JOURNAL OF BACTERIOLOGY, Mar. 1993, p. 1412–1422
0021-9193/93/051412-11$02.00/0

Vol. 175, No. 5

Molecular Analysis of the 3,6-Dideoxyhexose Pathway Genes of *Yersinia pseudotuberculosis* Serogroup IIA

ANNETTE C. KESSLER, ANTJE HAASE, AND PETER R. REEVES*

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

Received 27 August 1992/Accepted 3 January 1993

***Salmonella enterica* and *Yersinia pseudotuberculosis* are the only examples in nature known to use a variety of 3,6-dideoxyhexose derivatives as O antigen constituents. To allow a comparison of the responsible biosynthetic genes of the two organisms, we have sequenced a section of the *Y. pseudotuberculosis* serogroup IIA *rfb* region that contained the genes for the abequose biosynthetic pathway. Comparison of the identified genes with the *rfb* region of *S. enterica* LT2 showed that the two dideoxyhexose pathway gene clusters are related. The arrangement of the genes was largely conserved, and the G+C compositions of the two DNA regions were strikingly similar; however, the degree of conservation of nucleotide and protein sequences suggested that the two gene clusters have been evolving independently for considerable time. Hybridization experiments showed that the dideoxyhexose pathway genes are widespread throughout the various serogroups of *Y. pseudotuberculosis*.**

3,6-Dideoxyhexoses (DDH) are a group of unusual, highly immunodominant monosaccharides found almost exclusively within the O antigen component of lipopolysaccharide (LPS). Of eight possible derivatives, only five have so far been found to occur naturally: four of them (abequose, paratose, tyvelose, and colitose) within strains of *Salmonella enterica*, and all five (including ascarylose) within strains of *Yersinia pseudotuberculosis*. Isolated *Escherichia coli* and *Citrobacter* strains are also known to contain DDH, although in these cases the repertoire is limited to the formation of colitose or abequose, respectively. Only in one situation, in the parasitic *Parascaris* worm, has a DDH (in this case ascarylose) been found outside LPS (25). Within *Y. pseudotuberculosis*, DDH are particularly widespread as O antigen constituents. The antigenic typing scheme of *Y. pseudotuberculosis* has recently been revised (1), suggesting the division into seven major serogroups (I to VII). This division largely reflects the distribution of DDH derivatives throughout the species: group I and III strains possess paratose, group II strains possess abequose, group IV strains possess tyvelose, group VA strains possess ascarylose, and group VI and VII strains possess colitose (20, 35); group VI strains additionally incorporate yersiniose, a DDH-related sugar, into their LPS, while group VB strains contain 6-deoxy-L-altrose instead of a DDH. The assignment of strains containing similar DDH derivatives to different serogroups (e.g., I and III or VI and VII), as well as a further subdivision of some groups into subgroups (IA and -B; IIA, -B, and -C; IVA and -B; and VA and -B), is due to differences in sugar composition and arrangement of the remaining O unit (35).

In both *S. enterica* and *Y. pseudotuberculosis*, DDH are formed by the same biosynthetic-reaction sequences. Four DDH, abequose, tyvelose, paratose, and ascarylose, are synthesized by a common pathway proceeding from CDP-D-glucose (Fig. 1) (28); the fifth, colitose, is derived from GDP-D-mannose presumably via a similar reaction sequence (11). All enzymes involved in DDH biosynthesis in *S. enterica* are encoded within the *rfb* gene cluster, which is responsible for the formation of the O-specific subunit of the LPS (26).

The *rfb* gene clusters of several DDH-containing *S. enterica* strains have been cloned and sequenced (7, 8, 17, 24, 45, 50), and most of the DDH-related genes have been identified. In all cases investigated, the arrangement of the DDH pathway genes and their relative positions within the *rfb* region are conserved: immediately downstream of the rhamnose pathway genes, a block of four highly conserved genes is found, comprising *rfbF* and *rfbG* as well as two open reading frames (ORFs), orf7.6 and orf10.4, which are thought to correspond to the postulated DDH pathway genes *rfbI* and *rfbH* (8). Abequose-forming strains of serogroups B and C2 (serovars typhimurium and muenchen, respectively) were shown to possess another gene, *rfbJ*, coding for abequose synthase, which is located adjacent to this highly conserved block; although secondary structure predictions indicated very similar proteins, DNA and amino acid sequences of the two *rfbJ* genes had only low levels of similarity (only 36% identity at the amino acid level [8]). In strains forming tyvelose and paratose (serovars typhi and paratyphi of serogroups D and A, respectively), the *rfbJ* gene was replaced by a paratose synthase gene, *rfbS*, and a tyvelose epimerase gene, *rfbE* (46). The tyvelose epimerase gene in the paratose-producing group A strain, however, was found to be inactive because of a single point mutation. While *rfbS* still showed a low degree of similarity to *rfbJ* at the DNA sequence level, no counterpart to *rfbE* was found in the *rfb* region of the abequose-producing strains of *S. enterica*. Sequence analysis revealed unusually low G+C contents for all *rfb* regions investigated, suggesting a relatively recent transfer of the gene cluster to *S. enterica* from a nonenterobacterial donor with a low G+C content (8, 17, 22, 48).

We have previously reported the cloning of the *Y. pseudotuberculosis* serogroup IIA *rfb* region (19). Hybridization studies of this abequose-producing strain (M85) had shown that at least some of the M85 *rfb* genes are related to DDH pathway genes (*rfbF* and *rfbG*) of *S. enterica* LT2. In this study, we present the sequence and detailed analysis of the DDH pathway gene region of this *Y. pseudotuberculosis* strain. Most of its DDH pathway genes were clearly related

* Corresponding author.

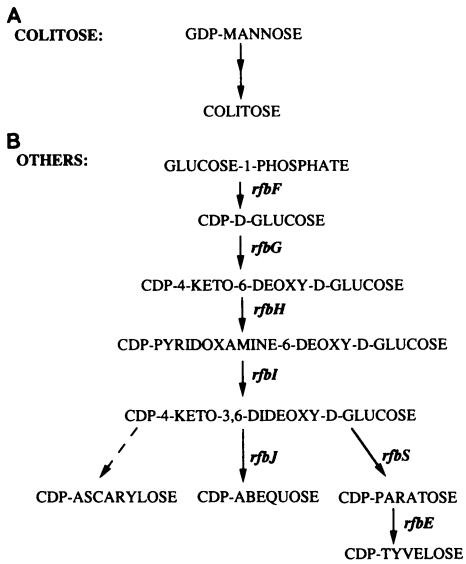

FIG. 1. Biosynthetic pathways for the formation of DDH. (A) Colitose pathway. (B) Pathway for formation of the derivatives abequose, ascarylose, paratose, and tyvelose. The *rfb* genes involved are indicated. They encode enzymes as follows: *rfbF*, glucose-1-P cytidylyltransferase; *rfbG*, CDP-D-glucose oxidoreductase; *rfbH*, CDP-4-keto-6-deoxy-D-glucose-3-dehydrase; *rfbI*, CDP-6-deoxy-$\Delta^{3,4}$ glucoseen reductase; *rfbJ*, CDP-abequose synthase; *rfbS*, CDP-paratose synthase; *rfbE*, CDP-2-tyvelose epimerase (26). The enzymes and the biochemical reaction steps involved in the formation of ascarylose from the common intermediate CDP-4-keto-3,6-dideoxy-D-glucose have not yet been elucidated.

to counterparts in *S. enterica*. G+C contents and codon usage data for the *Y. pseudotuberculosis* DDH pathway genes were found to be very similar to those for *S. enterica*, while the DNA and amino acid sequences have diverged considerably and the *rfbJ* gene had only low levels of similarity with the *rfbJ* genes of *S. enterica* identified previously. Possible histories of the DDH pathway genes are discussed. Hybridization experiments using the M85 DDH pathway genes as probes demonstrated widespread distribution of these genes throughout *Y. pseudotuberculosis*.

MATERIALS AND METHODS

Bacterial strains. *E. coli* K-12, *S. enterica*, and *Y. pseudotuberculosis* strains used are listed in Table 1. All strains were cultivated in NB broth as described by Maniatis et al. (27); *Y. pseudotuberculosis* was grown at 30°C, while all other strains were grown at 37°C.

Enzymes and chemicals. DNA polymerase I, RNase A, Klenow polymerase, exonuclease III, S1 nuclease, restriction enzymes, T4 DNA ligase, *Taq* polymerase, and deoxy- and dideoxynucleotide mixes were from Pharmacia LKB Biotechnology, Uppsala, Sweden, or Boehringer Mannheim Biochemicals, Indianapolis, Ind.; DNase I and other chemicals were from Sigma Chemical Co. and Ajax Chemicals, Sydney, Australia. Reinforced nitrocellulose membrane was from Schleicher & Schuell. An Applied Biosystems 370A DNA sequencer and dye-labelled M13 universal and reverse primers were from Applied Biosystems Inc. Radiochemicals ([α-^{32}P]dCTP) were from Bresatec Ltd., Adelaide, Australia.

Phages and plasmids. Plasmid pPR1197, containing the complete functional *rfb* region of *Y. pseudotuberculosis* M85 in the cosmid vector pPR691 (Fig. 2A) (19), was used as a DNA source for most subcloning procedures. Fragments for sequencing were cloned into either the plasmid vector pT7T3 19U (Pharmacia) or the phage vectors M13mp18 and M13mp19 (31). Helper phage M13KO7 was obtained from Pharmacia. DNA fragments to be used as hybridization probes were subcloned into pUC18 (51).

DNA techniques. Plasmid DNA preparation, single-stranded DNA preparation for clones in M13 vectors, agarose gel electrophoresis, radioactive labelling of DNA, autoradiography, ligation, and bacterial transformation using $CaCl_2$ were performed as described by Maniatis et al. (27). Transformation of *E. coli*-derived plasmid DNA into *S. enterica* M6 could not be carried out directly because of restriction systems in the *S. enterica* host. *E. coli*-derived plasmid DNA was, therefore, first transformed into the restriction-negative *S. enterica* P9121; the plasmid isolated from this strain could then easily be introduced into M6.

Propagation of M13KO7 and preparation of single-stranded DNA from pT7T3-derived deletion clones were carried out as described by the supplier, Pharmacia, or alternatively, by using the protocol of Blondel and Thillet (6). Hybridization conditions were those outlined by Howley et al. (16) for high stringency: the hybridization solution contained 50% (vol/vol) formamide, and hybridization was carried out at 37°C. The filter was washed three times for 5 min each time at room temperature in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS), then for 1 h in 1× SSC–0.1% SDS at 56°C, and finally for 15 min in 0.1× SSC–0.1% SDS at 56°C.

Sequencing strategy. A 7-kb *Pvu*II fragment covering the DNA region thought to contain most of the *Y. pseudotuberculosis* abequose pathway genes (as indicated by previous hybridization experiments [19]) was isolated from pPR1197 and ligated into the *Hin*dII polylinker restriction site of pT7T3 19U. Plasmids pPR1224 and pPR1225 were isolated; they carried the desired fragment in opposite orientations with respect to the universal priming site to allow sequencing of both DNA strands. An overview of the subcloning and sequencing strategy is given in Fig. 2B. By using *Bam*HI and *Sac*I as the 5′ and 3′ restriction enzymes, respectively, nested sets of deletions were then created in each plasmid with exonuclease III according to the protocol of Henikoff (15); plasmid pT7T3 19U carries the IG (intergenic) region from phage f1 which allows derivatives to be packaged as single-stranded DNA during superinfection with helper phage M13KO7 (47). Gaps in the assembled sequence were closed by subcloning suitable DNA restriction fragments into the sequencing vector M13mp18 or M13mp19 to yield the complete sequence from position 2.7 to position 9.7; the DNA sequence from position 1.75 to position 2.7 was also determined by subcloning into M13, while the area from 9.7 to 10.55 was sequenced by using DNA fragments subcloned into pUC18 as double-stranded templates. *E. coli* NM522 was used as the cloning host in all experiments involving pUC18- and pT7T3 19U-derived clones, while *E. coli* JM101 was used for all M13 clones. All DNA preparations were sequenced by the chain termination technique of Sanger et al. (37) by using fluorescent dye-labelled M13 universal or reverse primers and running the reaction mixture on an Applied Biosystems 370A sequencer.

Analysis of sequence data. Sequence data were analyzed by using the Australian National Genomic Information Service at Sydney University, which incorporates several sets of programs. Sequences were assembled by using the program SAP (40, 42); molecular weight and G+C content were

TABLE 1. Bacterial strains

Strain	Laboratory stock no.	Characteristics	Source or reference
Y. pseudotuberculosis			
	M85	Serotype IIA	D. Hughes, New South Wales Dairy Corporation laboratory
H102/88	M443	Serotype IA	S. Aleksic, Institute for Hygiene, Hamburg, Germany
H892/87	M444	Serotype IA	S. Aleksic
H749/89	M445	Serotype IB	S. Aleksic
H376/89	M446	Serotype IB	S. Aleksic
H165/891	M448	Serotype IIA	S. Aleksic
H1779	M449	Serotype IIB	S. Aleksic
H713/86	M451	Serotype III	S. Aleksic
H1091/90	M452	Serotype IVA	S. Aleksic
H1132/90	M453	Serotype IVA	S. Aleksic
H715/86	M454	Serotype IVB	S. Aleksic
H717/86	M455	Serotype IVB	S. Aleksic
H719/86	M456	Serotype VA	S. Aleksic
H1092/90	M457	Serotype VA	S. Aleksic
H450/86	M458	Serotype VB	S. Aleksic
H1117/90	M459	Serotype VB	S. Aleksic
H720/86	M460	Serotype VI	S. Aleksic
H1098/90	M461	Serotype VI	S. Aleksic
H455/86	M462	Serotype VII	S. Aleksic
H143/84	M463	Serotype IIA	S. Aleksic
H130/87 S	M464	Serotype IIA	S. Aleksic
H125/87 S	M465	Serotype IIB	S. Aleksic
H62/87 S	M466	Serotype IIC	S. Aleksic
H172/87 S	M467	Serotype IIC	S. Aleksic
H302/89	M468	Serotype III	S. Aleksic
H97/88	M469	Serotype III	S. Aleksic
H144/86	M470	Serotype VA	S. Aleksic
H2/87 S	M471	Serotype VA	S. Aleksic
H14/87 S	M472	Serotype VB	S. Aleksic
H132/87 S	M473	Serotype VB	S. Aleksic
H207/87 S	M474	Serotype VII	S. Aleksic
H721/86	M475	Serotype VII	S. Aleksic
E. coli			
JM101	P2398		51
P4554		JM109 carrying pPR1197	19
GB23152	P3898	*lacZ trpA kdgR recA1 rpsL25 hsdR trpR* Δ(*edd-zwf*)*22* Δ(*attA-rfbD*)	3
NM522	P4442	Δ*hsd-5* Δ(*lac-pro*)	Pharmacia
P4580		NM522 carrying pPR1224	This study
P4581		NM522 carrying pPR1225	This study
S. enterica			
P9029		*S. enterica* LT2 Δ(*his-rfb*)*388*	30
LB5000	P9121	*S. enterica* LT2 *leu hsdL* (r^- m^+) *trpD2 rpsL120 ilv-452 metE551 metA22 hsdA*(r^- m^+) *hsdB*(r^- m^+)	9
M6		*S. enterica* serovar dublin	S. Dixon, Institute of Medical and Veterinary Science, Adelaide, Australia; 1983

determined by using the program NIP (41); and secondary protein structures, hydrophobicity, and charge distribution were calculated by the program CHOU (GCG package [10]). The RNY (purine:*N*-pyrimidine) preference method of Shepherd (39) was used to identify ORFs. To allow DNA and amino acid sequence alignments, the programs GAP, BESTFIT (GCG package [10]), and SEQA (18) were used. Parameters for running BESTFIT were as follows: gap weight, 5.0; gap length weight, 0.3. Promoter search programs were generously provided by M. C. O'Neill. ALIGNIC2 searched for promoters on the basis of information content (33). MOLBVHI3 used the information content Berg-von Hippel function, and the program VONHIPIC was used to calculate the index for each potential promoter, as described by O'Neill (32).

Y. pseudotuberculosis nucleotide sequence data available from the GenBank and EMBL DNA sequence collections were extracted, noncoding sequences were removed, and the G+C contents of the three codon positions were calculated as described by Sueoka (44). All genes of the two subspecies (*Y. pseudotuberculosis* subsp. *pseudotuberculosis* and *Y. pseudotuberculosis* subsp. *pestis*) available at

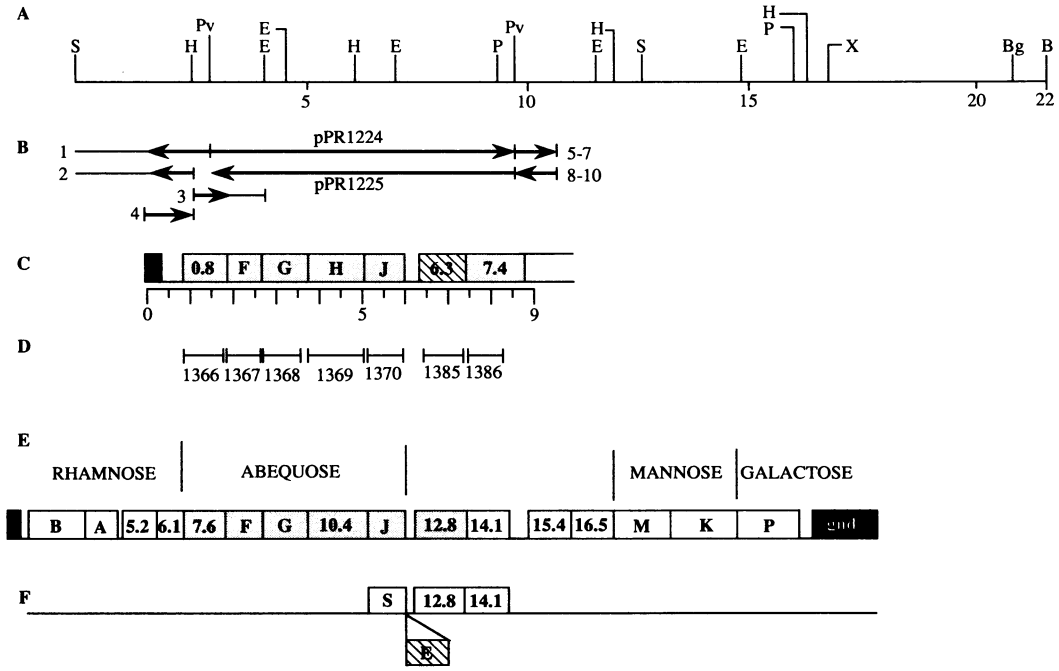

FIG. 2. (A) Restriction map of plasmid pPR1197, which contains the complete *Y. pseudotuberculosis* M85 *rfb* region (19). B, *Bam*HI; S, *Sal*I; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; Pv, *Pvu*II; X, *Xho*II. Numbers represent length in kilobases. (B) Sequencing strategy. The fragments are aligned with panel A. Some of the restriction sites used for subcloning of the smaller fragments were deduced from the DNA sequence and are not shown in panel A. A total of 8.8 kb (from positions 1.75 to 10.55) was sequenced. The *Pvu*II-*Pvu*II fragment (positions 2.7 to 9.7) was cloned into pT7T3 19U in both orientations with respect to the universal priming site; the two resulting plasmids were designated pPR1224 and pPR1225 and sequenced by using nested sets of deletions. Fragments 1 (*Sal*I-*Pvu*II; positions 0 to 2.7), 2 (*Sal*I-*Hin*dIII; positions 0 to 2.45), 3 (*Hin*dIII-*Eco*RI; positions 2.45 to 4.05), and 4 (*Hin*dII-*Hin*dIII; positions 1.75 to 2.45) were cloned into M13mp18 or M13mp19 and sequenced by using single-stranded templates. Arrows denote the direction and the length of the sequence obtained from a clone; sections of the clones that have not been sequenced are indicated by thinner lines. Fragments 5 (*Pvu*II-*Eco*RI; positions 9.7 to 11.2), 6 (*Hin*dII-*Eco*RI; positions 9.8 to 11.2), 7 (*Nde*-*Eco*RI; positions 10.2 to 11.2), 8 (*Nsi*I-*Pvu*II; positions 10.55 to 9.7), 9 (*Nde*I-*Pvu*II; positions 10.2 to 9.7), and 10 (*Hin*dII-*Pvu*II; positions 9.8 to 9.7) were cloned into pUC18 and sequenced by using double-stranded templates to give the DNA sequence of the region from 9.7 to 10.55 kb. (C) *Y. pseudotuberculosis* M85 (serogroup IIA) DDH pathway genes. Areas related to the *S. enterica* LT2 *rfb* region are stippled. Areas related to *S. enterica* Ty2 only are hatched. DNA regions outside *rfb* are indicated in black. Previous mutagenesis data indicated that the left-hand end of the M85 *rfb* region was approximately between kilobase positions 35 and 37 of the original cosmid clone pPR981 (19), which corresponds to kilobase positions 0 to 2 in the DNA sequence. Sequence analysis showed extensive noncoding DNA sequences upstream of orf0.8, indicating that the M85 *rfb* cluster, in fact, begins with orf0.8. The scale (in kilobases) used in this figure is based on DNA sequence data, which do not include the DNA region from position 0 to position 1.75 of the plasmid map in panel A. (D) DNA fragments used as probes. All fragments were subcloned into pUC18, in most cases by using restriction sites identified from the DNA sequence. The plasmids are pPR1366 (*Pvu*II-*Sca*I; 0.7 kb), pPR1367 (*Sty*I-*Sty*I; 0.7 kb), pPR1368 (*Dra*I-*Dra*I; 1 kb), pPR1369 (*Hin*dII-*Bal*I; 1.3 kb), pPR1370 (*Nsi*I-*Bal*I; 0.9 kb), pPR1385 (*Hin*dII-*Sca*I; 0.8 kb), and pPR1386 (*Kpn*I-*Pvu*II; 0.6 kb). (E) The *rfb* region of *S. enterica* LT2 (group B). Areas related to the *Y. pseudotuberculosis* M85 *rfb* region are stippled. (F) The *rfb* region of *S. enterica* Ty2 (group D). Only areas differing from *S. enterica* LT2 *rfb* are shown; the *rfbE* gene homologous to the M85 orf6.3 is hatched.

the time of data preparation were used (*kim-5*, *lcrD*, *lcrE*, *lcrF*, *lcrG*, *lcrH*, *lcrV*, *ompH*, *psaA*, *rplC*, *rplD*, *rplW*, *rplB*, *rpsS*, *yopA*, *yopE*, *yopH*, and four unidentified ORFs). Accession codes of the DNA sequences are YEPKIM5A, YEPLCRD, YEPLCRGVHP, YEPOMPH, YEPORF, YEPPSAA, YEPRPLDWB, YEPTPA, YEPVPIB1, YEPYOPA, YEPYOP5, YEPYOPH, and PD1THRMRP.

Construction of gene-specific DNA probes. Suitable restriction enzyme sites identified by DNA sequencing were used to subclone part of each of the M85 DDH pathway genes into pUC18 to yield plasmids pPR1366 to -1370, pPR1385, and pPR1386 (Fig. 2D).

Immunological techniques. The presence of an antigenic epitope cross-reacting with O4 antiserum was determined by slide agglutination as described by Leinonen (23). O4-specific antiserum was supplied by Wellcome Diagnostics, Dartford, England.

Staining of 6-phosphogluconate dehydrogenase. Plasmids to be tested for the presence of the *gnd* gene were transformed into *E. coli* GB23152. Extracts of the transformants were prepared as described by Selander et al. (38); 15 μl of each extract was spotted on Whatman 3MM filter and stained with 3 ml of staining solution (composition as suggested by the same authors). A red dye was formed on spots containing a functional 6-phosphogluconate dehydrogenase.

Nucleotide sequence accession number. The DNA sequence reported here has been assigned the GenBank accession number LO1777.

RESULTS AND DISCUSSION

Sequence of the *Y. pseudotuberculosis* M85 abequose pathway region and identification of ORFs. Both strands of the M85 *rfb* region from position 1.75 (*Hin*dII) to position 10.55

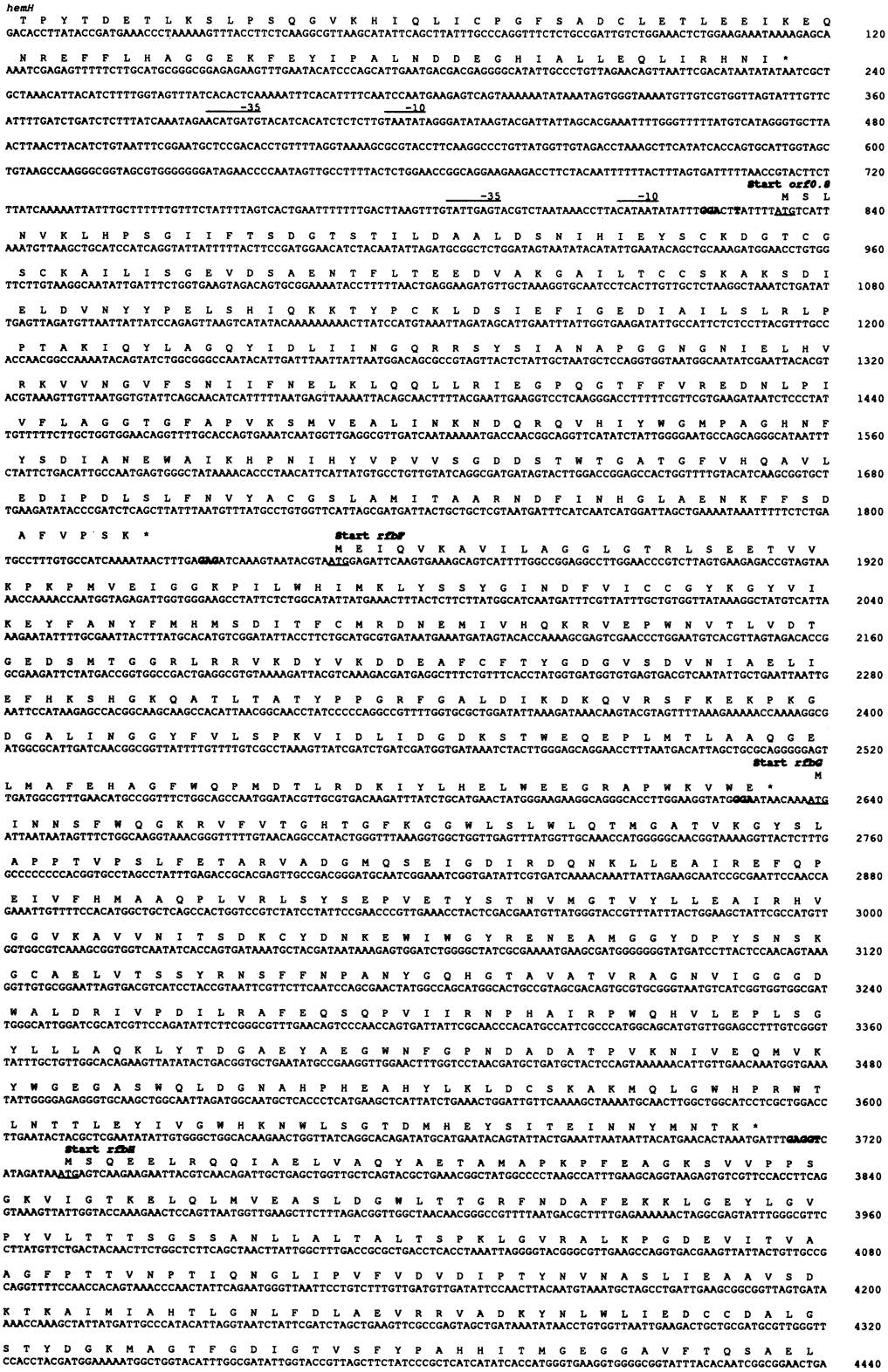

FIG. 3. DNA and deduced amino acid sequences of the *Y. pseudotuberculosis* M85 *rfb* section from map position 1.75 to map position 10.55. Presumptive start codons are underlined, and putative SD sequences are shown in boldface letters.

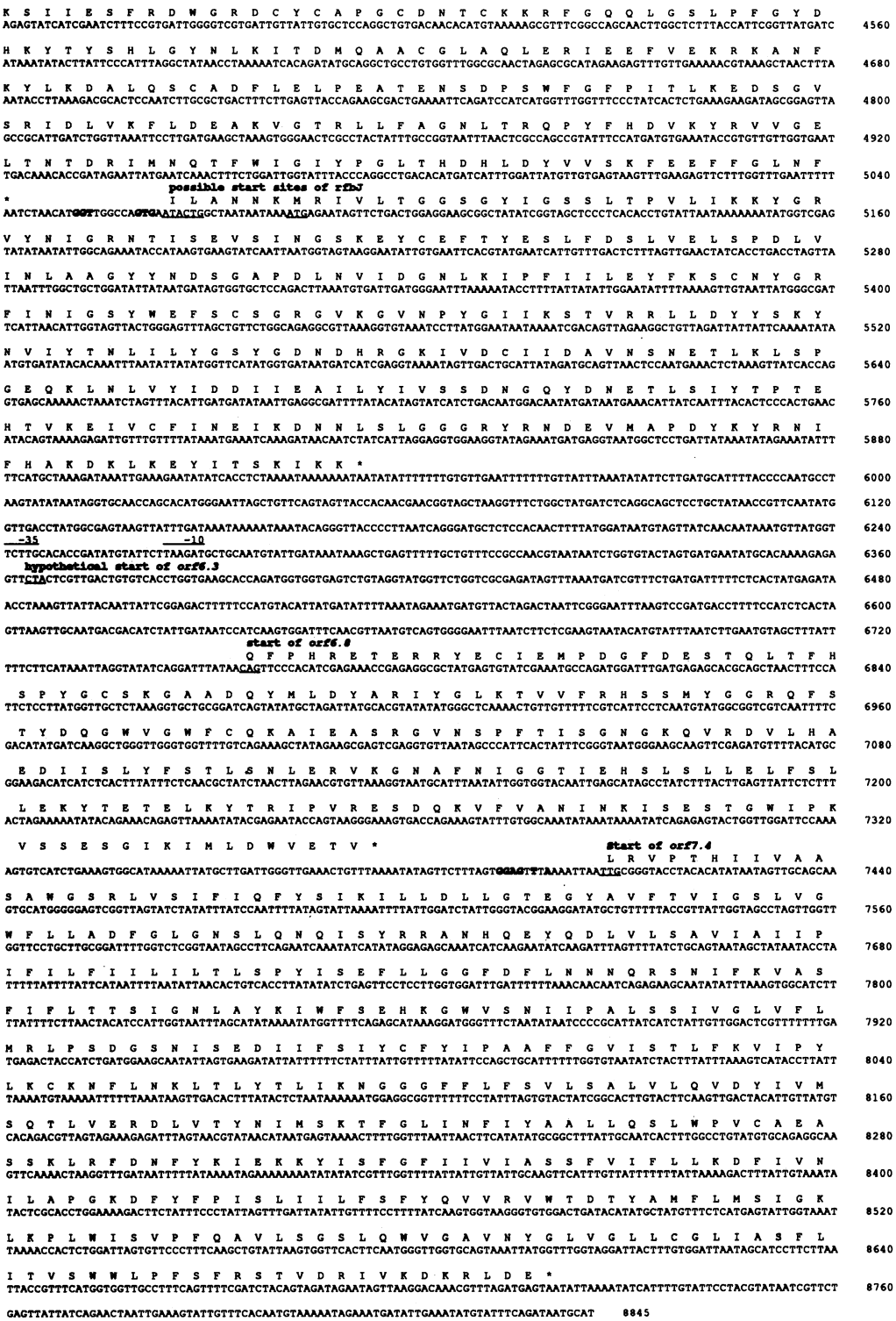

FIG. 3—*Continued.*

(*Nsi*I) relative to the pPR1197 restriction map (Fig. 2A) were sequenced. The complete DNA sequence obtained is shown in Fig. 3. Seven ORFs were found within this region; they were provisionally named orf0.8, orf1.8, orf2.6, orf3.7, orf5.1, orf6.8, and orf7.4 according to the position of the respective start codon within the DNA segment sequenced. Further DNA sequence analysis showed that orf6.8 is part of a longer former ORF starting approximately at position 6.3 that has undergone frameshift mutations such that only part of it is now recognizable as an ORF (see "Identification of

the *Y. pseudotuberculosis* M85 DDH pathway genes" below); it was, therefore, renamed orf6.3 and, as it is probably not functional, is not included in the following general discussion of ORFs. All ORFs found were transcribed in the same direction, from orf0.8 towards orf7.4. Presumptive ATG start codons were found for most ORFs, but the far less frequent initiation codon TTG was also observed (orf7.4). orf3.7 and orf7.4 were preceded by Shine-Dalgarno sequences (SD) located within reasonable distances (9 and 11 bp, respectively) from the respective presumptive start sites (Fig. 3). The DNA sequences preceding orf0.8, orf1.8, orf2.6, and orf5.1 showed only limited complementarity to the 3′ end of the *E. coli* 16S rRNA; however, such sequences have been shown to be functional in some cases (43). The exact initiation site of orf5.1 was not obvious: the only nucleotide sequence resembling an SD-like signal upstream of the first ATG start codon (located at position 5082 of the DNA sequence [Fig. 3]) was too far away (19 bp) to be likely to function, and the large distance (17 bp) of this sequence from the preceding stop codon would probably not allow reinitiation of ribosomal translation. Alternatively, codons ATA and CTG, located further upstream (positions 5064 and 5067, respectively), could be possible, although unlikely, sites for initiation of translation, as these codons have been shown to be functional in some rare cases (12). Both codons are preceded by the same potential SD with optimal or near optimal spacing (9 and 12 bp, respectively [Fig. 3]). This putative SD also seems close enough (8 bp) to the stop codon of the preceding gene to allow ribosomal reinitiation. To facilitate gene and protein comparisons in this study, CTG was used arbitrarily as the initiation codon for orf5.1.

Apart from orf3.7, which terminates with a TGA codon, all complete ORFs were found to end with a TAA stop codon. There are extensive noncoding sequences preceding orf0.8 (597 bp) and orf6.3 (430 bp), possibly indicating regulatory independence of the two gene blocks thus separated (see also "Promoters" below). Shorter noncoding segments were found between all other genes (26 bp between orf0.8 and orf1.8, 4 bp between orf1.8 and orf2.6, 16 bp between orf2.6 and orf3.7, 24 bp between orf3.7 and orf5.1 [if CTG is the initiation codon]), and 31 bp between orf6.3 and orf7.4. Within the region comprising orf0.8 to orf5.1, SD signals and initiation codons are close enough to the preceding stop codons to allow ribosomal reinitiation; the 31-bp gap between orf6.3 and orf7.4, however, appears to be too large to permit this.

Identification of the *Y. pseudotuberculosis* M85 DDH pathway genes. The DNA sequences of several DDH pathway genes of *S. enterica* are known. These genes include *rfbF* (coding for CDP-glucose-1-phosphate cytidylyltransferase), *rfbG* (coding for CDP-glucose-4,6-dehydratase), *rfbJ* (encoding CDP-abequose synthase), *rfbE* (encoding CDP-tyvelose epimerase), and *rfbS* (coding for CDP-paratose synthase [17, 50]). These genes were found clustered in the central section of the *S. enterica rfb* region, together with two additional genes, orf7.6 and orf10.4, with so far unidentified functions (Fig. 2E and F). The presence of orf7.6 and orf10.4 exclusively in DDH-forming strains (8, 17, 22, 24, 48) strongly suggests their involvement in DDH formation, and they are, therefore, assumed to represent the so far unidentified genes *rfbI* (coding for CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase) and *rfbH* (coding for CDP-4-keto-6-deoxy-D-glucose-3-dehydrase).

Comparison of the nucleotide sequences of the seven M85 ORFs with the *S. enterica* LT2 *rfb* gene cluster showed that the first block of five M85 ORFs, orf0.8, orf1.8, orf2.6, orf3.7, and orf5.1, possessed significant similarity to the CDP-abequose pathway genes orf7.6, *rfbF*, *rfbG*, orf10.4, and *rfbJ*, respectively (Fig. 2E) (17); also, orf7.4 of M85 and orf12.8 of LT2 were related on the DNA sequence level. Only orf6.3 did not have a related equivalent within the LT2 *rfb* region; however, it was found to be similar to the *rfbE* gene of *S. enterica* Ty2 (DNA sequence identity values are given in Table 2). Because of the presence of this *rfbE*-like orf6.3 situated between orf5.1 and orf7.4, the overall arrangement of the sequenced M85 *rfb* genes resembles the situation observed with the tyvelose-forming *S. enterica* Ty2 more closely than that with the abequose-forming strain LT2 (Fig. 2C and F). Alignment of the deduced amino acid sequences of the *Y. pseudotuberculosis* M85 and *S. enterica rfb* genes showed similar or even stronger sequence conservation at the protein level than at the DNA level for orf0.8, orf1.8 (*rfbF*), orf2.6 (*rfbG*), orf3.7, and orf6.3; for orf5.1 and orf7.4, amino acid sequence identities to the respective *S. enterica* counterparts were considerably below the DNA sequence similarity levels (Table 2). For *Y. pseudotuberculosis* genes with clear structural or functional similarities to previously identified *rfb* genes of *S. enterica*, the terminology used for the *S. enterica rfb* cluster was adopted (Fig. 2C).

TABLE 2. DNA and amino acid similarities between *S. enterica* and *Y. pseudotuberculosis* M85 DDH pathway genes

Y. pseudotuberculosis gene	*S. enterica* gene[a] (serogroup)	% DNA sequence identity	% Amino acid sequence similarity	% Amino acid sequence identity
orf0.8	orf7.6 (B)	58.9	68.8	51.2
rfbF	*rfbF* (B)	73.5	89.9	80.5
rfbG	*rfbG* (B)	70.3	82.6	72.5
rfbH	orf10.4 (B)	74.3	92.9	87.4
rfbJ	*rfbJ* (B)	46.1	48.4	22.8
rfbJ	*rfbJ* (C2)	42.5	50.2	25.6
rfbJ	*rfbS* (D)	45.1	50.9	24.2
orf6.3[b]	*rfbE* (D)	63.0	77.9	64.2
orf7.4	orf12.8 (B)	51.5	62.3	36.8

[a] *rfbS* and *rfbE* were from *S. enterica* Ty2 (serogroup D); other genes were from *S. enterica* LT2 (serogroup B) and *S. enterica* muenchen (serogroup C2).
[b] For orf6.3 DNA sequence alignments, the DNA sequence between positions 6364 and 7403 was used; for protein sequence comparison, the translation of the rudimentary ORF between positions 6755 and 7373 was used.

The possibility that the protein encoded by orf5.1 is an abequose synthase was supported by the identification of a motif within the amino acid sequence resembling an NAD-binding domain typical for abequose synthases (8, 50) and other NAD-linked dehydrogenases (34), which consists of a β-barrel of six β-sheets connecting α-helices. A functional test confirmed that the orf5.1 gene product exhibits abequose synthase activity: transformation of orf5.1 (as pPR1370 [Fig. 2D]) into a paratose-forming *S. enterica* serovar dublin strain (M6) resulted in the expression of the abequose-specific O4 antigen by the transformant, confirming that orf5.1 is the *rfbJ* gene of strain M85.

The N-terminal amino acid sequence of the *rfbH* gene product, CDP-4-keto-6-deoxy-D-glucose-3-dehydrase, from a *Y. pseudotuberculosis* serogroup VA strain (49) was compared with the deduced polypeptide sequences of the M85 ORFs. It was found to be identical to the N-terminal portion of the orf3.7 gene product, indicating that orf3.7 is the *rfbH* gene of strain M85. We can, therefore, conclude that orf10.4

TABLE 3. G+C contents and P1, P2, and P3 values of the *rfb* genes sequenced

Gene	Content[a]			
	G+C (0.465)	P1 (0.579)	P2 (0.401)	P3 (0.391)
orf0.8	0.376	0.497	0.373	0.259
LT2 orf7.6	0.405	0.525	0.380	0.304
rfbF	0.436	0.564	0.354	0.394
LT2 *rfbF*	0.473	0.592	0.373	0.322
rfbG	0.460	0.565	0.402	0.403
LT2 *rfbG*	0.437	0.579	0.382	0.338
rfbH	0.425	0.552	0.399	0.325
LT2 orf10.4	0.446	0.550	0.404	0.379
rfbJ	0.302	0.386	0.323	0.206
LT2 *rfbJ*	0.320	0.434	0.302	0.216
orf6.3[b]	0.375	0.489	0.391	0.308
Ty2 *rfbE*	0.355	0.437	0.368	0.268
orf7.4	0.308	0.391	0.325	0.242
LT2 orf12.8	0.31	0.382	0.324	0.221

[a] Averages for *Y. pseudotuberculosis* are given in parentheses. Average codon frequencies were calculated as described by Sueoka et al. (44). G+C, overall GC content for the gene(s) analyzed; P1, P2, and P3, corrected GC contents for individual codon positions.

[b] Codon preferences for orf6.3 were calculated by using the rudimentary ORF between sequence positions 6755 and 7373 as a basis.

of *S. enterica* LT2 is also an *rfbH* gene, given its high similarity to orf3.7 (Table 3).

The N-terminal amino acid sequence of the *rfbI* gene product CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase from the same *Y. pseudotuberculosis* group VA strain (14) was compared in a similar manner with the M85 ORFs and *S. enterica* LT2 *rfb* genes; however, no match within either of the regions could be found. Only about 50% of the M85 *rfb* region has been analyzed in this study, and it is therefore possible that the M85 *rfbI* gene is located outside the DNA segment investigated. However, almost all ORFs of the *S. enterica* LT2 *rfb* region have been identified and allocated a function, and orf7.6 remains the only likely candidate for *rfbI* in this strain. The M85 orf0.8 is homologous to orf7.6 of LT2 and is, therefore, still the best candidate for *rfbI* in M85.

A feature common to all *rfb* gene clusters of *S. enterica* studied so far is the presence of a gene that encodes a highly hydrophobic polypeptide with 12 predicted membrane-spanning segments (orf12.8 in strain LT2) (8, 17, 22, 24, 48). The function of this protein is unknown; it may assist in the export of the O antigen to the cell periphery. The polypeptide encoded by orf7.4 in *Y. pseudotuberculosis* M85 was also predicted to have 12 transmembrane segments, indicating similar physiological roles for the orf12.8 and orf7.4 proteins.

The initially identified orf6.8 (extending from sequence position 6755 to position 7372) was found to have significant similarity with the central part of the tyvelose epimerase gene *rfbE* of *S. enterica* Ty2 (serogroup D) (46). However, when upstream and downstream sequences were included in the DNA alignment, homologies to the *rfbE* gene were obvious up to position 6476 and down to position 7360 (of the 884 bases that could be aligned, 67.8% were identical). The *rfbE* start codon was aligned with position 6364 of the M85 sequence, and with reference to this hypothetical former start position at 6364, orf6.8 was renamed orf6.3. Several mutational events such as base deletions, insertions, and single base changes obviously have led to the destruction of the original M85 ORF down to position 6754 and, again, from position 7372 onward; given the nature and the significant number of the changes observed, it seems unlikely that the orf6.3 gene of strain M85 is still functional.

Relative position of the *Y. pseudotuberculosis* M85 *rfb* region on the chromosome. In *E. coli* and *S. enterica*, the *rfb* region has been shown to be closely linked to the *gnd* gene on the bacterial chromosome, which is immediately downstream of *rfb* (4, 7, 17, 21, 48), at approximately 44 min on the chromosome map of *E. coli* K-12 (2) and 42 min on the *S. enterica* LT2 map (36). To test whether this position is conserved for the *Y. pseudotuberculosis* M85 *rfb* region, the original cosmid pPR981, which contains the M85 *rfb* region plus about 25 kb of DNA downstream and 2.4 kb of DNA upstream (19), was transformed into *E. coli* GB23152 with a deleted *gnd*. Enzyme extracts of the transformants failed to exhibit 6-phosphogluconate dehydrogenase activity, while the wild-type M85 *gnd* gene was clearly functional, suggesting that the M85 *gnd* gene is not located immediately downstream of *rfb*.

A translation of a partially sequenced ORF preceding the M85 *rfb* region on clone pPR1197 (positions 1 to 331 of the DNA sequence) was compared with the GenBank and EMBL amino acid sequence collections and was found to be 68.1% identical to the *E. coli hemH* gene product ferrochelatase (29). In *E. coli*, this gene is located at approximately 11 min on the chromosome (2), at a considerable distance from the *rfb* region. Apparently, the relative position of the *Y. pseudotuberculosis* M85 *rfb* region on the chromosome is not conserved in comparison with *S. enterica* and *E. coli*.

Promoters. Computer searches using the programs provided by M. C. O'Neill (see Materials and Methods) revealed a large number of potential promoters, all with a Berg-von Hippel index too high to be likely to function unless under positive control. The intergenic gaps preceding orf0.8 and orf6.3 are the most likely positions for such regulatory regions. Three potential promoters were found in front of orf0.8. The −10 region of the top-ranked (Berg-von Hippel index, 3.4) promoter is located at 416 bp of the DNA sequence; a second promoter is located very close to the start of orf0.8, with a −10 region positioned at 809 bp (Berg-von Hippel index, 5.4). Within the large intergenic gap preceding orf6.3, one potential promoter at position 6264 (Berg-von Hippel index, 3.4) was found, with a spacing of 85 bp between the transcriptional start site and the CTA codon at the position equivalent to that of the *S. enterica rfbE* ATG codon. Therefore, it seems possible that the M85 *rfb* region is divided into at least two separately regulated segments, contrary to the situation observed with *S. enterica*: the data for all of the *rfb* regions in *S. enterica* investigated suggested transcription of the cluster as a single operon.

G+C content and codon usage. The overall G+C content of an organism is thought to be the result of a long-term bias in the mutation rates from G · C to A · T and A · T to G · C (44); any genes that have been introduced into a given organism would be expected to adapt to the species-specific level over time. We have previously reported that the G+C contents of genes within the *rfb* clusters of many *S. enterica* strains are much lower than the species average of 0.51; we also found their P1, P2, and P3 values, which are the corrected average G+C contents for bases 1, 2, and 3 of the codons used (44), to be characteristic of those observed for

TABLE 4. Distribution of DDH pathway genes among *Y. pseudotuberculosis* serogroups

Serogroup[a]	DDH derivative	Hybridization with probe[b]:						
		orf0.8	*rfbF*	*rfbG*	*rfbH*	*rfbJ*	orf6.3	orf7.4
IA	Paratose	+	+	+	+	−	−	−
IB	Paratose	+	+	+	+	−	−	−
IIA	Abequose	+	+	+	+	+	+	+
IIB	Abequose	+	+	+	+	+	+	+
IIC	Abequose	+	+	+	+	+	+	+
III	Paratose	+	+	+	+	−	−	−
IVA	Tyvelose	+	+	+	+	−	+	+
IVB	Tyvelose	+	+	+	+	−	+	+
VA	Ascarylose	+	+	+	+	−	−	−
VB	None[c]	+	+	−	−	−	−	−
VI	Colitose[d]	+	+	+	+	−	−	−
VII	Colitose	−	−	−	−	−	−	−

[a] All *Y. pseudotuberculosis* strains listed in Table 1 were included.
[b] + and −, hybridization and no hybridization, respectively.
[c] Serogroup VB strains do not possess any DDH but contain 6-deoxy-L-altrose.
[d] Serogroup VI strains contain the octose yersiniose in addition to colitose.

low-G+C content species. These data led to the conclusion that these *rfb* gene clusters evolved in a low-G+C content species before being transferred to *S. enterica* in a comparatively recent event (8, 17, 22, 46, 48). Genes of the M85 *rfb* region analyzed in this study have G+C contents ranging from 0.30 to 0.46, closely resembling the values obtained for the respective *rfb* genes of *S. enterica* LT2; codon frequencies of the M85 genes are also very similar to the *S. enterica* values (Table 3). The range of the G+C contents as well as the codon frequency values overlaps the average values for *Y. pseudotuberculosis*, which possesses a G+C content of 0.465 (5), while they are clearly set apart from the *S. enterica* average, which is 0.51. Therefore, the possibility that genes *rfbF*, *rfbG*, and *rfbH* and, perhaps also, orf6.3 and orf0.8 have originated in *Y. pseudotuberculosis* itself (or its ancestor) clearly has to be considered.

Relationships of the *Y. pseudotuberculosis* M85 and *S. enterica* DDH pathway gene clusters. The DDH pathway genes of *S. enterica* LT2 and *Y. pseudotuberculosis* M85 are clearly homologous gene clusters: the genes themselves are homologous, and the order of the genes is conserved in the two species, indicating descent of the two clusters from a common ancestor. The large number of changes (25 to 54% of the nucleotide residues and 23 to 77% of the amino acid residues) observed indicates that the separation itself is an ancient one, and the presence of genes with different G+C contents suggests that this ancestral gene cluster was itself assembled from more than one source. Particularly, the P3 values for the genes with a G+C content of 0.3 suggest that the recent history of both clusters has been in low-G+C content species and that they were transferred independently to yersinias and salmonellas. We also draw attention to the facts that *rfbJ* shows the most divergence from its homolog in *S. enterica* and that there are two highly divergent forms of *rfbJ* within *S. enterica*. We have no simple explanation for the very high variation of *rfbJ*.

Distribution of DDH pathway genes within *Y. pseudotuberculosis*. A set of probes (pPR1366 to -1370, pPR1385, and pPR1386 [Fig. 2D]) specific for the M85 DDH pathway genes was used to test chromosomal DNA preparations of strains representative of the serogroups of *Y. pseudotuberculosis*; the results are given in Table 4. All strains, except those of serogroups VB and VII, were found to contain the complete set of genes for the common part of the DDH biosynthesis pathway (*rfbF*, *rfbG*, *rfbH*, and orf0.8, which is likely to be *rfbI*). Strains of serogroup VII produce colitose; the colitose pathway is not closely related to the abequose pathway, and, as expected, the group VII strains do not hybridize to the four genes of the common pathway. Strains of serogroup VI contain two 3,6-dideoxysugar derivatives in their O antigen, the hexose colitose and the octose yersiniose, a 3,6-dideoxy-4C-(1-hydroxyethyl)-D-*xylo*-hexose (13). It seems probable that the serogroup VI genes hybridizing to the M85 probes are involved in the formation of yersiniose; this would imply that yersiniose is formed via the same pathway as abequose, tyvelose, paratose, and ascarylose up to the common intermediate CDP-4-keto-3,6-dideoxy-D-glucose, from which the final product yersiniose could be made easily, e.g., by addition of pyruvate under decarboxylation.

Serogroup VB strains contained only orf0.8 and *rfbF*-like genes; these strains possess 6-deoxy-L-altrose as an immunodominant sugar but no DDH. It is not known how 6-deoxy-L-altrose is formed in vivo; the roles of the two genes orf0.8 and *rfbF*, if any, in the biosynthesis of this sugar therefore remain unknown.

orf6.3 equivalents, the like of the *S. enterica* Ty2 gene *rfbE*, were found within all other serogroup II strains (subgroups IIA, IIB, and IIC) tested as well as all tyvelose-producing strains (serogroup IV), which necessarily possess a tyvelose epimerase gene, *rfbE*. The significant homology of orf6.3 to group IV genes supports the idea that the nonfunctional orf6.3 once encoded an epimerase; however, a tyvelose epimerase would be redundant in an abequose-producing strain. It remains speculative whether the inactive orf6.3 gene in abequose-forming *Y. pseudotuberculosis* strains is a leftover of a tyvelose epimerase gene from times when DDH gene regions were assembled from various precursors and has since become superfluous in an abequose context or whether its functional ancestor was able to use CDP-abequose as a substrate to form a so far unreported sixth DDH derivative by epimerization at carbon 2.

No *rfbE* gene was found within any of the paratose-forming strains of *Y. pseudotuberculosis*, indicating that the formation of paratose in these strains is an original trait, not a secondary effect of mutational inactivation of the *rfbE* gene, as shown for paratose-forming strains of *S. enterica* (46). As in *S. enterica*, the final specificity for a particular DDH derivative appears to be based on the alternative presence of the respective synthase genes: only abequose-forming strains were found to possess the *rfbJ* gene.

Overall, the genes of the common part of the DDH pathway were found to be almost ubiquitous within *Y. pseudotuberculosis*. This widespread distribution supports the idea that the ability to form DDH is not a recently acquired trait for *Y. pseudotuberculosis* but has been present in the species from an early point onward in its evolution. Further weight to this argument is added by the fact that *Y. pseudotuberculosis* contains a variety of these rather unusual and rare sugars that is unsurpassed by that of any other organism.

ACKNOWLEDGMENTS

This work was supported by a grant from the Australian Research Council.

We thank S. Aleksic for generously providing the *Y. pseudotuberculosis* strain collection.

REFERENCES

1. **Aleksic, S., G. Suchan, J. Bockemühl, and V. Aleksic.** 1991. An extended antigenic scheme for Yersinia pseudotuberculosis. Contrib. Microbiol. Immunol. **12**:235–238.
2. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
3. **Barcak, G. J., and R. E. Wolf, Jr.** 1988. Growth-rate-dependent expression and cloning of *gnd* alleles from natural isolates of *Escherichia coli*. J. Bacteriol. **170**:365–371.
4. **Bastin, D. A., L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. Mol. Microbiol. **5**:2223–2231.
5. **Bercovier, H., H. H. Mollaret, J. M. Alonso, J. Brault, R. G. Fanning, A. G. Steigerwalt, and D. J. Brenner.** 1980. Intra- and interspecies relatedness of *Yersinia pestis* by DNA hybridization and its relationship to *Yersinia pseudotuberculosis*. Curr. Microbiol. **4**:225–229.
6. **Blondel, A., and J. Thillet.** 1991. A fast and convenient way to produce single stranded DNA from a phagemid. Nucleic Acids Res. **19**:181.
7. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1991. Cloning of the *rfb* gene cluster of a group C2 *Salmonella*: comparison with the *rfb* regions of groups B and D. Mol. Microbiol. **5**:1873–1881.
8. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6**:1385–1394.
9. **Bullas, L. R., and J.-I. Ryu.** 1983. *Salmonella typhimurium* LT2 strains which are r^- m^+ for all three chromosomally located systems of DNA restriction and modification. J. Bacteriol. **156**:471–474.
10. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
11. **Elbein, A. D., and E. C. Heath.** 1965. The biosynthesis of cell wall lipopolysaccharide in *Escherichia coli* II guanosine diphosphate 4-keto-6-deoxy-D-mannose, an intermediate in the biosynthesis of guanosine diphosphate colitose. J. Biol. Chem. **240**: 1926–1931.
12. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
13. **Gorshkova, R. P., V. A. Zubkov, V. V. Isakov, and Y. S. Ovodov.** 1983. Structural features of O-specific polysaccharide from lipopolysaccharide of *Yersinia pseudotuberculosis* VI serovar. Bioorg. Khim. **9**:1068–1073.
14. **Han, O., V. P. Miller, and H. Liu.** 1990. Mechanistic studies of the biosynthesis of 3,6-dideoxyhexoses in *Yersinia pseudotuberculosis*. J. Biol. Chem. **265**:8033–8041.
15. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
16. **Howley, P. M., M. A. Israel, M. F. Law, and M. A. Martin.** 1979. A rapid method for detecting and mapping homology between heterologous DNAs. J. Biol. Chem. **254**:4876–4883.
17. **Jiang, X. M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5**:695–713.
18. **Kanehisha, M. J.** 1982. Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. **10**:183–196.
19. **Kessler, A., L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and genetic characterization of the *rfb* region from *Yersinia pseudotuberculosis* serovar IIA, which determines the formation of the 3,6 dideoxyhexose abequose. J. Gen. Microbiol. **137**:2689–2695.
20. **Komandrova, N. A., R. P. Gorshkova, V. A. Zubkov, and Y. S. Ovodov.** 1989. The structure of the O-specific polysaccharide chain of the lipopolysaccharide of *Yersinia pseudotuberculosis* serovar VII. Bioorg. Khim. **15**:104–110.
21. **Lee, S. J., L. K. Romana, and P. R. Reeves.** 1992. Cloning and structure of group C1 O-antigen (*rfb* gene cluster) from *Salmonella enterica* serovar *montevideo*. J. Gen. Microbiol. **138**:305–312.
22. **Lee, S. J., L. K. Romana, and P. R. Reeves.** 1992. Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain. J. Gen. Microbiol. **138**: 1843–1855.
23. **Leinonen, M.** 1985. Serological methods for the study of bacterial surface antigens, p. 179–206. *In* T. K. Korhonen, E. A. Dawes, and P. H. Mäkelä (ed.), Enterobacterial surface antigens: methods for molecular characterisation. Elsevier Science Publishers, Amsterdam.
24. **Liu, D., N. K. Verma, L. K. Romana, and P. R. Reeves.** 1991. Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J. Bacteriol. **173**:4814–4819.
25. **Lüderitz, O., O. Westphal, A. M. Staub, and H. Nikaido.** 1971. Isolation and chemical and immunological characterization of bacterial lipopolysaccharides, p. 145–233. *In* G. Weinbaum, S. Kadis, and S. J. Ajl (ed.), Microbial toxins, vol. 4. Academic Press, Inc., New York.
26. **Mäkelä, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. I. Elsevier Science Publishers, Amsterdam.
27. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
28. **Matsuhashi, S., M. Matsuhashi, and J. L. Strominger.** 1966. Enzymatic synthesis of cytidine diphosphate 3,6 dideoxyhexoses. I. Overall reactions. J. Biol. Chem. **241**:4267–4274.
29. **Miyamoto, K., K. Nakahigashi, K. Nishimura, and H. Inokuchi.** 1991. Isolation and characterization of visible light-sensitive mutants of *Escherichia coli* K12. J. Mol. Biol. **219**:393–398.
30. **Nikaido, H., M. Levinthal, K. Nikaido, and K. Nakane.** 1967. Extended deletions in the histidine-rough-B region of the *Salmonella* chromosome. Proc. Natl. Acad. Sci. USA **57**:1825–1832.
31. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26**:101–106.
32. **O'Neill, M. C.** 1989. Consensus methods for finding and ranking DNA binding sites—application to *Escherichia coli* promoters. J. Mol. Biol. **207**:301–310.
33. **O'Neill, M. C., and F. Chiafari.** 1989. *Escherichia coli* promoters. II. A spacing class-dependent promoter search protocol. J. Biol. Chem. **264**:5531–5534.
34. **Rossman, M. G., A. Liljas, C. I. Brenden, and L. J. Banaszak.** 1975. Evolutionary and structural relationships among dehydrogenases, p. 61–102. *In* P. D. Boyer (ed.), The enzymes, 3rd ed. Academic Press, Inc., New York.
35. **Samuelsson, K., B. Lindberg, and R. R. Brubaker.** 1974. Structure of O-specific side chains of lipopolysaccharides from *Yersinia pseudotuberculosis*. J. Bacteriol. **117**:1010–1016.
36. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition VII. Microbiol. Rev. **52**:485–532.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51**:873–884.
39. **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. USA **78**:1596–1600.
40. **Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10**:4731–4751.
41. **Staden, R.** 1984. Computer methods to locate signals in nucleic acid sequences. Nucleic Acids Res. **12**:505–519.

42. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14**:217–231.
43. **Stormo, G. D., T. D. Schneider, and L. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.
44. **Sueoka, N.** 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85**:2653–2657.
45. **Verma, N. K., N. B. Quigley, and P. R. Reeves.** 1988. O-antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. J. Bacteriol. **170**:103–107.
46. **Verma, N., and P. Reeves.** 1989. Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D salmonellae. J. Bacteriol. **171**:5694–5701.
47. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
48. **Wang, L., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics **130**:429–443.
49. **Weigel, T. M., L. Liu, and H. Liu.** 1992. Mechanistic studies of the biosynthesis of 3,6-dideoxyhexoses in *Yersinia pseudotuberculosis*: purification and characterization of CDP-4-keto-6-deoxy-D-glucose-3-dehydrase. Biochemistry **31**:2129–2139.
50. **Wyk, P., and P. Reeves.** 1989. Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B salmonellae: homology with galactose epimerase. J. Bacteriol. **171**:5687–5693.
51. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.