***Escherichia coli* K12 regains its O antigen**

Dan Liu and Peter R. Reeves

Author for correspondence: Peter R. Reeves. Tel: +61 2 6922536. Fax: +61 2 6924571.

Department of Microbiology
(G08), The University of
Sydney, Sydney NSW 2006,
Australia

Extant *Escherichia coli* K12 strains are phenotypically rough, their lipopolysaccharide having a complete core structure, but no O antigen. We used DNA hybridization and DNA sequencing to show that the rough phenotype of this strain is due to the presence of one of two independent mutations in the *rfb* gene cluster. The *rfb-50* mutation, consisting of an IS5 insertion at the downstream end of *rfb*, is present in strain EMG2, which is representative of most K12 derivatives. The *rfb-51* mutation is a deletion at the upstream end of *rfb*, and was found in strain WG1. A gene cloned from strain WG1 could complement the *rfb-50* mutation in strain EMG2, and the complemented strain produced O antigen which was typed as O16 with cross reaction to O17.

Keywords: *Escherichia coli* K12, lipopolysaccharide, O antigen, *rfb* gene mutations

INTRODUCTION

Escherichia coli is one of the best studied species in terms of biochemistry and genetics. Most of the work has been done with strain K12, which was first isolated in 1922 and used as a standard *E. coli* strain at Stanford University (Bachmann, 1987). It was used by Gray & Tatum (1944) to make auxotrophic mutants, and soon after by Lederberg & Tatum (1946) and Tatum & Lederberg (1947) for their work on recombination in bacteria. All surviving strains of *E. coli* K12 appear to lack O antigen and there is no reference to the strain having carried an O antigen in the period since 1944 (Gray & Tatum, 1944) when it was first reported in the scientific literature.

The O antigen is part of the lipopolysaccharide (LPS) and is attached to the lipid A-core component, which is synthesized separately. Mutants which lack O antigen can be defective either in the O antigen (*rfb*) gene cluster or in the LPS core (*rfa*) gene cluster, as in *rfa* mutants O antigen cannot be added to the incomplete core. *E. coli* K12 has a complete core structure (Prehm *et al.*, 1976; Schmidt, 1973), and K12 strains carrying cloned *rfb* gene clusters from various sources add O antigen to the LPS core (Al-Hendy *et al.*, 1991; Jiang *et al.*, 1991; Valvano & Crosa, 1989). It thus appears that *E. coli* K12 underwent mutation in *rfb* prior to its adoption as a strain for laboratory study. The genes for a rhamnose-biosynthesis pathway (*rfbA*,

rfbB, *rfbC* and *rfbD*) have been located near the histidine (*his*) operon and gluconate dehydrogenase (*gnd*) gene, and are considered to be part of the *rfb* gene cluster of *E. coli* K12 (Nikaido *et al.*, 1965). A complete *rfb* region would have glycosyltransferases and perhaps other sugar biosynthetic pathway genes, but none have yet been identified. A copy of the IS5 element was reported 0.5 kb upstream from the promoter of the *gnd* gene in *E. coli* K12 (Barcak & Wolf, 1988), which places it between the identified *rfb* genes and the *gnd* gene.

In this study we show that independent *rfb* mutations occurred in two lines of *E. coli* K12 strains: we present a comparison of the *rfb* region of several strains of K12, and show that a plasmid carrying part of the *rfb* region of K12 strain WG1 can complement the *rfb* mutation of K12 strain EMG2. We also present the sequence of part of the *rfb* region of two K12 strains, one with and one without the IS5 insertion, and discuss the mutational events which occurred in the *rfb* regions of different *E. coli* K12 strains.

METHODS

Media, bacterial strains, plasmids and bacteriophages. Bacteria were grown in L-broth (constituents from Difco) or on nutrient agar plates with or without antibiotics. CY broth [10 g Casamino acids (Difco), 5 g yeast extract (Oxoid), 3 g NaCl and 2 g KCl in 1 litre of water, with Tris/HCl (pH 7.4) and MgCl₂ added after autoclaving to final concentrations of 30 mM and 10 mM, respectively] was used for phage propagation. The bacterial strains and plasmids used for this study are described in Table 1. A set of λ phage clones (Kohara *et al.*, 1987) covers the whole *E. coli* K12 chromosome. Four of the clones, K350, K351, K352 and K353, from the 44 min

The GenBank accession numbers for the sequence data reported in this paper are L19537 for the *rfb* sequence from strain WG1 and L19538 for the *Sqx* sequence.

Table 1. Bacterial strains and plasmids

Strains/ plasmids	Lab stock no.	Characteristics*	Reference or source
<i>E. coli</i> K12 strains			
C600	P334	F ⁻ <i>thr-1 leuB6 tonA1 lacY1 tonA21 supE44 rfbD</i> λ^-	Appleyard (1953)
58-161F ⁻ -S ^R	P501	<i>relA rpsL100 metB1</i>	Hayes (1953)
W1485F ⁺	P955	<i>supE</i> λ^- made from WG1 (but see text)	Lederberg & Lederberg (1953)
So874	P4052	<i>lacZ4503 try-355 upp-12 relA rpsL150 Δ(<i>shcB-rfb</i>)</i>	Neuhard & Thomassen (1976)
NM538	P4396	<i>glnV hsdR</i> λ^-	Frischauf <i>et al.</i> (1983)
EMG2	P4809	Wild-type strain (CGSC4401)	Bachmann (1972)
WG1	P4810	Wild-type strain (CGSC5073)	Bachmann (1972)
	P4863	W1485 carrying plasmid pPR1474	This study
CR63	P4881	<i>lamB63 supD60</i> λ^- (CGSC3591)	Appleyard <i>et al.</i> (1956)
KL983	P4882	<i>lacY1 mglP1 xyl-7</i> λ^- (CGSC4240)	Low (1973)
J5-3	P4884	<i>pro-22 metF63</i> (CGSC6850)	Clowes & Rowley (1954)
Plasmids			
pMN1		6.1 kb <i>Hind</i> III fragment containing the <i>gnd</i> gene from K12 strain AB311 in pBR322	Nasoff & Wolf (1980)
pPR685		<i>cps</i> genes of <i>S. enterica</i> LT2 in pUC8	Stevenson <i>et al.</i> (1991)
pPR963		<i>rfbA</i> , <i>rfbC</i> and <i>rfbD</i> genes of <i>S. enterica</i> LT2 in pUC19	X. M. Jiang (this lab.), unpublished
pPR1010		<i>rfbB</i> gene of <i>S. enterica</i> LT2 in pUC9	Jiang <i>et al.</i> (1991)
pPR1167		3.1 kb <i>Hind</i> III fragment from K352 in pUC18	This study
pPR1168		4.1 kb <i>Eco</i> RI- <i>Hind</i> III fragment from K352 in pUC18	This study
pPR1461		8 kb <i>Eco</i> RI fragment from WG1 in pUC18 in orientation for expression	This study
pPR1462		8 kb <i>Eco</i> RI fragment from WG1 in pUC18 in incorrect orientation for expression	This study
pPR1474		Deletion derivative of pPR1462 (3 kb <i>Sph</i> I fragment deleted)	This study
pPR1491		2.2 kb <i>Hind</i> III- <i>Pst</i> I fragment from pPR1461, in pUC18	This study

* CGSC, *E. coli* Genetic Stock Center.

region were used in this study (Fig. 1). Phage lysates were prepared by the 'small-scale liquid propagation' method (Sambrook *et al.*, 1989), modified by using one fresh plaque as inoculum and propagating in strain NM538 in 30 ml of CY broth, and DNA was extracted from the phage.

Enzymes and reagents. All the enzymes and dNTPs were from Boehringer Mannheim or Pharmacia LKB, and [α -³²P]dCTP and [³⁵S]dATP α S was from Bresatec. Sequencing was done using a kit from USB Cleveland, and oligonucleotides were synthesized using an Applied Biosystems DNA synthesizer.

DNA methods. The preparations of chromosomal and plasmid DNA were as described by Bastin *et al.* (1991), and Southern hybridization was as described by Sambrook *et al.* (1989). The dideoxy-chain-termination method (Sanger *et al.*, 1977) was

used for DNA sequencing, using M13mp18 or M13mp19 and universal primer unless otherwise stated. PCR (Saiki *et al.*, 1988) was employed to amplify DNA fragments.

Computer analysis. The DNA sequence was analysed using the Australian National Genomic Information Service (ANGIS) at Sydney University (Reisner *et al.*, 1993), which incorporates several sets of programs. The deduced amino acid sequences were analysed using programs in the GCG package: Bestfit (Devereux *et al.*, 1984) for comparing the sequences and Peplot (Chou & Fasman, 1978) for the prediction of secondary structure. Potential transmembrane segments were predicted using programs based on the methods of Eisenberg *et al.* (1984) and the program ALOM (NIH programs: M. Kanehisa, National Institutes of Health, USA) based on the method of Klein *et al.* (1985). Sequence databases were searched using the NCBI BLAST program (Altschul *et al.*, 1990; Gish & States, 1993).

Silver stain for LPS. LPS preparations from whole cell membranes, SDS-PAGE and silver staining were carried out as previously described by Brown *et al.* (1991).

Serology. Strain P4863 was typed by Dr Skov Sørensen of The International *Escherichia* and *Klebsiella* Centre, Statens Serum-institut in Copenhagen, Denmark.

RESULTS AND DISCUSSION

Location of the *rfb* region in the *E. coli* K12 physical map

In *Salmonella enterica*, the capsular polysaccharide (*cps*) genes lie upstream of *rfb*, while the *gnd* gene and *his* operon are downstream (Jiang *et al.*, 1991; Stevenson *et al.*, 1991). We expect the *rfb* genes to be located at the equivalent position in *E. coli* K12.

The *his* operon and the *gnd* gene of *E. coli* K12 are at positions 2100.1–2108 and 2109.6–2111, respectively, on the physical map (Kohara *et al.*, 1987; Rudd *et al.*, 1992) (see Fig. 1). This region is spanned by Kohara clones K350–K353.

A 6.1 kb *Hind*III fragment from Kohara clones K351 hybridized with plasmid pMN1 (Nasoff & Wolf, 1980), which carries the *gnd* gene of K12; this result agrees with the finding of Rudd *et al.* (1992).

In order to locate the *rfb* region on the physical map (Kohara *et al.*, 1987), Kohara clones were hybridized with DNA carrying *rfb* genes and *rfb* flanking regions from *S. enterica* LT2. pPR1010 (Jiang *et al.*, 1991) carries the *rfbB* gene of LT2. This plasmid hybridized with a 4.1 kb *Eco*RI–*Hind*III fragment from K351 and K352 in the overlap region, and an additional 3.1 kb *Hind*III fragment from K352 (map positions 2116.8–2124) (Fig. 1). The plasmid pPR963, carrying the *rfbA*, *rfbC* and *rfbD* genes of LT2, hybridized with the same 4.1 kb *Eco*RI–*Hind*III fragment. Sequencing from a *Hind*III site (Fig. 1) located the *rfbB* gene of *E. coli* K12, which had 89% identity with *rfbB* of *S. enterica* strain LT2 at the amino acid level for the 561 bp sequenced (data not shown). The sequence also indicates that transcription of the *rfbB* gene of K12 is towards *gnd*, in agreement with transcription of *rfb* in LT2 (Jiang *et al.*, 1991). The *Hind*III site is not present in LT2, but the location can be identified as the sequence AAACGTG, which differs by two bases from the *Hind*III recognition sequence, at bases 4714–4719 in the LT2 sequence (Jiang *et al.*, 1991). If we presume that the *rfbB* genes in LT2 and K12 have the same size (906 bp), then the *rfbB* gene of K12 is located from positions 2121.3–2120.3 in the physical map (Fig. 1). If the other rhamnose biosynthetic pathway genes have the same size and organization in K12 as in LT2, they will locate between map positions 2121.3 and 2117.8 (Fig. 1). The distance from *rfbB* to *gnd* is thus about 12 kb (Fig. 1).

pPR685, carrying *cps* genes of LT2 (Stevenson *et al.*, 1991), hybridized with a 7.9 kb *Eco*RI(vector)–*Hind*III fragment from K352, and with both a 6.5 kb

Fig. 1. Restriction map and genes of the *rfb* region of *E. coli* K12 strain W3110. The restriction map and scale are from the genomic physical map (Kohara *et al.*, 1987) as modified by Rudd and co-workers (Bouffard *et al.*, 1992; Rudd *et al.*, 1992). The scale is from EcoMap6 produced by the program GeneScape (Bouffard *et al.*, 1992). The restriction sites with an asterisk are not present in the physical map as published, but were found in Kohara clones in this study: a 2.2 kb *Hind*III fragment present in K351 is missing in the published physical map, and is indicated by a break on the scale. The positions and directions of sequencing are indicated by arrows. Kohara clones and plasmid probes are displayed above the map. Restriction sites: B, *Bam*HI; E, *Eco*RI; H, *Hind*III.

Fig. 2. Strategy for DNA sequencing part of the *rfb* gene cluster. (a) Fragments 1, 2, 3 and 4 from pMN1 (from strain AB311), and fragment 5 from K351 (from strain W3110), were cloned into M13mp18 or M13mp19 for sequencing the regions flanking the ISS element. (b) Fragments 6, 7 and 8 were cloned from pPR1461 into M13mp18 or M13mp19 for sequencing the *rfb* region of strain WG1, and fragment 9 is a 261 bp *Eco*RV fragment from pPR1461. The positions and directions of sequencing are indicated by arrows; arrows with an asterisk indicate that synthetic oligonucleotide primers were used for sequencing. Arrows in the boxes indicate transcription of genes or ORFs, and the areas shaded are those common to strains pMN1/K351 and pPR1461. Restriction sites: Bg, *Bgl*III; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sph*I.

Fig. 3. Southern hybridization of K12 strains with (a) a *gnd* gene probe and (b) an IS5 probe. Chromosomal DNA was digested with *Eco*RI (lanes 1–4) and *Hind*III (lanes 5–8). Lanes: 1 and 5, C600; 2 and 6, EMG2; 3 and 7, WG1; 4 and 8, SØ874. The *gnd* probe was a 1.1 kb PCR product, and the IS5 probe was an *Eco*RI–*Bgl*II fragment from pMN1.

*Eco*RI(vector)–*Bam*HI fragment and adjacent 3.1 kb *Bam*HI fragment from K353: the three fragments span from positions 2124.3 to 2136.5 (Fig. 1). The start of the *cpsB* gene was about 110 bp to the left of the *Bam*HI site (Fig. 1), as determined by sequencing from that site (data not shown), and the sequence showed 83% identity to the *cpsB* sequence of LT2 (from bases 48 to 185) (Stevenson *et al.*, 1991) (data not shown).

These data allowed us to locate the *his*, *gnd*, *rfbB* and *cpsB* genes on the Kohara map (Fig. 1). The arrangement of these genes is the same as in the corresponding region of *S. enterica* (Jiang *et al.*, 1991; Stevenson *et al.*, 1991). This arrangement also agrees with the gene order in the *E. coli* K12 linkage map, version 7 (Bachmann, 1983): note that in the later version (Bachmann, 1990) the *cps* genes were incorrectly located.

Sequencing the insertion site of the IS5 element

The IS5 element found between the *gnd* gene and the known *rfb* genes in *E. coli* K12 (Barcak & Wolf, 1988) could be within the *rfb* gene cluster and the cause of the loss of *rfb* function. This raised the possibility of reconstructing the original sequence. We therefore sequenced the DNA flanking this IS5 element using plasmid pMN1: the sequencing strategy is shown in Fig. 2. The region downstream between the IS5 element and

the *gnd* gene has already been sequenced (Miller *et al.*, 1988; Nasoff *et al.*, 1984), but we resequenced the junction for confirmation, as part of it was critical to our conclusions. There is a 480 bp sequence in this region containing the 3' end (116 amino acids) of an ORF directly adjacent to the inverted repeat sequence of IS5, and a 127 bp non-coding sequence between this ORF and the *gnd* gene (Fig. 2).

657 bp of non-coding sequence was obtained upstream of the IS5 element (Fig. 2). We decided to extend the upstream sequence to reach coding DNA sequence, but as we had sequenced to the end of the pMN1 insert, we used a 2.2 kb *Hind*III fragment from Kohara phage K351 which was directly adjacent to the 6.1 kb *Hind*III fragment which hybridized to pMN1 (Fig. 1). The 510 bp sequence obtained upstream from the *Hind*III site was all potential coding sequence (Fig. 2). We believed that the 657 bp and 510 bp sequences were contiguous (but see below), and concluded that *E. coli* K12 has the structure shown in Fig. 2(a) in the vicinity of the IS5 element near the *gnd* gene, although we obtained the sequence from two sources. IS5 transposition causes a 4-base duplication at the sites of insertion (Engler & van Bree, 1981), but in this case the adjacent sequences are CTAG upstream and TTAA downstream, suggesting that IS5 induced a deletion after insertion (Calos & Miller, 1980) with loss of the 4-base duplicated sequence from one end. If there has been a

Fig. 4. A reconstruction of the *rfb* region of the original *E. coli* K12. The upper part of the figure shows the IS5 insertion, which constitutes mutation *rfb-50* present in most K12 strains, and the lower part shows the *rfb-51* deletion (Δ) found in strain WG1. The plasmids and fragments which were used as probes are also shown in the upper part. The IS5 probe was an *EcoRI*–*Bgl*II fragment from pMN1, the *gnd* probe was a 1.1 kb PCR product, and fragment A comprised 2.2 kb and 1.6 kb *Hind*III fragments isolated from K351. The plasmids cloned from strain WG1 are displayed in the lower part of the figure. The restriction sites with asterisks are present on strain W3110 but not WG1: all other restriction sites are common to strains WG1 and W3110. Restriction sites: E, *EcoRI*; H, *Hind*III; P, *Pst*I; S, *Sph*I.

deletion, then reconstruction of the pre-insertion sequence is not possible.

Attempts to find other strains producing the 'K12' O antigen

As it was not possible to reconstruct the pre-deletion sequence using the above sequence data, we looked instead for other strains which might carry *rfb* genes which can complement the mutation in strain K12. We looked at two natural isolates (ECOR 2 and ECOR 3) which are apparently very closely related to K12, but both were found to have *rfb* regions very different from that of K12 (D. Liu & P. Reeves, unpublished data).

K12 strain WG1 has an independent *rfb* mutation

We decided to look at early K12 derivatives. Dr B. J. Bachman of the *E. coli* Genetics Stock Center sent us two strains: one was WG1, the 'wild-type' strain from J. Lederberg's Wisconsin laboratory, and the other EMG2, ostensibly the same strain as it reached the W. Hayes' laboratory. To our surprise, strain WG1 differed in the *rfb* region from EMG2, AB311(pMN1), W3110 (Kohara phage) and all other K12 strains subsequently tested.

Southern hybridization of *EcoRI*- and *Hind*III-digested chromosomal DNA with a *gnd* gene probe showed up the expected 4.1 kb *EcoRI* and 6.1 kb *Hind*III fragments in strain EMG2, but in strain WG1, 8 kb *EcoRI* and 4.8 kb *Hind*III fragments hybridized (Fig. 3a). When IS5 DNA

Fig. 5. Silver staining of LPS. Membrane extracts were electrophoresed in 18% SDS-PAGE and silver stained. Extracts are from: 1, EMG2; 2, WG1; 3, EMG2(pPR1461); 4, EMG2(pPR1462); 5, EMG2(pPR1474); 6, EMG2(pPR1491).

was used as a probe, EMG2 and WG1 gave different patterns with the 4.1 kb *EcoRI* and 6.1 kb *Hind*III fragments of EMG2 hybridizing to the probe, but not the 8 kb *EcoRI* and 4.8 kb *Hind*III fragments of WG1 (Fig. 3b). Strain Sö874, which lacks the *rfb* region and the *gnd* gene and was used as a negative control, lacked the four fragments: strain C600, the standard K12 strain of this laboratory used as a positive control, resembled strain EMG2 in all significant aspects. These results indicated that there is no IS5 insertion upstream of the *gnd* gene in WG1.

More probes (Fig. 4) were used to hybridize DNA from the four strains. The results (data not shown) indicate that strains C600 and EMG2 are identical throughout the *rfb* region, whereas strain WG1 differs substantially. Plasmid pPR1168, which carries the 3' half of the *rfbB* gene and the *rfbC*, *rfbA* and *rfbD* genes (Fig. 4), hybridized with DNA from strains C600 and WG1, although the fragments differed in size (data not shown). However, pPR1167, which carries the 5' half of the *rfbB* gene and 2.5 kb upstream of *rfbB* (Fig. 4), hybridized to strain C600 but not to strain WG1 (data not shown). These results indicate that strain WG1 carries a substantial deletion which includes *rfbB*.

The presence of two distinct mutations in strains WG1 and EMG2 indicates that the original K12 strain had a complete *rfb* region and that the two strains which descend from it have independent mutations which lead to the same rough phenotype. We have named the

ORF224
 A V F V Q Q F W M K E K P I K K Y S I N N I I V S R P E I
 CTGCAAGTGTITGTTCAACAATTTCTGGATGAAAGAAAAATTTATCAAGAAATATTCTATAAATAACATCATTGTGAGTCGGCCAGAAATTA 90
PstI
 K L S D K S Q L T D D D S Q P K N N P S E L T I F Y P A V P
 AATTATCTGATAAAAGCCAACCTTACTGATGATGATTCTCAATTTAAGAAATAACCCCTCTGAGTTGACAATATTTTACCCCTGCTGTTCCAC 180
 R V F K N Y E L I I S A A R K L K E Q S N I K F L L T I S G
 GAGTATTTAAAAATTACGAGCTTATTATTAGTGCAGCAAGGAAATTTGAAAGAACAATCCAATATTAAATTTCTGCTTACTATCATGTTGTA 270
 T E N A Y A K Y I I S L A E G L D N V H F L O Y L D K E K I
 CAGAAAAATGCGTATGCAAAATATATTATCAGTCTTGCAGAGGACTGGATAATGTTTCATTTCCTCGGGTACTTGGATAAAGAAAAATCG 360
 D H C Y N I S D I V C F P S R L E T W G L P L S E A K E R G
 ATCATTGTTAATAATTTTTCAGATATAGTTTGTTCCTCTAGGTTAGAAACATGCGGATTCGCTTGTCTGAAGCTAAAGAGCGAGGTA 450
 K W V L A S D P P P T R E T L G S Y E K K A F F D S N N D D
 AGTGGGTATFAGCATCAGATTTCCCATTTACTAGAGAAACTCTTGGTAGTTATGAAAAAGAAAGCTTTTGTGATTCTAATAACGATGACA 540
 M L V K L I I D F K K G N L K K D I S D A N P I Y R N E N V
 TGTAGTTAAACTTATTATTGACTTCAAAAAAGGTAACCTCAAAAAAGATATCTCTGATGCAAAATTTTCAATTTATCGTAATGAAATGTAT 630
HindIII
 L V G F D E L V N F I T E E H * M V Y I I I V S H G H E D Y
 TAGTTGGGTTTGTGAACTAGTTAATTTTATTACTGAAGAACATTGAAATGGTATATATAAATCGTTTCCCACGGACATGAAGACTAC 720
 I K K L L N A D D E H Y K I I V R D N K D S L L L K Q
 ATCAAAAAATTTACTCGAAATCTTAATGCTGACGATGAGCACTACAAGATTATCGTACGCGACAACAAAGACTCTCTATTATTGAAACAA 810
 I C Q H Y A G L D Y I S G G V Y G F G H N N N I A V A Y V K
 ATATGCCAGCATATGACAGGCTGGACTATATTAGTGGAGGTGTATACGGCTTTGGTCAATAATAATATATTGCGGTGGCGTATGTAAAG 900
 E K Y R P A D D D Y I L P L N P D I I M K H D L L T Y I K
 GAAAAATATAGACCCGAGATGATGATTACATTTTGTGTTTGAATCCCGATATCATCATGAAGCATGATGATTGCTGACATATATTTAA 990
 Y V E S K R Y A F S T L C L F R D E A K S L H D Y S V R K P
 TATGTCGAAAGTAAGCGTTATGCTTTTGTACATTATGCTGTTCGAGATGAAGCGAAATCTTTACATGATTATTCGTAAGAAAAATTT 1080
 P V L S D P I V S P M L G I N K T K I P K E S I Y S D T V V
 CCTGTGCTTTCTGATTTTATTGTGTCATTATGTTAGGGAATTTTAAACAAAAATTCCTAAAGAAAGTATCTATTCTGATACGGTGTGT 1170
 D W C A G S P M L V R P S D F V R V N G F D Q G Y F M Y C E
 GATTGGTGGCAGGATCATTTATGCTGGTACGTTTTCAGATTTTGTGCGTGAATGGCTTCGATCAAGGTTACTTTATGTACTGTGAA 1260
 D I D L C L R L S L A G V R L H Y V P A F H A I H Y A H H D
 GATATTGACCTGTGCTTGAAGCTTAGCCTGGCTGGTGTGAGACTTCATTATGTTCCCGCTTTTCATGCGATACATTATGCTCATCATGAC 1350
 N R S F P S K A F R W H L K S T F R Y L A R K R I L S N R N
 AATCGAAGTTTTTTTCAAAAGCCTTCAGATGGCACTTAAAAAGTACTTTTAGATATTTAGCCAGAAAACGTATTTTATCAATCGCAAC 1440
 F D R I S S V F H P *
 ITTGATCGAATTTTCATCAGTTTTTACCCTGAATATAAAGCCGTAAGCATATAAGCATGGATAAGCTATTTATACTTTAAATAGTACTTT 1530
 M
 GTATACTTATTTCGAACATTCAGGCCGCGAGCATTACAGCGGGTGATCACACCTGACAGGAGTATGTAATG 1603
 start of the *gnd* gene

Fig. 6. DNA sequence and deduced amino acid sequence of part of the *rfb* region from strain WG1. The IS5 insertion site in *rfb-50* strains is indicated in bold and by shading. This sequence is also found in Kohara phage K351, but with the IS5 insertion, and the sequence from positions 1121 to 1603 is identical to that found in pMN1 downstream of the IS5 element.

mutations in EMG2 and WG1 *rfb-50* and *rfb-51*, respectively.

Cloning of *rfb* genes from WG1

Chromosomal DNA of strain WG1 was digested with *EcoRI*, electrophoresed and fragments in the size range 7–9 kb isolated from the gel and cloned into pUC18. A small bank of 650 colonies was constructed in JM101 and screened by colony blotting using the *gnd* gene as a probe. Five colonies hybridized with the probe and two, which carried the desired 8 kb *EcoRI* insert in different orientation, were designated pPR1461 and pPR1462 (Fig. 4).

Complementation of the *rfb* gene cluster of *rfb-50* K12 strains

The above plasmids were used to see if DNA from strain WG1 could complement the *rfb-50* lesion. Plasmids pPR1461 and pPR1462 were transferred separately into strains EMG2 and W1485, and in each case the transformants produced O antigen as determined by LPS

electrophoresis on SDS-PAGE (data for EMG2 in Fig. 5). The results indicate that the gene(s) cloned from WG1 could complement the *rfb-50* mutation. Plasmids pPR1474 and pPR1491, carrying smaller segments (Fig. 4), gave similar results (Fig. 5). Strain P4863, which is W1485 carrying pPR1474, was found to be O16:K⁻:H48, and also reacted with anti-O17 sera.

The smallest clone that could complement the *rfb-50* mutation in strain EMG2 was pPR1491 with a 2.2 kb insert, of which about 1.1 kb is DNA from the 5' end of the *gnd* gene; the remaining 1.1 kb presumably complements the mutation.

Sequence of part of the *rfb* region of strain WG1

We sequenced the region which complemented strains EMG2 and W1485 (Fig. 2). A DNA sequence of 1603 bp was obtained which contained two ORFs (Fig. 6), one with 224 amino acids (*orf224*) and the other with 264 amino acids (*orf264*).

We assume that the region of strain WG1 which restores

rfb function to strain EMG2 carries authentic K12 sequence, and used this to reconstruct the original K12 sequence (Fig. 6). The reconstructed sequence contains two ORFs, *orf224* and *orf264*. By comparison we can see that pMN1 lacks a 603 bp segment upstream of the IS5 element but has 657 bp segment of unknown origin which we have named Sqx (sequence X) (see GenBank accession number L19538).

The *rfb-50* mutation

The detailed studies on the *rfb-50* mutation reported above were on plasmid pMN1 and Kohara clone K351. We used the IS5 probe to hybridize *Eco*RI and *Hind*III chromosomal DNA from K12 strains C600, EMG2, W1485, J5-3, CR63, 58-161F⁻-S^R and KL983, and in each case the 4.1 kb *Eco*RI fragment and 6.1 kb *Hind*III fragment hybridized (data not shown). This result shows that all these strains carry the *rfb-50* mutation with IS5 insertion. It appeared at first that the IS5 insertion had induced a 603 bp deletion and an insertion of the 657 bp Sqx segment. A search of databases revealed no sequences related to Sqx. However, probing with Sqx (fragment 3, Fig. 2) showed that Sqx was present in all the above K12 strains, including WG1 and Sö874, on a 7.2 kb *Hind*III fragment but not, as expected, on the 6.1 kb *Hind*III fragment. This 7.2 kb *Hind*III fragment also hybridized to the IS5 probe (data not shown). It now appears that in most K12 strains, the Sqx segment is part of the 7.2 kb *Hind*III fragment, which carries a different IS5 insertion than that inserted into *rfb*. When a 261 bp *Eco*RV fragment from the *rfb* sequence absent in pMN1 (fragment 9, Fig. 2) was used as probe, it hybridized to the 6.1 kb *Hind*III fragment for all the above K12 strains, including the 6.1 kb *Hind*III fragment in K351 (data not shown). Sequencing of the end of the 6.1 kb *Hind*III fragment from K351 confirmed the hybridization result and showed the 4 bp duplication of the IS5 insertion site TTAA (Fig. 6). These results indicate that the *rfb-50* mutation present in most K12 strains comprises a simple IS5 element inserted in the downstream end of the *rfb* gene cluster. All K12 strains carry the Sqx sequence adjacent to a different IS5 element.

We suggest that the situation in pMN1 arose by genetic rearrangement such as deletion or recombination (causing inversion) involving the two IS5 elements. However, we have not looked at K12 strain AB311 (the source of pMN1) and thus we are unable to comment on the situation in that strain. The coincidence that the *Hind*III site upstream of the IS5 elements is at 657 bp in Sqx and 603 bp in *rfb* meant that there was only 54 bp difference in the approximately 6.1 kb *Hind*III segments from pMN1 and Kohara phage K351 (and all *rfb-50* K12 strains) which hybridized to IS5 probe, leading to delay in proper interpretation of the data.

orf224 and *orf264*

The deduced amino acid sequences of *orf224* and *orf264* were examined: there were no significant potential

transmembrane segments found in either, and the predicted secondary structures consisted of alternating α helices and β sheets. The functions of the proteins are not yet known, and their amino acid sequences do not match those of any genes in the databases. There is no obvious start site for *orf224*, which probably begins before the *Pst*I site.

Origins of K12 strain WG1

In this paper we have presented data which enable us to reconstruct the original *rfb* region and express the O antigen of *E. coli* K12, which, until 1993, had probably not been expressed in any K12 strain since at least the 1940s.

We have looked at nine strains of K12: WG1 and EMG2, considered to be wild-type strains, and C600, W3110, J5-3, CR63, KL983, 58-161F⁻-S^R and W1485, derived from several early mutants (Bachmann, 1972). All but WG1 carry *rfb-50*. WG1 was the name given to the wild-type K12 strain in Lederberg's Wisconsin laboratory: Dr B. J. Bachmann has kindly investigated the origins of WG1, and finds that WG1 at Wisconsin came from at least two sources. M. L. Morse obtained a stock directly from C. E. Clifton in 1948 and took it with him when he went to Lederberg's laboratory in 1951. Lederberg also obtained a K12 strain directly from Clifton in 1951 probably as a replacement for a stock obtained earlier, presumably via E. L. Tatum. We understand that all stocks of wild-type K12 in the Wisconsin laboratory were known as WG1, with no indication in the records that they were distinguished. Strains 58-161 and C600, which carry the *rfb-50* mutation, derive from early Tatum mutants 58 and 679, respectively (Gray & Tatum, 1944), and the Tatum stock may have been the only one used up to 1951, other than in the Morse laboratory. The Tatum K12 stock must then have carried *rfb-50* while one or both of the strains sent from Stanford by Clifton in 1948 and 1951, both in the Lederberg Wisconsin laboratory from 1951, probably carried the *rfb-51* mutation. W1485 is derived from WG1, but clearly came from a different stock of WG1 than that now in the *E. coli* Genetic Stock Center. It may be that two or more K12 stocks were kept at Stanford and that they independently developed *rfb* mutation or, alternatively, one or more of the strains given out by Clifton were smooth and mutated afterwards. It seems unlikely that a smooth form will be found in the Tatum line as we have checked several early strains, but we have not as yet been able to check the other line(s).

The other *rfb* mutation known in *E. coli* K12 is '*rfbD1*', which was found in the early Tatum strain 697-680 (Bachmann, 1987; Nikaido *et al.*, 1965). This mutation can be complemented by the *rfbC* gene from *Shigella dysenteriae* 1 (Klena & Schnaitman, 1993), and should therefore be known as *rfbC1*: it is not related to *rfb-50* or *rfb-51*. Strain C600, which was derived from strain 697-680, has '*rfbD1*' and *rfb-50*, and so it was not a candidate strain for complementation of *rfb-50* with the plasmids cloned from WG1. Strains W1485 and EMG2, which are from

different lines of K12, were therefore chosen for the complementation experiments presented above.

E. coli K12 seems certain to be the first organism with a fully sequenced genome. This emanates from the choice of K12 by Tatum for his studies on gene enzyme relationships. *E. coli* is a clonal species and clones can be niche-adapted (Reeves, 1992). Thus, although K12 is well suited to study of the generic properties of *E. coli* and of life itself, part of the genome may be related to adaptation to a specific niche. Presence of complete LPS is almost universal in *E. coli* on isolation, and O-antigen specificity may well be important in niche adaptation.

It is now possible to restore *orf264* in strains such as EMG2 by recombination from WG1 and the *rfb-51* mutation in K12. There may well be other significant mutations which have accumulated since 1922, but as there are related strains which are very similar in all but the *rfb* region (D. Liu & P. Reeves, unpublished data), it should be possible to determine this as occasion demands. The study of *E. coli* K12 and its fully sequenced genome need not be limited by an inability to reconstruct the original sequence in those regions where it has changed since isolation.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Australian Research Council. We especially thank Dr B. J. Bachmann for her very useful advice about the history of strain *E. coli* K12 and for providing K12 strains for this study, a referee who prompted us to look at a range of K12 strains, and Dr Uffe B. Skov Sørensen of The International *Escherichia* and *Klebsiella* Centre, Statens Seruminstitut in Copenhagen for typing strain P4863. We also acknowledge the preliminary work carried out by students Maria Romeo and Paul Bernard, and Dr Antje Haase for assistance with the manuscript.

REFERENCES

- Al-Hendy, A., Toivanen, P. & Skurnik, M. (1991). Expression and cloning of *Yersinia enterocolitica* O:3 *rfb* gene cluster in *Escherichia coli* K12. *Microb Pathog* **10**, 47–59.
- Altschul, A. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Appleyard, R. K. (1953). Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K-12. *Genetics* **39**, 440–452.
- Appleyard, R. K., McGregor, J. F. & Baird, K. M. (1956). Mutation to extended host range and the occurrence of phenotypic mixing in the temperate coliphage. *Virology* **2**, 565–574.
- Bachmann, B. J. (1972). Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* **36**, 525–557.
- Bachmann, B. J. (1983). Linkage map of *Escherichia coli* K-12, edition 7. *Microbiol Rev* **47**, 180–230.
- Bachmann, B. J. (1987). Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, vol. 2, pp. 1190–1219. Edited by F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.
- Bachmann, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol Rev* **54**, 130–197.
- Barcak, G. J. & Wolf, R. E., Jr (1988). Growth-rate-dependent expression and cloning of *gnd* alleles from natural isolates of *Escherichia coli*. *J Bacteriol* **170**, 365–371.
- Bastin, D. A., Romana, L. K. & Reeves, P. R. (1991). Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* 0111 strain. *Mol Microbiol* **5**, 2223–2231.
- Bouffard, G., Ostell, J. & Rudd, K. E. (1992). GeneScope: a relational database of *Escherichia coli* genomic map data for Macintosh computers. *Comput Applic Biosci* **8**, 563–567.
- Brown, P. K., Romana, L. K. & Reeves, P. R. (1991). Cloning of the *rfb* gene cluster of a group C2 *Salmonella* strain: comparison with *rfb* regions of groups B and D. *Mol Microbiol* **5**, 1873–1881.
- Calos, M. P. & Miller, J. H. (1980). Transposable elements. *Cell* **20**, 579–595.
- Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol* **47**, 45–148.
- Clowes, R. C. & Rowley, D. (1954). Some observations on linkage effects in genetic recombination in *Escherichia coli* K-12. *J Gen Microbiol* **11**, 250–260.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.
- Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**, 125–142.
- Engler, J. A. & van Bree, M. P. (1981). The nucleotide sequence and protein-coding capability of the transposable element IS5. *Gene* **14**, 155–163.
- Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* **170**, 827–842.
- Gish, W. & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nature Genet* **3**, 266–272.
- Gray, C. H. & Tatum, E. L. (1944). X-ray induced growth factor deficiencies in bacteria. *Proc Natl Acad Sci USA* **30**, 404–410.
- Hayes, W. (1953). Observations on a transmissible agent determining sexual differentiation in *Bacterium coli*. *J Gen Microbiol* **8**, 72–88.
- Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Mol Microbiol* **5**, 695–713.
- Klein, P., Kanehisa, M. & Delisi, C. (1985). The detection and classification of membrane spanning proteins. *Biochim Biophys Acta* **815**, 468–476.
- Klena, J. D. & Schaitman, C. A. (1993). Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. *Mol Microbiol* **9**, 393–402.
- Kohara, Y., Akiyama, K. & Isono, K. (1987). The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495–508.

- Lederberg, E. M. & Lederberg, J. (1953).** Genetic studies of lysogenicity in *Escherichia coli*. *Genetics* **38**, 51–64.
- Lederberg, J. & Tatum, L. (1946).** Gene recombination in *Escherichia coli*. *Nature* **158**, 558.
- Low, B. (1973).** Rapid mapping of conditional and auxotrophic mutants of *Escherichia coli* K-12. *J Bacteriol* **113**, 798–812.
- Miller, R. D., Dykhuizen, D. E. & Hartl, D. L. (1988).** Fitness effects of a deletion mutation increasing transcription of the 6-phosphogluconate dehydrogenase gene in *Escherichia coli*. *Mol Biol Evol* **5**, 691–703.
- Nasoff, M. S. & Wolf, R. E., Jr (1980).** Molecular cloning, correlation of genetic and restriction maps, and determination of the direction of transcription of *gnd* of *Escherichia coli*. *J Bacteriol* **143**, 731–741.
- Nasoff, M. S., Baker, H. V., II & Wolf, R. E., Jr (1984).** DNA sequence of the *Escherichia coli* gene, *gnd*, for 6-phosphogluconate dehydrogenase. *Gene* **27**, 253–264.
- Neuhard, J. & Thomassen, E. (1976).** Altered deoxyribonucleotide pools in P2 educants of *Escherichia coli* K-12 due to deletion of the *ded* gene. *J Bacteriol* **126**, 999–1001.
- Nikaido, H., Nikaido, K. & Rapin, A. M. C. (1965).** Biosynthesis of thymidine diphosphate L-rhamnose in *Escherichia coli* K-12. *Biochim Biophys Acta* **3**, 548–551.
- Prehm, P., Schmidt, G., Jann, B. & Jann, K. (1976).** The cell-wall lipopolysaccharide of *Escherichia coli* K-12. *Eur J Biochem* **70**, 171–177.
- Reeves, P. R. (1992).** Variation in O antigens, niche specific selection and bacterial populations. *FEMS Microbiol Lett* **100**, 509–516.
- Reisner, A. H., Bucholtz, C. A., Smelt, J. & McNeil, S. (1993).** Australian national genetic information system. In *Proceedings of the 26th Annual Hawaii International Conference in Systems Sciences*, vol. 1, pp. 595–602. Edited by T. N. Mudge, V. Milutinovic & L. Munter. Los Alamos: IEEE Computer Society.
- Rudd, K. E., Bouffard, G. & Miller, W. (1992).** Computer analysis of *E. coli* restriction map. In *Genome Analysis*, vol. 4, pp. 1–38. Edited by K. E. Davies & S. M. Tilghman. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Saiki, R. K., Gelfand, D. H., Stofell, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988).** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977).** DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Schmidt, G. (1973).** Genetical studies on the lipopolysaccharide structure of *Escherichia coli* K12. *J Gen Microbiol* **77**, 157–160.
- Stevenson, G., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991).** The *cps* gene cluster of *Salmonella* strain LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the *rfb* gene cluster. *Mol & Gen Genet* **227**, 173–180.
- Tatum, E. L. & Lederberg, J. (1947).** Gene recombination in bacterium *Escherichia coli*. *J Bacteriol* **53**, 673–684.
- Valvano, M. A. & Crosa, J. H. (1989).** Molecular cloning and expression in *Escherichia coli* K-12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E. coli* O7:K1. *Infect Immun* **157**, 937–943.

Received 24 June 1993; revised 24 August 1993; accepted 9 September 1993.