacteriol.* **124** : 942-958.

Smith, S.M. and Stocker, B.A.D. 1962. Colicinogeny and recombination. *Brit. Med. Bull.* **18** : 46-51.

Smith, G.E. and Summers, M.D. 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl paper. *Anal. Biochem.* **109** : 123-129.

Southern, E. 1975. Detection fo specific sequence among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98** : 503-517.

Staden, R. 1980. A new computer method for the storage and manipulation of DNA gel reading data. *Nucl. Acids Res.* **8** : 3673-3694.

Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucl. Acids Res.* **10** : 4731-4751.

Staden, R. and McLachlan, A.D. 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucl. Acids Res.* **10** : 141-156.

Staden, R. 1984. Computer methods to locate signals in nucleic acid sequences. *Nucl. Acids Res.* **12** : 505-519.

Staden, R. 1986. The current status and probability of our sequence handling software. *Nucl. Acids Res.* **14** : 217-231.

Staub, A.M. and Bagdian, G. 1966. Etudes immunochimiques sur les *Salmonella*. XII. Analyse immunologique des facteurs 27A, 27B en 27D. *Ann. Inst. Pasteur Paris* **110** : 849-853.

Sternberg, N., Tiemeier, D. and Enquist, L. 1977. *In vitro* packaging of a λ dam vector containing EcoRI DNA fragments of *Escherichia coli* and phage P1. *Gene* **1** : 255-280.

Stocker, B.A.D., Staub, A.M., Tinelli, R. and Kopacka, B. 1960. Etude immunochimique sur les *Salmonella*. VI. Etude de l'antigene I present sur deux *Salmonella* des groupes B et E4. *Ann. Inst. Pasteur Paris.* **98** : 505-523.

Stocker, B.A.D. and Mäkelä, P.H. 1971. Genetic aspects of biosynthesis and structure of *Salmonella* lipopolysaccharide, pp 369-438. *In* G. Weinbaum, S. Kadis and S.J. Ajl (eds.), *Microbial Toxins,* vol. 4, Bacterial endotoxins. Academic Press, New York.

Stocker, B.A.D. and Mäkelä, P.H. 1978. Genetics of the (gram negative) bacterial surface. *Proc. Roy. Soc. Lond. B.* **202** : 5-30.

Stoleru, G.H., Le Minor, L. and Lheritier, A.M. 1976. Polynucleotide sequence divergence among strains of *Salmonella* sub-genus IV and closely related organisms. *Ann. Microbiol.* (Paris) **127A** : 477-486.

Stone, K.J. and Strominger, J.L. 1971. Mechanism of action of bacitracin: compexation with metal ion and C55-isoprenyl pyrophosphate. *Proc. Natl. Acad. Sci. U.S.A.* **68** : 3223-3227.

Storm, D.R., Rosenthal, K.S. and Swanson, P.E. 1977. Polymyxin and related peptide antibiotics. *Ann. Rev. Biochem.* **46** : 723-763.

Strain, S.M., Fesik, S.W. and Armitage, I.M. 1983a. Characterization of lipopolysaccharide from a heptoseless mutant of *Escherichia coli* by carbon 13 nuclear magnetic resonance. *J. Biol. Chem.* **258** : 2906-2910.

Strain, S.M., Fesik, S.W. and Armitage, I.M. 1983b. Structure and metal-binding properties of lipopolysaccharides from heptoseless mutants of *Escherichia coli* studies by ^{13}C and ^{31}P nuclear magnetic resonance. *J. Biol. Chem.* **258** : 13466-13477.

Sueoka, N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. Natl. Acad. Sci. U.S.A.* **48** : 582-592.

Sugimoto, K. and Okazaki, R. 1967. A new rhamnosyl compound in lipopolysaccharide preparation of *Escherichia coli*. *J. Biochem.* (Tokyo) **62** : 373-383.

Sunshine, M.G. and Kelly, B. 1971. Extent of host deletions associated with bacteriophage P2-mediated eduction. *J. Bacteriol.* **108** : 695-704.

Sutherland, I.W. 1977. Bacterial exopolysaccharides - their nature and production, pp 27-96. *In* I. Sutherland (ed.), *Surface carbohydrates of the procaryotic cell*. Academic Press, London.

Svenson, S.B. and Lindberg, A.A. 1983. Artificial *Salmonella* vaccines, pp 120-143. *In* K Ishizeka, P. Kallos, B.H. Waksman and A.L. de Weck (eds.), *Progress in Allergy,* vol. 33, Karger, Basel.

Syvanen, M. 1986. Cross-species gene transfer : a major factor in evolution? *Trends Genet.* **2** : 63-66.

Takayama, K., Qureshi, N., Ribi, E. and Cantrell, I.L. 1984. Separation and characterization of toxic and nontoxic forms of lipid A's. *Infect. Dis.* **6** : 439-443.

Takeshita, M. and Mäkelä, P.H. 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.

Takeuchi, Y. and Nikaido, H. 1981. Persistance of segregated phospholipid domains in phospholipid-lipopolysaccharide mixed bilayers: studies with spin-labelled phospholipids. *Biochemistry* **20** : 523-529.

Teuber, M. 1970. Lysozyme dependent production of spheroplasts-like bodies from polymyxin B treated *Salmonella typhimurium. Arch. Microbiol.* **70** : 139-146.

Teuber, M. and Bader, J. 1976. Action of polymyxin on bacterial membranes. Binding capacities for polymyxin B of inner and outer membranes isolated from *Salmonella typhimurium* G30. *Arch. Microbiol.* **109** : 51-58.

Thomshich, V., Gorshkova, R.P., Elkin, Y.N. and Ovodov, Y.S. 1976. Lipopolysaccharide from *Yersinia pseudotuberculosis,* type 1B. A structure study of O-specific chains. *Eur. J. Biochem.* **65** : 193-199.

Tinelli, R. and Staub, A.M. 1960. Etude immunochimique sur les *Salmonella*. VII. Etude des products de l'hydrolyse acide menagee du polyoside extrait de *S. tyhpi.* 1re partie: Analyse de l'antigene 0:12 du tableau de Kauffman-White. *Bull. Soc. Chim. Biol.* **42** : 583-599.

Torres-Cabasa, A.S. and Gottesman, S. 1987. Capsule synthesis in *Escherichia coli* K-12 is regulated by proteolysis. *J. Bacteriol.* **169** : 981-989.

Troy, F.A., Frerman, F.E. and Heath, E.C. 1971. The biosynthesis of capsular polysaccharide in *Aerobacter aerogenes. J. Biol. Chem.* **246**: 118-133.

Troy, F.A., Visay, I.K. and Tesche, N. 1975. Role of undecaprenyl phosphate in synthesis of polymers containing sialic acid in *Escherichia coli. J. Biol. Chem.* **250** : 156-163.

Uchida, T., Matsumoto, T. and Sasaki, T. 1974. Mutants of group D 1 *Salmonella* carrying the somatic antigen of group A organisms : isolation and serological characterization. *J. Bacteriol.* **117** : 8-12.

Uetake, H., Nakagawa, T. and Akiba, T. 1955. The relationship of bacteriophage to antigenic changes in group E *Salmonellas*. *J. Bacteriol.* **69** : 571-579.

Unger, F.M. 1981. The chemistry and biological significance of 3-deoxy-D-manno-2-octulosonic acid (KDO). *Adv. Carbohydr. Chem. Biochem.* **38** : 323-388.

Vaara, M. 1981. Increased outer membrane resistance of ethylenediaminetetraacetate and cations in novel lipid A mutants. *J. Bacteriol.* **148** : 426-434.

Vaara, M. and Vaara, T. 1981. Outer membrane permeability barrier disruption by polymyxin in polymyxin-susceptible and resistant *Salmonella typhimurium*. *Antimicrob. Agents Chemother.* **19** : 578-583.

Vaara, M. and Nikaido, H. 1984. Molecular organization of bacterial outer membrane, pp 1-45. *In* E.T. Reitschel (ed.), *Handbook of endotoxin,* vol. 1. Elselvier Science Publishers, Amsterdam.

Valtonen, M.V., Sarvas, M. and Mäkelä, P.H. 1976. T2 lipopolysaccharide antigen of *Salmonella* : genetic determination of T2 and properties of the T2, T2, S and T2, SR forms. *Infect. Immun.* **13** : 1647-1653.

Vieira, J. and Messing, J. 1982. The pUC plasmids, and M13 mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primer. *Gene* **19** : 259-268.

Watanabe, H. and Timmis, K.N. 1984. A small plasmid in *Shigella dysenteriae* 1 specifies one or more functions essential for O antigen production and bacterial virulence. *Infect. Immun.* **43** : 391-396.

Weiner, I.M., Higushi, T., Osborn, M.J., Rothfield, L., Saltmarsh-Andrew, and Horecker, B.L. 1965. Biosynthesis of bacterial lipopolysaccharide. V. Lipid-linked intermediates in the biosynthesis of the O-antigen groups of *Salmonella tyhpimurium*. *Proc. Natl. Acad. Sci. U.S.A.* **54** : 228-234.

Weisgerber, C. and Jann, K. 1982. Glucosyldiphosphoundecaprenol, the mannose acceptor in the synthesis of the 09 antigen of *Escherichia coli*. *Eur. J. Biochem.* **127** : 165-168.

Wilber, W.J. and Lipman, D.J. 1983. Rapid similarity searches of nucleic acid and protein databanks. *Proc. Natl. Acad. Sci. U.S.A.* **80** : 726-730.

Wilkinson, R.G., Gemski, P. Jr. and Stocker, B.A.D. 1972. Non-smooth mutants of *Salmonella typhimurium:* differentiation by phage sensitivity and genetic mapping. *J. Gen. Microbiol.* 70 : 527-554.

Wilkinson, R.G. 1977. Composition and structure of bacterial lipopolysaccharides, pp 97-176. *In* I. Sutherland (ed.), *Surface carbohydrates of the procaryotic cell.* Academic Press, New York.

Winblad, S. 1968. Studies on O-antigen factors of *'Yersinia enterocolitica'. Series Immunbiol. Standard.* 9 : 337-342.

Wright, A., Dankert, M., Fennessay, P. and Robbins, P.W. 1967. Characterization of a polyisoprenoid compound functional in O antigen biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 57 : 1798-1803.

Wyk, P. 1988. Genetics of abequose biosynthesis in the *rfb* region of *Salmonella typhimurium* LT2. Ph.D. thesis, University of Adelaide, Adelaide.

Yamamoto, T., Gojobori, T. and Yokota, T. 1987. Evolutionary origin of pathogenic determinants in enterotoxigenic *Escherichia coli* and *Vibrio cholerae* O1. *J. Bacteriol.* 169 : 1352-1357.

Yanisch-Perron, C., Vieira, S. and Messing, J. 1985. Improved M13 phage cloning vectors and host strains : nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33** : 103-119.

Yanofsky, C., Platt, T., Crawford, I.P., Nichols, B.P., Christie, G.E., Horowitz, H., Vancleemput, M. and Wu, A.M. 1981. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli. Nucl. Acids Res.* **9** : 6647-6668.

Yuan, R. and Horecker, B.L. 1968. Biosynthesis of the O antigen from *Citrobacter* 139. *J. Bacteriol.* **95** : 2242-2248.

Yuasa, R., Levinthal, M. and Nikaido, H. 1969. Biosynthesis of cell wall lipopolysaccharide in mutants of *Salmonella. J. Bacteriol.* **100** : 433-444.

Yuasa, R., Nakae, K. and Nikaido, H. 1970. Structure of cell wall lipopolysaccharide from *Salmonella typhimurium.* Structure of lipopolysaccharide from a semirough mutant. *Eur. J. Biochem.* **15** : 63-71.

Zinder, N. 1957. Lysogenic conversion in *Salmonella typhimurium. Science* **126** : 1237.

Zinkewich-Peotti, K. and Fraser, J.M. 1988. New locus for exopolysaccharide overproduction in *Escherichia coli* K-12. *J. Bacteriol.* **170** : 1405-1407.