JOURNAL OF BACTERIOLOGY, Sept. 1994, p. 5483–5493
0021-9193/94/$04.00+0
Copyright © 1994, American Society for Microbiology
Vol. 176, No. 17

Studies of the Biosynthesis of 3,6-Dideoxyhexoses: Molecular Cloning and Characterization of the *asc* (Ascarylose) Region from *Yersinia pseudotuberculosis* Serogroup VA

JON S. THORSON,† STANLEY F. LO,‡ OLIVIER PLOUX,§ XUEMEI HE, AND HUNG-WEN LIU*

Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

Received 27 December 1993/Accepted 16 June 1994

The 3,6-dideoxyhexoses are found in the lipopolysaccharides of gram-negative bacteria, where they have been shown to be the dominant antigenic determinants. Of the five 3,6-dideoxyhexoses known to occur naturally, four have been found in various strains of *Salmonella enterica* (abequose, tyvelose, paratose, and colitose) and all five, including ascarylose, are present among the serotypes of *Yersinia pseudotuberculosis*. Although there exists one report of the cloning of the *rfb* region harboring the abequose biosynthetic genes from *Y. pseudotuberculosis* serogroup IIA, the detailed genetic principles underlying a 3,6-dideoxyhexose polymorphism in *Y. pseudotuberculosis* have not been addressed. To extend the available information on the genes responsible for 3,6-dideoxyhexose formation in *Yersinia* spp. and facilitate a comparison with the established *rfb* (O antigen) cluster of *Salmonella* spp., we report the production of three overlapping clones containing the entire gene cluster required for CDP-ascarylose biosynthesis. On the basis of a detailed sequence analysis, the implications regarding 3,6-dideoxyhexose polymorphism among *Salmonella* and *Yersinia* spp. are discussed. In addition, the functional cloning of this region has allowed the expression of E_p (α-D-glucose cytidylyltransferase), E_{od} (CDP-D-glucose 4,6-dehydratase), E_1 (CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose-3-dehydrase), E_3 (CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase), E_{ep} (CDP-3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose-5-epimerase), and E_{red} (CDP-3,6-dideoxy-L-*glycero*-D-*glycero*-4-hexulose-4-reductase), facilitating future mechanistic studies of this intriguing biosynthetic pathway.

A large degree of bacterial immunological diversity in gram-negative species is attributed to the portion of the lipopolysaccharides known as the O antigen. Exposed on the cell envelope, the O antigen is the dominant surface entity which is polymorphic and provides the basis for the serological classification of the family *Enterobacteriaceae* (4, 16, 26). Among the large number of monosaccharides found as components of O-specific polysaccharides, derivatives of the 3,6-dideoxyhexoses have drawn special attention because of their highly immunogenic characteristics (22, 43, 45). It has been shown that only a limited number of species of gram-negative bacteria, all in the family *Enterobacteriaceae*, are able to produce these unusual sugars, and, of the eight possible stereoisomers of 3,6-dideoxyhexoses, only five have been identified in nature. Although colitose (3,6-dideoxy-L-*xylo*-hexose) and abequose (3,6-dideoxy-D-*xylo*-hexose, compound I [see Fig. 1]) can be detected in the lipopolysaccharide of some *Escherichia coli* serotypes and *Citrobacter* strains, respectively, the five 3,6-dideoxyhexoses known to occur naturally are present mainly in *Salmonella* and *Yersinia* spp. as major antigenic determinants (23). For example, abequose, colitose, tyvelose (3,6-dideoxy-D-*arabino*-hexose, compound II), and paratose (3,6-dideoxy-D-*ribo*-hexose, compound III) have been found in strains of *Salmonella enterica* (34), while all five, including ascarylose (3,6-dideoxy-L-*arabino*-hexose, compound IV), are present among serotypes of *Yersinia pseudotuberculosis* (1, 13, 22, 31). Considering the evolutionary distance between the genera *Salmonella* and *Yersinia* (7), the simultaneous occurrence of this otherwise unknown polymorphism is rather intriguing.

Inspired by the uniqueness of their occurrence in nature and their intriguing immunological effects, substantial efforts have been devoted to study the biosynthesis of 3,6-dideoxyhexoses (10, 12). It has been shown that they are synthesized via a complex series of enzymatic reactions starting, in most cases, from CDP-D-glucose (compound VI) derived from glucose-1-phosphate (compound V) in an α-D-glucose cytidylyltransferase (E_p)-catalyzed reaction (29, 35). The subsequent steps, exemplified by the formation of CDP-ascarylose (compound IV), involve an irreversible intramolecular oxidation-reduction catalyzed by an NAD^+-dependent CDP-D-glucose 4,6-dehydratase (E_{od}) (40, 48). The nascent product, CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose (compound VII), is then converted to CDP-3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose (compound VIII) in two consecutive enzymatic reactions (11) mediated by CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose-3-dehydrase (E_1), a pyridoxamine 5′-phosphate (PMP)-linked iron-sulfur-containing catalyst (36, 37, 42), and CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase (E_3), a [2Fe-2S]-containing flavoprotein (21, 24, 25). As delineated in Fig. 1, the final conversions are carried out by an epimerase (CDP-3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose-5-epimerase [E_{ep}]), which inverts the configuration at C-5, followed by a reductase (CDP-3,6-dideoxy-D-*glycero*-L-*glycero*-4-hexulose-4-reductase [E_{red}])-catalyzed stereospecific reduction at C-4 to give the desired product, CDP-ascarylose (compound IV) (38). As a part of our continuous effort to study the biosynthesis of 3,6-dideoxyhexoses, we have recently cloned the entire *asc* (ascarylose) gene cluster of *Y. pseudotuberculosis*. Three of the six presented *asc* open reading frames

* Corresponding author. Phone: (612) 625-5356. Fax: (612) 626-7541.
† Present address: Department of Chemistry, University of California at Berkeley, Berkeley, Calif.
‡ Present address: Department of Pathology and Internal Medicine, Washington University, St. Louis, MO 63130.
§ Permanent address: Centre National de la Recherche Scientifique, Paris, France.

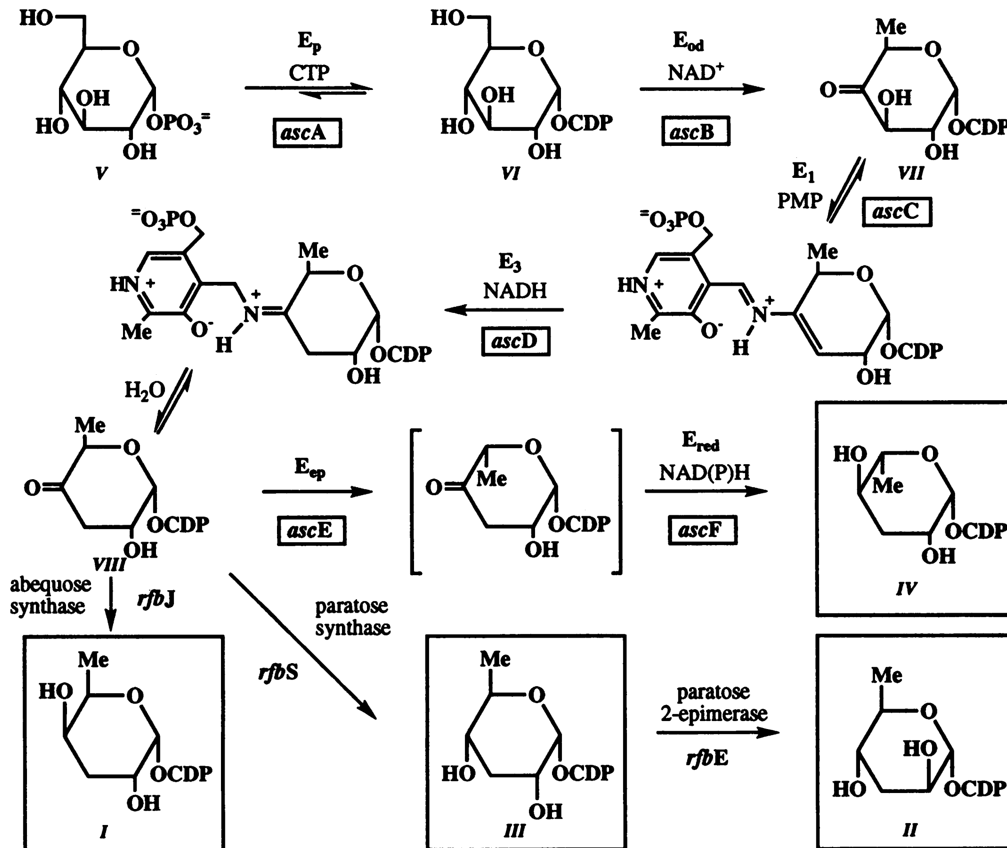

FIG. 1. Biosynthesis of 3,6-dideoxyhexoses from CDP-D-glucose. The corresponding genes associated with the various steps are indicated. Compounds: I, CDP-abequose; II, CDP-tyvelose; III, CDP-paratose; IV, CDP-ascarylose; V, α-D-glucose-1-phosphate; VI, CDP-D-glucose; VII, CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose; VIII, CDP-3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose.

(ORFs) were located and identified as the genes coding for E_{od} (*ascB*), E_1 (*ascC*), and E_3 (*ascD*) by Southern hybridization with oligonucleotide probes designed from the corresponding purified wild-type enzymes (38). These assignments were later substantiated via expression and purification to homogeneity of these gene products (21, 36, 40). Furthermore, the identification via expression and activity assay of the remaining three genes, *ascA* (E_p), *ascE* (E_{ep}), and *ascF* (E_{red}), had also assisted in establishing the order of the latter three genes in the *asc* gene cluster and in completing their indisputable assignments (35, 38). Detailed in this paper is a full account of the cloning, sequencing, and function identification of the entire *asc* gene cluster and its comparison with the *Salmonella* spp. *rfb* clusters. As described below, although the genetic principles underlying the 3,6-dideoxyhexose variation in *Y. pseudotuberculosis* are still in their infancy, the insights gained from these studies provide a strong foundation for constructing an evolutionary model of the polymorphism of this class of unusual sugars.

MATERIALS AND METHODS

General. The bacterial strain, *Y. pseudotuberculosis* serotype VA, was kindly provided by Otto Lüderitz of the Max Planck Institute for Immunobiology, Germany. The *E. coli* strains HB101 and DH5α were purchased from Bethesda Research Laboratories (Gaithersburg, Md.), and XL1-Blue and Y1088 were purchased from Stratagene (La Jolla, Calif.). The phenyl-Sepharose, Sephadex G-100, fast-protein liquid chromatography (FPLC) Superdex-200 and mono Q columns are products of Pharmacia (Piscataway, N.J.). The Sequenase version 2.0 DNA sequencing kit, M13 sequencing primers, 7-deaza-dGTP sequencing mixtures, DNA-modifying enzymes, their respective buffers, and standard plasmids and products used in recombinant DNA work were purchased from suppliers such as United States Biochemical Corp. (Cleveland, Ohio), Promega (Madison, Wis.), Stratagene, and Bethesda Research Laboratories. The λ ZAP II phage vector system, *Eco*RI λ gt11 cloning kit, and Gigapack II Plus λ phage packaging kit were from Stratagene. Duralose, nitrocellulose, and Elutip-D minicolumns were products of Schleicher and Schuell (Keene, N.H.), electrophoretic reagents were from Beckman Instruments (Fullerton, Calif.), and Zeta Probe membrane was purchased from Bio-Rad (Richmond, Calif.). The radiolabeled [α-^{32}P]dATP (800 Ci/mmol), [γ-^{32}P]ATP (>5,000 Ci/mmol), and [α-^{35}S]dATP (>1,000 Ci/mmol) were purchased from Amersham (Arlington Heights, Ill.). Oligonucleotide primers and probes were synthesized by the University of Minnesota Microchemical Facility in the Institute of Human Genetics, the Biomedical Research Institute of Saint Paul Children's Hospital, or National Biosciences (Plymouth, Minn.). Sequence analysis and manipulation were performed with the IntelliGenetics program (Suite 5.4) through the Molecular Biology Computing Center at the University of Minnesota. The wild-type CDP-D-glucose 4,6-dehydratase (E_{od}), CDP-6-deoxy-L-

threo-D-*glycero*-4-hexulose-3-dehydrase (E_1), and CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase (E_3) used in assay procedures and as standards were isolated from the same *Y. pseudotuberculosis* strain by published procedures (21, 42, 48). The protease inhibitors used in enzyme purification were prepared as previously described (21). All protease inhibitors, molecular weight standards, and DEAE-cellulose and most biochemicals were purchased from Sigma (St. Louis, Mo.), while all other chemicals were of analytical reagent grade or the highest quality commercially available.

Genomic DNA isolation and Southern blotting. The general methods and protocols for recombinant DNA manipulations were those described by Sambrook et al. (30) and Ausubel et al. (2). *Y. pseudotuberculosis* genomic DNA was isolated with a slightly modified CsCl DNA preparation described by Ausubel et al. (2). Southern blotting was initiated by digestion of the isolated DNA with various restriction endonucleases, electrophoresis through a 0.8% agarose gel, and transblotting to a Zeta Probe membrane. Membranes containing blotted DNA were preincubated for a period of 1 to 2 h in a solution of 0.5 M Na_2HPO_4 (pH 7.2) containing 1 mM EDTA, 7% sodium dodecyl sulfate (SDS), 5× Denhardt's solution, and 0.5 mg of denatured salmon sperm DNA per ml. This was followed by hybridization overnight at 42°C in the same solution containing 1× Denhardt's solution with an appropriate DNA probe. The degenerate oligonucleotide JST1 (5′-AA[AG]-ACC-GT[TC]-AC[AGTC]-TT[TC]-[GT]C[AGTC]-AA-3′) was prepared on the basis of the N-terminal amino acid sequence of purified wild-type CDP-D-glucose 4,6-dehydratase (E_{od}) from *Y. pseudotuberculosis* (48). The oligonucleotide probe SFL4, based on the N-terminal amino acid sequence of the wild-type E_3 isolated from the same bacterial strain (21), has the sequence 5′-AA[CT]-GT[AGTC]-AA[AG]-CT[ACGT]-CA[CT]-CC-3′. These probes allowed the identification of one DNA fragment which carried *ascA* (E_p), *ascB* (E_{od}), *ascD* (E_3), and part of *ascC* (E_1). To facilitate chromosome walking and the production of a subgenomic library containing the entire coding region (*ascC*) for E_1, a third probe, JST2, with the sequence 5′-CCTATTCCGAGCCTGTTGAA-3′, was designed on the basis of the 3′-end nucleotide sequence of *ascB*. Finally, to complete the *asc* cluster, a fourth probe, JST4 (5′-GTTGAGTGTGATTCTGG-3′), which was designed specifically from the 3′ region beyond *ascC*, was used for screening. For Southern hybridizations, all probes were labeled to a specific activity of 7×10^8 dpm/μg with [γ-^{32}P]ATP and T4 polynucleotide kinase. The membranes hybridized with JST1 were washed twice at 50°C, for 30 min each time, in 1% SDS–2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and twice at 48°C, for 45 min each time, in 0.1% SDS–0.1× SSC. The membranes hybridized with SFL4 were washed three times at room temperature, for 20 min each time, in 1% SDS–2× SSC and once at 42°C, for 45 min, in 1% SDS–0.1× SSC. Membranes hybridized with JST2 were washed three times at room temperature, for 20 min each time, in 1% SDS–2× SSC and once at 42°C, for 45 min, in 0.1% SDS–0.1× SSC. Blots done with JST4 were washed three times at room temperature, for 20 min each time, in 1% SDS–2× SSC and once at 45°C, for 45 min, in 0.1% SDS–0.1× SSC.

Subgenomic library constructions and clone selections. Three subgenomic libraries were constructed. Library YPT1 was generated, as previously described, by cloning of the 3.3-kb *Hin*dIII restriction fragments identified by JST1 hybridization into the λ ZAP II vector (35). From this size-selected library, the desired recombinant plasmid, designated pYPT1, was derived from selected phage by in vivo excision mediated by the addition of helper phage according to the manufacturer's instructions.

The second subgenomic library was constructed on the basis of the screening results with JST2, which revealed consistent hybridization of this radiolabeled probe with an approximately 2.7-kb *Eco*RI restriction band within the genomic digest. A library of the *Eco*RI size-selected DNA fragments, designated YPT3, was then constructed by ligation of the inserts (50 ng) with λ gt11 arms (1 μg) as per the manufacturer's instructions. Packaging and plating with *E. coli* Y1088 was also carried out as suggested by the manufacturer. Approximately 500 plaques were screened under conditions similar to those described for YPT1, and the washes were identical to those used for Southern hybridization. The insert isolated from a positively identified plaque of YPT3 was subsequently inserted into pUC19 to yield pJT17 and pJT18.

The final library (YPT4) designed to complete the cloning of the *asc* cluster was constructed by the isolation of the 3.6-kb *Hin*dIII fragment identified by Southern hybridization with JST4, treatment with calf intestine phosphatase, ligation into *Hin*dIII-digested pUC19, and transformation into *E. coli* DH5α. Approximately 1,000 bacterial colonies were lifted onto nitrocellulose filters and screened with ^{32}P-labeled JST4 under conditions identical to those previously used for Southern hybridization. The plasmid DNAs from positive colonies were isolated from *E. coli* by the alkali method (3), found to be identical, and subsequently designated pJT23 or pJT24, depending on the orientation of the insert.

Restriction analysis and plasmid construction. DNA from the isolated clones was subjected to single, double, or triple digestions with various restriction endonucleases, and the size patterns of the resulting fragments were analyzed by electrophoresis on a 0.8% agarose gel. Comparison of patterns with known reference points within the vector DNA led to the construction of a linear map of the restriction sites within the cloned inserts. Standard recombinant DNA techniques were used for all plasmid manipulations (2, 30).

DNA sequencing. Nucleotide sequencing of the various plasmids was carried out directly on double-stranded templates with commercially available M13 forward and reverse primers by the dideoxy chain termination method (32). On the basis of this approach, the entire sequence of the *asc* cluster (see Fig. 3) was elucidated. Greater than 90% of the sequencing was accomplished with unidirectional (14) or bidirectional deletions; the remainder of the sequencing was completed with synthetic oligonucleotides.

Database analysis. Computer analysis of the resulting sequence information was performed with IntelliGenetics software (Suite 5.4). Initial databank searches with DNA or amino acid query sequences utilized the FASTDB program (8) and the Genetics Computer Group TFASTA and FASTA programs (9). Detailed alignments were derived from the IntelliGenetics ALIGN and DDMATRIX programs and the GENALIGN and PILEUP programs.

Protein expression. The expression system, pJT12, designed for E_p (*ascA*) was constructed by subcloning the *Sca*I-*Kpn*I fragment of pYPT1 into pUC19, in which the direction of transcription of the insert is aligned with the *lacZ* promoter of the plasmid (35). In a similar manner, E_{od} (*ascB*) expression was first attempted, with limited success, by placing the *Pst*I-*Hin*dIII fragment of pYPT1 into pUC19, resulting in pJT7. Further improvement to maximize expression of E_{od} was achieved by fill-in and blunt end ligation of the *Pst*I-*Acc*I fragment of pJT7 into pUC19 to yield pJT8. However, the best expression was obtained with pJT15, which was constructed by subcloning the *Hin*dIII-*Bgl*II fragment of pJT7 into pUC19.

The E_1 (*ascC*) expression system (pJT18) was directly produced by subcloning the *Eco*RI fragment from the YPT3 clone into pUC19 (36). Similarly, the E_3 (*ascD*) expression construct (pSFL28) was prepared by reversing the direction of transcription of the pYPT1 insert (21). Finally, pJT24 (E_{ep}/E_{red}) was directly obtained upon isolation of positive colonies from the JST4 screening. The separate gene cassettes for E_{ep} and E_{red} (pJT27 and pJT29, respectively) were constructed by PCR amplification utilizing primers appropriately designed to provide convenient restriction sites directly adjacent to the desired gene. A schematic representation of essential expression plasmids is given below (see Fig. 2), and the corresponding nucleotide and deduced amino acid sequences of these constructs are designated (see Fig. 3). For each construct, the plasmids were transformed into *E. coli* HB101 and expression was determined by SDS-polyacrylamide gel electrophoresis (PAGE). The desired enzymes were isolated and confirmed by activity assays.

Enzyme assays. α-D-Glucose-1-phosphate cytidylyltransferase (EC 2.7.7.33) activity was determined spectrophotometrically (35) by a slightly modified method of Rubenstein and Strominger (29). Analogously, the E_{od} activity was determined by measuring the formation of CDP-4-keto-6-deoxy-D-glucose, which exhibits a characteristic absorption at 320 nm under alkaline conditions (48).

The assay routinely used to assess E_1 activity relied on the estimation of the extent of tritium release from 4′-[^{3}H]PMP during catalysis (42). Since the substrate of E_1 is not readily available, it has to be prepared in situ from CDP-D-glucose prior to each assay. The newly developed large-scale preparation involved the incubation of purified E_{od} (900 μg) with CDP-D-glucose (21 mg, 37.2 μmol) and NAD^+ (5 mg, 7.4 μmol) in 3 ml of 50 mM potassium phosphate buffer, pH 7.5, at 37°C for 1 h. The product was purified in a fast-protein liquid chromatograph equipped with a mono Q 10/10 column using the solvent systems A (20 mM Tris-HCl, pH 7.4) and B (A plus 0.5 M NaCl, pH 7.4), with the profile 0% B from 0 to 4 min and a linear gradient of 0% to 30% B from 4 min to 24 min, increasing to 100% B at 25 min, with a 5-min final wash. The injection volume was 3 ml, the flow rate was 4 ml min^{-1}, and the detector was set at 280 nm. To determine whether complete conversion to the E_{od} product had occurred, an aliquot was subjected to C_{18} high-performance liquid chromatography using the solvent systems C (200 mM triethylamine phosphate, pH 5.5) and D (C plus 50% CH_3CN) at a flow rate of 1 ml min^{-1}, with the profile 0% D from 0 to 4 min, a linear gradient of 0% to 30% D from 4 min to 24 min, and a linear gradient of 30% D to 100% D from 24 to 30 min, followed by a final wash for 5 min at 100% D. The E_{od} product formed was quantitated by measuring its characteristic absorption at 320 nm (ε, 5,600 $M^{-1} \cdot cm^{-1}$) under alkaline conditions. An aliquot (40 μl) of this mixture was then added to an assay solution containing the labeled PMP coenzyme (0.04 μg, 0.15 nmol, 9.98 nCi) and the appropriate amount of E_1 in 50 mM potassium phosphate buffer (pH 7.5; final volume, 200 μl). The reaction was allowed to incubate at room temperature for 1 h, followed by the addition of activated charcoal (10% solution, 200 μl) and vigorous mixing for 1 min. The mixture was centrifuged at 14,000 × *g* for 2 min, and the supernatant (50 μl) was removed and analyzed by scintillation counting. The readings were calibrated against controls prepared in parallel without substrate or enzyme.

A gas chromatography-mass spectrometry (GC-MS) assay was also developed for both E_1 and E_3 activity determination. Incubation was initiated by mixing appropriate amounts of E_1 and E_3 with 50 μl of the purified E_{od} product (as described above), PMP (10 nmol), and NADH (100 nmol) in a total volume of 200 μl of 50 mM potassium phosphate buffer, pH 7.5. As previously described (42), the reaction was quenched with $NaBH_4$, and the mixture was boiled for 1 min and centrifuged to remove protein precipitate. The supernatant was acidified to pH 2.0 (HCl), boiled for 10 min, neutralized to pH 7.0 (KOH), and lyophilized. The solid residue was redissolved in NH_4OH (1 M, 0.6 ml), reduced with $NaBH_4$ (in dimethyl sulfoxide), and slowly quenched with glacial acetic acid. 1-Methylimidazole (2.51 mmol) in 6 ml of acetic acid was then added, and resulting mixture was stirred at room temperature for 1 h, followed by quenching with 5.0 ml of methanol at 0°C (5). The products were extracted with chloroform, dried, filtered, and concentrated. The samples were then subjected to GC-MS (electron impact and chemical ionization) analysis.

The activities of E_{ep} and E_{red} were analyzed by the same GC-MS assay developed to determine E_1 and E_3 activities (21, 42). A typical procedure involved the incubation of CDP-D-glucose (0.76 mg, 13.4 mM) and NAD^+ (9.95 μg, 150 μM) with E_{od} (30 μg, 3.53 μM, 4.5 U) in 50 mM potassium phosphate buffer, pH 7.5 (total volume, 100 μl), for 30 min at 37°C, followed by the addition of a solution of 50 mM potassium phosphate (200 μl, pH 7.5) containing PMP (0.92 mg; final concentration, 10.8 mM), E_1 (160 μg, 2.6 U), NADH (3.8 mg, final concentration, 17.9 mM), and E_3 (3 μg, 5.4 U). Incubation was continued at 37°C for 1 h. To this mixture was added a mixture of 3 mg of NADH (8.46 mM), 3 mg of NADPH (7.2 mM), and appropriate amounts of the desired crude epimerase and/or reductase preparations in 500 μl of the same phosphate buffer. The subsequent manipulations were identical to those described above for the E_1/E_3 activity, except for the use of $NaBD_4$ instead of $NaBH_4$ as the reducing agent (38).

Protein purification. While the *ascE*- and *ascF*-encoded proteins have not been purified to homogeneity, those encoded by *ascA* (E_p) and *ascD* (E_3) have been reported earlier (21, 35). Purification of the wild-type *ascB* (E_{od})- and *ascC* (E_1)-encoded proteins was achieved by previously described procedures (42, 48), and that of recombinant E_{od} and E_1 was effected by protocols summarized below. All operations were carried out at 4°C. Unless otherwise specified, all buffers contained 1 mM EDTA to prevent enzyme inhibition by exogenous trace metals during purification.

Purification of recombinant CDP-D-glucose 4,6-dehydratase (*ascB*, E_{od}). Growth of *E. coli*(pJT15) cells (*ascB*, E_{od}). An overnight culture of the HB101(pJT15) bacteria was grown in Luria-Bertani (LB) medium supplemented with ampicillin (100 μg/ml) at 37°C and diluted 10-fold into the same medium without antibiotics, which was then incubated for 24 h in an Incubator-Shaker (Lab-Line) with vigorous agitation (225 rpm) at 37°C. The organisms were harvested by centrifugation at 4,000 × *g* for 10 min at 4°C, and the collected cells were washed twice with cold 50 mM potassium phosphate buffer (pH 7.5), followed by immediate disruption. The typical yield was 9.1 g (wet weight) of cells per liter of culture.

Step 1. Crude extracts. Cells from 1 liter of culture (9.1 g of packed cells) were resuspended in four times their volume (36.4 ml) of 50 mM potassium phosphate buffer (pH 7.5). The cells were disrupted by sonication with three 50-s bursts, with 2-min intervals between bursts, with the VirSonic model 300 sonicator at 70% output. The temperature of the extracts was controlled so as not to exceed 4°C during this process. Cellular debris was removed by centrifugation at 10,000 × *g* for 30 min. The supernatant was diluted with the same buffer to 126 ml and was designated the crude extracts.

Step 2. Streptomycin sulfate treatment. In a dropwise fashion, streptomycin sulfate (5% aqueous solution) was added

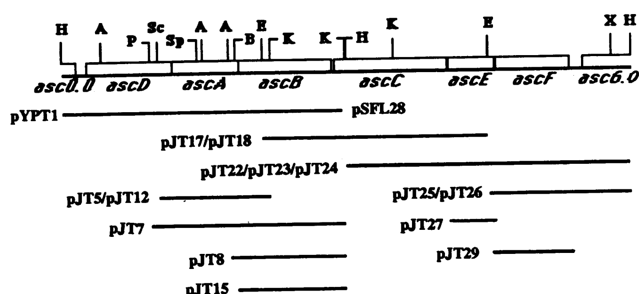

FIG. 2. *asc* region of *Y. pseudotuberculosis* VA. ORFs which have been identified are given their *asc* designations. Selected restriction sites (A, *Acc*I; B, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; Sc, *Sca*I; Sp, *Sph*I; X, *Xba*I) are shown above the map. Various plasmids discussed in the text are also included.

to the crude extract to give a final concentration of 0.8%. After standing for 1 h with stirring, the precipitate was eliminated by centrifugation at 13,000 × *g* for 1 h. The supernatant was diluted with 14.2 ml of 1 M potassium phosphate buffer (pH 7.5) and carried on to the next step.

Step 3. Ammonium sulfate precipitation. Solid ammonium sulfate was added slowly to the enzyme solution from step 2 to give a final concentration of 65% saturation. After the addition was complete, the cloudy solution was stirred overnight. The precipitated proteins were collected by centrifugation at 13,000 × *g* for 1 h and redissolved in a minimum amount of 100 mM potassium phosphate buffer (16.3 ml, pH 7.5). This solution was dialyzed against 1 liter of the same buffer for 12 h, with three changes of buffer.

Step 4. DEAE-cellulose chromatography. The dialysate from step 3 (29 ml) was applied to a column of DE-52 (2.5 by 38 cm) preequilibrated with 100 mM potassium phosphate buffer, pH 7.5. The column was washed with 80 ml of the same buffer, followed by elution with a linear gradient between 500 ml of 100 mM potassium phosphate buffer and 500 ml of 300 mM potassium phosphate buffer, pH 7.5. The flow rate was 33 ml h^{-1}, with fractions of 5.5 ml collected during the entire gradient elution. The contents of fractions 55 to 102 were collected, concentrated to 20 ml, and carried on to the next step.

Step 5. FPLC Superdex-200 gel filtration chromatography. Aliquots (2.2 ml) of the material from step 4 were applied to a FPLC Superdex-200 HR (10/30) column which had been previously equilibrated with 50 mM potassium phosphate buffer, pH 7.5. The sample was eluted with the same buffer at a flow rate of 1 ml · min^{-1}. The desired protein was found to have a retention time of 14 min under these conditions. The collected protein solutions were combined, concentrated to 1.5 ml, and subjected to the next purification step.

Step 6. FPLC mono Q chromatography. The enzyme solution from step 5 was further purified by FPLC with an instrument equipped with a mono Q HR (10/10) column using the solvent systems A (20 mM Tris · HCl, 1 mM EDTA, pH 7.5) and B (A plus 0.5 M NaCl, pH 7.5) with the profile 0% B from 0 min to 4 min and a linear gradient of 0% to 100% B from 4 min to 24 min, with a 6-min final wash (100% B). The flow rate was 3 ml · min^{-1}. The pooled active enzyme solution was concentrated (to 8.1 ml), aliquoted, and stored at −85°C.

Purification of recombinant CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose-3-dehydrase (*ascC*, E_1). Growth of *E. coli*(pJT18) (*ascC*, E_1). An overnight culture of strain HB101(pJT18) was typically grown in LB medium supplemented with ampicillin (100 μg/ml) and $(NH_4)_2Fe(SO_4)_2$ (50 mg $liter^{-1}$) at 37°C and diluted 10-fold in the same medium, without antibiotic, which was then incubated for 24 h in an Incubator-Shaker with vigorous agitation (225 rpm) at 37°C. The organisms were harvested by centrifugation at 4,000 × *g* for 10 min at 4°C, and the collected cells were washed twice with cold 50 mM potassium phosphate buffer (pH 7.5), followed by immediate disruption. The typical yield was 7.5 g (wet weight) of cells per liter of culture.

Step 1. Crude extracts. Cells from 1 liter of culture (7.5 g of packed cells) were resuspended in four times their volume (30 ml) of 50 mM potassium phosphate buffer (pH 7.5). The cells were disrupted by sonication with three 40-s bursts, with 2-min intervals between bursts. The temperature of the extracts was controlled so as not to exceed 4°C during this process. Cellular debris was removed by centrifugation at 10,000 × *g* for 20 min. The supernatant was diluted with the same buffer to 100 ml and was designated the crude extracts.

Step 2. Streptomycin sulfate treatment. In a dropwise fashion, streptomycin sulfate (19 ml, 5% aqueous solution) was added to the crude extract to give a final concentration of 0.8%. After standing for 1 h with stirring, the precipitate was eliminated by centrifugation at 13,000 × *g* for 30 min. The supernatant was diluted with 12 ml of 1 M potassium phosphate buffer (pH 7.5) and carried on to the next step.

Step 3. Ammonium sulfate precipitation. Solid ammonium sulfate was added slowly to the enzyme solution from step 2 to give a final concentration of 65% saturation. After the addition was complete, the cloudy solution was stirred for 3 to 4 h. The precipitated proteins were collected by centrifugation at 13,000 × *g* for 30 min and redissolved in a minimum amount of 200 mM potassium phosphate buffer (16 ml), pH 7.5. This solution was dialyzed against 1 liter of the same buffer for 12 h, with four changes of buffer.

Step 4. DEAE-cellulose chromatography. The dialysate from step 3 (16 ml) was applied to a column of DE-52 (2.5 by 50 cm) preequilibrated with 200 mM potassium phosphate buffer, pH 7.5. The column was washed with 50 ml of the same buffer, followed by elution with a linear gradient between 250 ml of 200 mM potassium phosphate buffer, and 250 ml of 400 mM potassium phosphate buffer, pH 7.5. The flow rate was 40 ml · h^{-1}, with fractions of 5 ml being collected during the entire gradient elution. Three separate fractions were collected and carried on to the next step. Fraction A consisted of tubes 30 to 45 (116 mg, concentrated to 4 ml), fraction B consisted of tubes 46 to 63 (68.2 mg, concentrated to 2.3 ml), and fraction C consisted of tubes 64 to 94 (35.3 mg, concentrated to 1.2 ml). All three fractions contained the desired protein but had different iron-sulfur and PMP contents.

Step 5. Sephadex G-100 gel filtration. To prevent overloading, only 1 ml of the solution from fraction A, B, or C from step 4 was loaded onto a Sephadex G-100 column (1.5 by 170 cm) and eluted with 50 mM potassium phosphate buffer (pH 7.5) per run. The flow rate was 15 ml · h^{-1}, and 2.5-ml fractions were collected throughout. The active protein from fractions 35 to 47 was concentrated (approximately 30 mg ml^{-1}), aliquoted (500 μl per aliquot), and stored at −85°C. This procedure was repeated with the remainder of the solutions from step 4 as necessary. Often the solutions of step 4 were preincubated with excess PMP prior to gel filtration.

SDS-PAGE. The level of expression of the gene products and their relative abundance in the crude protein extracts were assessed by SDS-PAGE. Electrophoresis was carried out in the discontinuous buffer system of Laemmli (19), and the separating gel and stacking gel were 13% and 6% polyacrylamide, respectively. Prior to electrophoresis, 1 ml of overnight culture was centrifuged for 2 min at 14,000 × *g*. The supernatant was

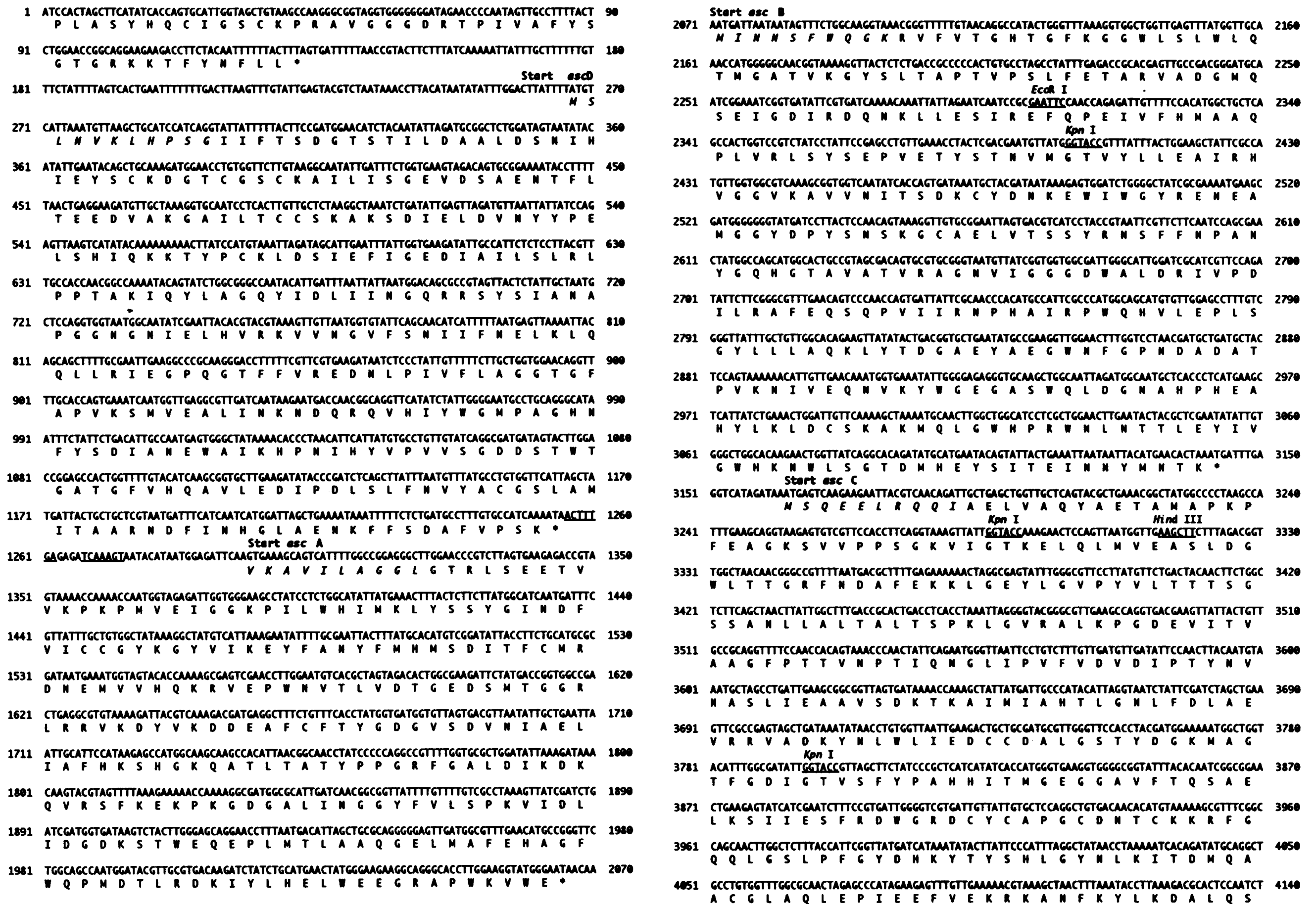

FIG. 3. DNA and deduced amino acid sequences of the *asc* region of *Y. pseudotuberculosis* VA. Regions confirmed by amino acid sequencing of purified proteins are italicized, potential DNA terminator stem-loop structures are underlined, and a few restriction sites of importance are also underlined and labeled above.

then removed and added to a solution (100 μl) containing 3% SDS, 5% 2-mercaptoethanol, 10% glycerol, and 1% dye, followed by vigorous vortexing and incubation at 100°C for >10 min. Gels were stained with Coomassie blue (41) and destained with acetic acid-ethanol-H_2O (15:20:165, by volume).

RESULTS

Cloning and localization of the *asc* gene cluster. Oligonucleotide screening of 750 plaques of YPT1, constructed by cloning of the 3- to 4-kb *Hin*dIII genomic digests into a λ ZAP II vector, resulted in the isolation of four positive, identical clones (0.53%) whose 3.3-kb insert (pYPT1) was later found to harbor the complete *ascA* (E_p), *ascB* (E_{od}), and *ascD* (E_3) genes (21, 35). Since only a truncated portion of the *ascC* (E_1) gene was present in pYPT1, a λ gt11 subgenomic library (YPT3) with a plating efficiency of 2×10^6 and a blue-to-white ratio of 1:35 was constructed from a 2- to 4-kb region of *Eco*RI-digested genomic DNA which was selected by Southern hybridization with JST3. Oligonucleotide screening of 500 plaques from this library led to the isolation of three positive clones (0.6%). Recombinant phage DNA isolation and restriction analysis found these three clones to be identical. Thus, subsequent subcloning, from a single clone, into pUC19 yielded pJT17 and pJT18, which differed only in insert orientation as determined by restriction mapping. To facilitate further chromosomal walking, preliminary pJT18 sequence information was utilized to design JST4. Southern hybridization with JST4 revealed a consistent hybridization pattern with *Hin*dIII-digested DNA fragments of approximately 3.6 kb. Therefore, library YPT1 (1,000 plaques) was rescreened with JST4, and five positive clones were isolated (0.5%). Through in vivo excision, these five clones were converted to the plasmid form of the pBluescript vector and were found to be identical by restriction analysis and Southern blotting. Thus, subsequent manipulation was performed on a single clone (pJT22). Meanwhile, a pUC19 library containing the *Hin*dIII 3- to 4-kb range was constructed (YPT4). Screening of this library with JST4 led to the identification of pJT23 and pJT24, differing only in the orientation of their inserts. However, as expected, the inserts of pJT23 and pJT24 were identical to that of pJT22. The complete physical map presented in Fig. 2 was established by detailed restriction enzyme analysis of plasmids pYPT1, pJT18, and pJT22. All constructs generated in this work are depicted in Fig. 2.

Nucleotide and amino acid sequences of the *asc* cluster. The nucleotide and deduced amino acid sequences of the *asc* cluster are presented in Fig. 3. The *asc* subcluster, a portion of the larger O antigen cluster of *Y. pseudotuberculosis*, begins at position 267 of the 6.6-kb fragment of cloned DNA and continues through position 5932. Analysis of the entire se-

```
4141 TGCGCTGACTTCATTGAGTTACCAGAAGCGACTGAAAATTCAGATCCATCATGGTTTGGTTTCCCTATCACTCTGAAAGAAGATAGCGGA 4230
     C  A  D  F  I  E  L  P  E  A  T  E  N  S  D  P  S  W  F  G  F  P  I  T  L  K  E  D  S  G

4231 GTTAGCCGCATTGATCTGGTTAAATTCCTTGATGAAGCTAAAGTGGGAACTCGCCTACTATTTGCCGGTAATTTAACTCGCCAGCCGTAT 4320
     V  S  R  I  D  L  V  K  F  L  D  E  A  K  V  G  T  R  L  L  F  A  G  N  L  T  R  Q  P  Y

4321 TTCCATGATGTGAAATACCGTGTTGTTGGTGAATTGACAAACACCGATAGAATTATGAATCAAACTTTCTGGATTGGTATTTACCCAGGC 4410
     F  H  D  V  K  Y  R  V  V  G  E  L  T  N  T  D  R  I  M  N  Q  T  F  W  I  G  I  Y  P  G
                                                                                 Start  ascE
4411 CTGACACATGATCATTTGGATTATGTCGTGAGTAAGTTTGAAGAGTTCTTTGGTTTGAATTTTTAATTGAGGAATTGGGGGTAATAGTCC 4500
     L  T  H  D  H  L  D  Y  V  V  S  K  F  E  E  F  F  G  L  N  F  *           L  G  V  I  V  P

4501 CCCACTATCTTATGATTTTTAAAAAATTGGATATTGAGGGTTGTTACCTTATAGAATTTAATAAATTCATTGATTCCCGAGGAACGTTTG 4590
       H  Y  L  M  I  F  K  K  L  D  I  E  G  C  Y  L  I  E  F  N  K  F  I  D  S  R  G  T  F  V

4589 TAAAAACATTTCACTCGGACTTTTTTTCTGAGAATGGAATAGTTTTGGATATGCGAGAGGAGTTTTATTCAATATCAGCTAAGAATGTTA 4680
       K  T  F  H  S  D  F  F  S  E  N  G  I  V  L  D  M  R  E  E  F  Y  S  I  S  A  K  N  V  I

4681 TACGGGGTATGCATTTTCAAATGCCTCCAGCTGAACATGACAAACTCGTCTATTGTGTAAATGGAGCTGTTCTTGATGTTATCTTAGATA 4770
       R  G  M  H  F  Q  M  P  P  A  E  H  D  K  L  V  Y  C  V  N  G  A  V  L  D  V  I  L  D  I

4771 TAAGAAAAGATTCTAAAACATATGGGGAGTATTTTAGTATTGAATTAAGTTATGAAAATAGCTTAGCACTATGGGTGCCAAAAGGTTTAG 4860
       R  K  D  S  K  T  Y  G  E  Y  F  S  I  E  L  S  Y  E  N  S  L  A  L  W  V  P  K  G  L  A

4861 CCCATGGATTTTTATCACTTGCAGATAATTCAATAATGTTTTATAAAACAAGTTCTGTTCACAATGTTGAGTGTGATTCTGGTATTAAAT 4950
       H  G  F  L  S  L  A  D  N  S  I  M  F  Y  K  T  S  S  V  H  N  V  E  C  D  S  G  I  K  W
      EcoR I
4951 GGAATTCATTTGGATTTAAATGGCCGATTGATAATCCAATTATATCTGAGAAAGATAATTCTCTTTGCTATTTTGATGAGTTTGATAGTT 5040
       N  S  F  G  F  K  W  P  I  D  N  P  I  I  S  E  K  D  N  S  L  C  Y  F  D  E  F  D  S  S
                  Start  ascF
5041 CATTCTAAAGAGAGTATTGATGAAATTATTAATTACAGGCGTCTCCGGTTATATAGGTTCTCATTTGGTTAATTACTTGGCAAACTTAGG 5130
       F  *           M  K  L  L  I  T  G  V  S  G  Y  I  G  S  H  L  V  N  Y  L  A  N  L  G

5131 TGGGTATGAAATTTATGGTATATCAAGAAATGAGATATTAGACCAGGATATTAACCAACTTTTATTAAACATAAAAATTTTCCAACTGGA 5220
      G  Y  E  I  Y  G  I  S  R  N  E  I  L  D  Q  D  I  N  Q  L  L  L  N  I  K  I  F  Q  L  D

5221 TAGAGATTCGCTTCCTGATATATTAAAGCGTGTTCGTCCCGATGTTGTCATTCATTTAGCATCATGTTTTCTTTCTCAACACTCGTATAA 5310
      R  D  S  L  P  D  I  L  K  R  V  R  P  D  V  V  I  H  L  A  S  C  F  L  S  Q  H  S  Y  K

5311 AAACATCAAAGAAATAATAAAAAGTAATGTTGAGTTTCCAACTGAGTTACTTGAGGCAATGAATGATGTTGGCGTAAAGAAAATAATAAA 5400
      N  I  K  E  I  I  K  S  N  V  E  F  P  T  E  L  L  E  A  M  N  D  V  G  V  K  K  I  I  N

5401 TACAGGGACATCATGGCAATGTTTTAATTCTGACACTTATAATCCTGTAAATTTATATGCTGCAAGCAAGCAAGCATTTGAGGATATCCT 5490
      T  G  T  S  W  Q  C  F  N  S  D  T  Y  N  P  V  N  L  Y  A  A  S  K  Q  A  F  E  D  I  L

5491 TAAATTTTATATAAACGCAGAGGGTTTTTCTGCTATAAATTTGAAACTTTTTGATACTTACGGTGGGGTGGATAAAAGAAGAAAGTTGAT 5580
      K  F  Y  I  N  A  E  G  F  S  A  I  N  L  K  L  F  D  T  Y  G  G  V  D  K  R  R  K  L  I

5581 ATCATTGTTAGATGATATTGCTAAAAATAATAAACAGTTAGATATGTCACCTGGAGAGCAGTTACTAGACTTGGTTCATATAAATGATGT 5670
      S  L  L  D  D  I  A  K  N  N  K  Q  L  D  M  S  P  G  E  Q  L  L  D  L  V  H  I  N  D  V

5671 ATGTAGAGCATTTAAAATAGCTATAGATAAATTATGTGAGCTTCCATCTGAGTATGTTGTTTCTTATGGCGTATCAAATAAATATAGAGT 5760
      C  R  A  F  K  I  A  I  D  K  L  C  E  L  P  S  E  Y  V  V  S  Y  G  V  S  N  K  Y  R  V

5761 TACGTTGAAGGAGTTGGTAAGTATTTATGAACGAGTTAATAATGTGAAATTAAATATTAACTTTGGAACGCGTGAATATCGAAATAGAGA 5850
      T  L  K  E  L  V  S  I  Y  E  R  V  N  N  V  K  L  N  I  N  F  G  T  R  E  Y  R  N  R  E

5851 AGTTATGGTGCCATGTACTAACATACAAAACTTGCCAGATTGGGAAGTTGTTATACCATTATCGCAAGGTTTAAAATATTAAAAATTCTT 5940
      V  M  V  P  C  T  N  I  Q  N  L  P  D  W  E  V  V  I  P  L  S  Q  G  L  K  Y  *

5941 ATCCTCTTATTATTTAATGGTGATTTCTCAAATTATTAGTACTCTCTTGGGAATGTAGTTAATACAGCACATATATATCACGGTTGAACC 6030

                                          Start orf6.0
6031 GTTAAATTTGGTTGTAATATAACTGTAGGGTAAAATGAAAATAAGTTCTAAAATGACATGGGGGTTATTAAATCAAGGGTGTAATATATT 6120
                                          M  K  I  S  S  K  M  T  W  G  L  L  N  Q  G  C  N  I  F

6121 TATTCCATTTGTAATAATGTTTTTGTCTTACAAATATCTCCCCTTAGAAATGGCTAGTGTATGGGTTATATTCTTATCGATGATATCACT 6210
      I  P  F  V  I  M  F  L  S  Y  K  Y  L  P  L  E  M  A  S  V  W  V  I  F  L  S  M  I  S  L

6211 GATTACTTTGTTTGATTTTGGTTTGTCACCAGCAATAGTCAGAAATGTTAGCTATGTGATATCTGGAGCGCAAAATCTAGTTAAAAGAGG 6300
      I  T  L  F  D  F  G  L  S  P  A  I  V  R  N  V  S  Y  V  I  S  G  A  Q  N  L  V  K  R  G

6301 AATAGATGATATTATTATAAAAGATGTAATATCCTATCCTCTTCTATCTAGATTGCTATTTGATATAAAACGAATTTATTTATATCTATC 6390
      I  D  D  I  I  I  K  D  V  I  S  Y  P  L  L  S  R  L  L  F  D  I  K  R  I  Y  L  Y  L  S

6391 TCTAATAGCATTTTTTATTATAGTTATTGGCGGCGTTTGGTATTTTTATTATATATCACCTCTGGATATAAAAAACGAAGTTGCATATTC 6480
      L  I  A  F  F  I  I  V  I  G  G  V  W  Y  F  Y  Y  I  S  P  L  D  I  K  N  E  V  A  Y  S

6481 TTGGCTTTTATTTTCAAGTGCGCTAATAATAAATCTTATTTATTTATATTACGTTCCTGTATTAACGGGCTTGGGTGAAATCGAAAGTTC 6570
      W  L  L  F  S  S  A  L  I  I  N  L  I  Y  L  Y  Y  V  P  V  L  T  G  L  G  E  I  E  S  S

6571 TTATAAAGCAAATGTTTTTGGGAGAATCATTTGGTTTTTTTTAAG
      Y  K  A  N  V  F  G  R  I  I  W  F  F  L
```

FIG. 3—*Continued.*

quence reveals the presence of the six complete ORFs schematically illustrated in Fig. 2. The identity of each gene is discussed in the following sections. The unidentified ORFs are named according to the start position (in kilobases) on the map. The presumed start site and, where present, the Shine-Dalgarno sequence for each ORF are given in Fig. 4, together with other properties. There exist two truncated ORFs within this cloned fragment, one prior to *ascD* and one beyond the postulated *ascF*. Among the start codons of the six complete ORFs, three lie within 10 bases or fewer after the stop codon of the previous ORF. One starts with a GTG (*ascA*), and another probably starts with a TTG (*ascE*). In fact, the start codons for four of the identified genes (*ascA*, *ascB*, *ascC*, and *ascD*) have been confirmed by amino acid sequencing of the purified proteins. However, a potential ATG start exists for both *ascA* genes. All eight are presumably initiated by translational coupling, as observed in the *rfb* cluster (15), although most also have quite good Shine-Dalgarno sequences. The usage of terminator codons is in agreement with the usual *E. coli* preferences, with TAA being used five times as a single

```
                               GTG
                               TTG
                            AAAATGAAA    TTAA
Genes     -----------                    ++++      Space between
 or       21111111111--------- +++++++++1111           SD and
ORFs      0987654321098765432101234567890123        first codon


ascD       atatatttGGActtattttATGtcattaaaTgt             8
rfb7.6     tgttaacaGAGtaagcAtcGTGtctcataTTAt             8

ascA       atacataatGGAGattcAAGTGAAAgcagTcAt             6
rfbF       ataatattGAGGataatttATGAAAgcggTcAt             7

ascB       aaggtatgGGAataacAAAATGattaataatAg             8
rfbG       aaaacctgGGAGtaactAgATGattgataaaAA             7

ascC       gatttGAGGtcatagatAAATGagtcaagaagA            10
rfb10.4    actttaaGGAatcaaagtAATGacagcaaaTAA             9

ascE       tgaatttttaattGAGGAATTGggggtaaTagt             2

ascF       cattctaaagaGAGtattgATGAAAttattaAt             5
rfbE       taattgaaGAGGaagggAAATGaagcttttaAt             7

orf6.0     aatataactgtAGGgtAAAATGAAAataagTtc             5
rfb12.8    TAAGtcGtTtAattgcAtAGTGAAAgttcaatt             8
```

FIG. 4. Comparison of the *asc* and *rfb* initiation regions. Bases as part of the potential Shine-Dalgarno sites (any bases that complement the 13 nucleotides, TAAGGAGGTGATC, at the 3′ end of the 16S RNA) or the presumed start codon (which is also underlined) are capitalized. Although a GTG start is assigned to *ascA*, a potential ATG start exists for the *ascA* gene. Capital letters are also used for the second codon, where it is AAA (the preferred second codon), and components of the sequences TTAA and AAA from +10 to +13 and from −1 to −3, respectively. The termination codon of the preceding gene is indicated by italics if it is in the region shown. The reference sequences involved are also shown above the set of sequences.

stop codon. The two additional termination sites include a single TGA and a coupled TAG-TGA cluster.

Gene expression and protein purification. The expression of the first four genes present in the *asc* cluster was accomplished either by redirecting the target gene to align with the direction of transcription of the promoter or by bringing it closer to the promoter of the vector (Fig. 2). The resulting constructs, pJT12, pJT15, pJT18, and pSFL28, designed for the expression of E_p, E_{od}, E_1, and E_3, respectively, were found to be overexpressed after transformation into the appropriate *E. coli* strains. Purification of the recombinant E_p and E_3 has been separately reported (21, 35), and the comparison of expression is summarized in Fig. 5.

Identification of ascarylose biosynthetic genes. As described above, the genes *ascA*, *ascB*, *ascC*, and *ascD* had been subcloned, overexpressed, and purified (Fig. 5); through careful characterization of the resulting enzymes by specific activity assays, we were able to assign them as the genes encoding E_p, E_{od}, E_1, and E_3, respectively. These assignments were further substantiated by comparing the N-terminal amino acid sequences of the purified proteins to those derived from the corresponding nucleotide sequences. In addition, translation of the fifth ORF (*ascE*, 3′ from *ascC*) resulted in a protein (21,663-Da subunit size from DNA translation; subunit molecular size [from Fig. 5], 22 kDa) whose sequence data reveal the absence of any common cofactor binding motif. Since the postulated C-5 epimerization of the 4-keto-3,6-dideoxy-D-glucose derivative (compound VIII) is expected to proceed via an enolization-reprotonation mechanism and thus to require no cofactor, this ORF (*ascE*) most likely encodes E_{ep} (38). Examination of the final complete ORF (3′ from the assigned

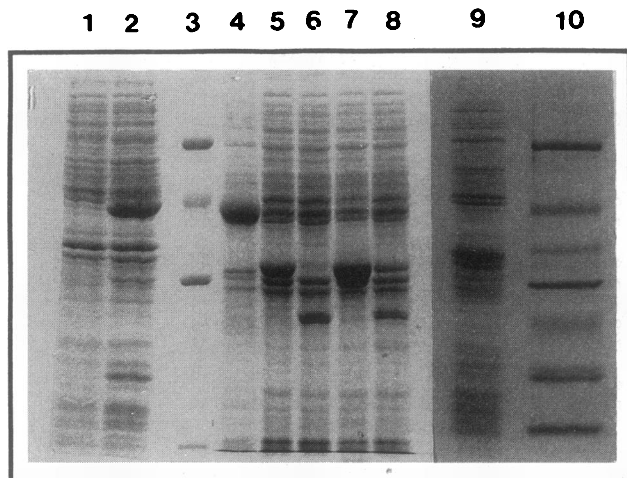

FIG. 5. Expression in *E. coli* of the cloned *asc* gene products as observed by SDS-PAGE. Lanes: 1, pUC18 crude extracts (control); 2, pJT24 crude extracts (E_{ep}, E_{red}, and truncated E_1); 3, molecular size markers (bovine serum albumin, 66 kDa; egg albumin, 45 kDa; carbonic anhydrase, 29 kDa; and α-lactalbumin, 14.2 kDa); 4, pJT18 crude extracts (E_1); 5, pJT8 crude extracts (E_{od}); 6, pJT12 crude extracts (E_p); 7, pJT15 crude extracts (E_{od}); 8, pJT7 crude extracts (E_{od} and E_p); 9, pSFL28 crude extracts (E_3); and 10, molecular size markers (bovine serum albumin, 66 kDa; egg albumin, 45 kDa; glyceraldehyde-3-phosphate dehydrogenase, 36 kDa; carbonic anhydrase, 29 kDa; trypsinogen, 24 kDa; β-lactoglobulin, 18.4 kDa; and α-lactalbumin, 14.2 kDa).

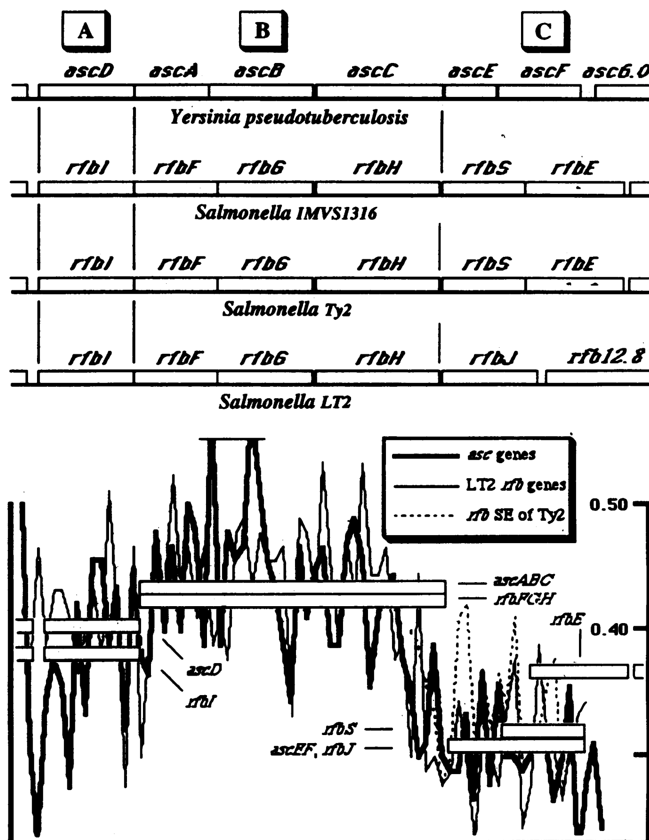

FIG. 6. 3,6-Dideoxyhexose gene clusters in *Y. pseudotuberculosis* and *Salmonella* strains. The G+C content is given in a plot directly beneath the schematic comparison. Each point represents an average G+C content of a span of over 90 bp along the cluster. As discussed in the text, the three bars crossing over the plot cover regions with distinct G+C contents. Since the G+C content of the *rfbFGH* series from *Salmonella* strain Ty2 is essentially identical to the corresponding region in *Salmonella* strain LT2, they have not been included in the graphical comparison.

ascE gene) unveils a gene that, when translated, contains the necessary nicotinamide binding motif (44) in a protein with a subunit size of 33,220 Da (subunit molecular size [from Fig. 5], 35 kDa). Hence, this gene was assigned as *ascF*, which could code for E_{red} (38). Support for the E_{ep} and E_{red} assignments was provided by the GC-MS assay, in which the homogeneous E_p, E_{od}, E_1, and E_3 were used to generate compound VIII, followed by the incubation with the crude *ascE* and *ascF* products via expression from pUC18-based plasmids (pJT27 and pJT29) in *E. coli* HB101. The fact that reduction of the resulting CDP-ascarylose (compound IV) with $NaBD_4$ produced a monodeuterated ascarylitol tetraacetate, while in the presence of denatured E_{ep} and E_{red} or in the HB101(pUC18) blank only the dideuterated species was formed (38), allowed an indisputable differentiation among the enzymatic products and thereby strongly supported the existence of both E_{ep} and E_{red} in the expressed extracts.

DISCUSSION

It has been suggested that all 3,6-dideoxyhexose biosynthetic pathways (excepting that for colitose) share the common intermediate (compound VIII), diverging beyond (Fig. 1) (10, 12). For paratose (compound III), C-4 reduction is catalyzed by paratose synthase to the corresponding ribose configuration. Subsequently, the CDP-paratose-2-epimerase-catalyzed C-2 epimerization of paratose results in the formation of tyvelose (compound II). In the biosynthesis of abequose, abequose synthase-catalyzed C-4 reduction yields abequose (compound I). Ascarylose (compound IV) biosynthesis, on the other hand, results from an initial C-5 epimerization followed by reduction with the C-4 stereochemistry identical to that given by abequose synthase. One of the current aims in O-antigen research is the development of a comprehensive perspective of O-antigen genetics with a focus on the evolutionary origin of the O-antigen variation and its role in pathogenesis. Initiation towards this goal began with the cloning and analysis of the *Salmonella typhimurium rfb* (O-antigen) cluster, which led to the elucidation of the entire sequence of this region and the identification of 12 of the 15 genes present in the cluster (6, 15, 46). Subsequently, the deoxy and dideoxy genes within the *rfb* clusters of *S. typhimurium* (group B, which contains abequose) were compared to those within *Salmonella paratyphi* (group A, which contains paratose), *Salmonella typhi* (group D, which contains tyvelose), and, most recently, *Y. pseudotuberculosis* M85 (serogroup IIA, which contains abequose) (17, 18). As shown in Fig. 6, the 3,6-dideoxyhexose biosynthetic genes were found to be clustered and relatively conserved throughout all four serogroups (20). It should be noted that many of these ORFs had been identified: *rfbF* for glucose-1-phosphate cytidylyltransferase (E_p), *rfbG* for CDP-D-glucose 4,6-dehydratase (E_{od}), *rfbJ* for abequose synthase, *rfbS* for paratose synthase, and *rfbE* for CDP-paratose-2-epimerase (found in both group A and group D, but functional only in *S. typhi*, group D). However, due to the lack of facile assays, the cloned genes essential for the culminating C-3 deoxygenation in the biosynthesis of 3,6-

TABLE 1. G+C contents for *asc* and *rfb* genes and codon positions

Gene	G+C content	Avg G+C content of codon position[a]:		
		P_1	P_2	P_3
ascE	0.318	0.434	0.301	0.215
ascF	0.312	0.422	0.281	0.242
rfbJ	0.320	0.434	0.302	0.216
rfbS	0.333	0.414	0.352	0.240
*rfb*12.8	0.310	0.382	0.324	0.221
ascD	0.379	0.496	0.371	0.266
rfbI	0.405	0.525	0.380	0.304
rfbE	0.355	0.436	0.368	0.269
ascA	0.441	0.569	0.361	0.387
ascB	0.424	0.453	0.401	0.400
ascC	0.424	0.549	0.398	0.329
rfbF	0.437	0.592	0.373	0.322
rfbG	0.437	0.579	0.382	0.338
rfbH	0.446	0.550	0.404	0.379

[a] The values of the first, second, and third codon positions (P_1, P_2, and P_3, respectively) are corrected average G+C contents of the three codon positions that are calculated from 56 of 64 triplets (33). Because of the inequality of A and G at the third position, the three stop codons (TAA, TAG, and TGA) and the three codons for isoleucine (ATT, ATC, and ATA) were excluded in the calculation of P_3. The two single codons for methionine (ATG) and tryptophan (TGG) were also excluded in the calculation of all three (P_1, P_2, and P_3). Genes with similar G+C contents are grouped.

dideoxyhexoses had evaded designation. Interestingly, data bank searches using the translated sequences of the *asc* gene products revealed extensive residue homology (51% for *ascD-rfb*7.6, 80% for *ascA-rfbF*, 72% for *ascB-rfbG*, and 86% for *ascC-rfb*10.4) with the *S. typhimurium rfb* cluster (38). As a result, we can now confidently assign and subsequently refer to the *S. typhimurium rfb* ORF 10.4 as coding for the *Salmonella* E_1 equivalent (*rfbH*) and *rfb* ORF 7.6 as coding for the *Salmonella* E_3 equivalent (*rfbI*) in the abequose, paratose, and tyvelose biosynthetic pathways (Fig. 1).

Detailed analysis of the *asc* gene cluster also revealed that its G+C content is relatively low, and Fig. 6 offers a direct comparison with the corresponding *rfb* regions. As illustrated in Fig. 6 and Table 1, the G+C content of discrete regions of the *asc* cluster varies, ranging from 0.32 to 0.45. Since this ratio is a genome characteristic for a given bacterial species (0.48 for *Yersinia* spp. and 0.50 to 0.52 for *Salmonella* spp. [47]) and its

TABLE 2. Yields of wild-type *asc* products from *Y. pseudotuberculosis*

Gene	Enzyme	No. of steps	Yield[a]	Sp act
ascA	E_p	3	0.006	17.8[b]
ascB	E_{od}	6	0.003	188[c]
ascC	E_1	4	0.006	4.5[d]
ascD	E_3	4	0.0007	23.9[e]
ascE	E_{ep}	ND[f]		
ascF	E_{red}	ND		

[a] Milligrams of protein per gram (wet weight) of cells.
[b] Micromoles of product per milligram per hour (41a).
[c] Micromoles of product per milligram per hour (48).
[d] Measured by the [^{3}H]PMP tritium release assay in nanomoles per milligram per hour (42).
[e] Micromoles of dichlorophenolindolphenol (DCPIP) consumed per milligram per minute (21).
[f] ND, not determined.

TABLE 3. Yields of cloned *Yersinia asc* products expressed in *E. coli*

Gene	Enzyme	No. of steps	Yield[a]	Sp act
ascA	E_p	2	5.0	152[b]
ascB	E_{od}	3	14.1	176[c]
ascC	E_1	1	8.8	20.3[d]
ascD	E_3	4	0.4	32.4[e]
ascE	E_{ep}	ND[f]	0.8[g]	ND
ascF	E_{red}	ND	0.3[g]	ND

[a] Milligrams of protein per gram (wet weight) of cells.
[b] Micromoles of product per milligram per hour (35).
[c] Micromoles of product per milligram per hour (this study).
[d] Measured by the [^{3}H]PMP tritium release assay in nanomoles per milligram per hour (this study).
[e] Micromoles of DCPIP consumed per milligram per minute (21).
[f] ND, not determined.
[g] Based upon SDS-PAGE (Fig. 5).

divergence in different species has been attributed to the variation in the mutation rates of (A/T) to (G/C) and (G/C) to (A/T) base pairs (34), the discrepancy in G+C content within the *asc* gene cluster strongly suggests the presence of several sets of genes in this cluster, each with a different origin or history. In addition, the G+C content of the third base (P_3) of many synonymous codons under mutational pressure often deviates more from the organism's typical G+C ratio than those found for the other two bases (P_1 and P_2) (28, 33). Examination of the codon usage patterns summarized in Table 1 further supported the hypothesis that the dideoxy sugar biosynthetic genes consist of groups of genes with distinctive characteristics. As shown in Table 1 and Fig. 6, the genes *ascE*, *ascF*, *rfbJ*, *rfbS*, and *rfb*12.8 (function unknown) have G+C contents of approximately 0.32 and display similar codon usage (region C). Upstream, the next group of genes has a substantially higher G+C content, estimated at 0.43. This region (region B), from *ascA* to *ascC* in *Yersinia* serogroup VA (found to be essentially identical to *rfbF* to *rfbH* in *Yersinia* serogroup IIA [18]) and from *rfbF* to *rfbH* in *Salmonella* spp., is relatively homogeneous among all strains. Finally, further upstream (region A) lies the gene encoding the CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase, *ascD* in *Yersinia* serogroup VA (found to be essentially identical to *rfbI* in *Yersinia* serogroup IIA [18]) and *rfbI* in *Salmonella* spp. On the basis of Table 1, the gene encoding CDP-paratose 2-epimerase (*rfbE* from *S. typhi*), located downstream from *rfbFGH*, appears to be a member of this class as well.

On the basis of the G+C contents, the genes of the ascarylose pathway can be divided into three groups, with the first gene (*ascD*) present in the 0.38 segment, the second three genes (*ascABC*) in the 0.43 cluster, and the last two genes (*ascEF*) in the 0.32 set. Assuming that the directional mutation pressure is alike in both *Yersinia* and *Salmonella* spp., the central three genes (*ascABC* or *rfbFGH*) may have entered the corresponding organism first, since these most closely resemble the genomic G+C content of the host organism. Furthermore, in view of the profound similarities of the G+C contents exhibited by these genes, this cluster may have been incorporated into *Yersinia* and *Salmonella* spp. at approximately the same time period. Alternatively, a recombination event between these organisms, resulting in a transfer of this region, could also lead to the high homogeneity found in region B of Fig. 6. Since the *ascA* and *ascB* (or *rfbFG*) gene products catalyze the conversion of α-D-glucose-1-phosphate (compound V) to CDP-6-deoxy-L-*threo*-D-*glycero*-4-hexulose (compound VII), which is the key biosynthetic precursor for most

deoxyhexoses (10, 12), the early entry of these genes into the host organisms is essential for the structural variance of O antigen. Another member of this cluster, the *ascC* gene, encodes a PMP-dependent iron-sulfur enzyme (E_1) critical to C-3 deoxygenation. Interestingly, we have recently discovered an evolutionary link between *ascC* and a number of the genes believed to participate in 2,6- and 4,6-dideoxyhexose pathways as well as to some pyridoxal 5'-phosphate (PLP)- and PMP-dependent transaminases (39). Although all of the consensus sequences lack the putative [2Fe-2S] domain, on the basis of their strong homologies with E_1 the individual roles of these proteins may exhibit an ancestral relationship to the E_1's PMP-dependent catalysis. Thus, the preevolved organism, containing only the region B cluster (*ascABC* or *rfbFGH*), of which the prebiotic *ascC* or *rfbH* gene product may have participated as a transaminase instead of a dehydrase, could have potentially provided an amino sugar moiety in the O-antigen repeat.

It is difficult to determine the time point at which the remaining genes were introduced. Yet, by assuming that the directional mutation pressure is similar within the genera *Yersinia* and *Salmonella*, it can be postulated that *rfbI* was introduced into *Salmonella* spp. prior to the introduction of *ascD* into *Yersinia* serotype VA (or *rfbI* into *Yersinia* serotype IIA). Although the G+C contents of *ascD* (0.38) and *rfbI* (0.41) from *Salmonella* serotype IIA are relatively close, the low homology (51%) between these genes excludes a recombination event between *Yersinia* and *Salmonella* spp. as the possible source of these essential genes. If this had occurred, the codon usage and G+C contents of *ascD* and *rfbI* should have been essentially identical. As is illustrated in Fig. 1, the serological specificity of the corresponding strain is governed by the remaining genes, which code for the final enzymes responsible for the conversion of CDP-3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose (compound VIII) to paratose (*rfbS*), tyvelose (*rfbS* and *rfbE*), abequose (*rfbJ*), or ascarylose (*ascE* and *ascF*). The relationships among these genes are difficult to interpret; however, it is evident that all (except *rfbE*) are relatively homogeneous, as judged by the G+C content criteria. Although *rfbE* more closely resembles the E_3 genes (*ascD* and *rfbH*), the moderate residue homology observed between *ascF* and *rfbE* (which share identical N-terminal amino acid sequences for the first eight residues) suggests that *rfbE* may be evolutionarily linked to *ascF*. Furthermore, although genetic information on dideoxyhexose formation in *Yersinia* spp. is currently lacking, the abrupt divergence of region C can also be observed upon comparison of *Yersinia* serogroups IIA (abequose [18]) and VA (this study).

As outlined in Tables 2 and 3 and Fig. 5, the expression of the *asc* genes in both their native host (*Y. pseudotuberculosis*) and the current host used for the recombinant studies (*E. coli*) displayed a diverse range of yields of the corresponding enzymes. The greatly enhanced expression of the desired enzymes in *E. coli* may be ascribable to the use of a high-copy-number HB101(pUC)-based system for expression; however, the relative yield of these gene products via the recombinant systems remained practically the same as that found for the wild types. Although the level of expression is determined by an intricate combination of many factors, such as promoter strength and regulation, transcription efficiency, secondary structure and stability of the transcript, etc., perhaps a partial explanation for the variation in *asc*-encoded product formation relies on the distinct preference of the codon usage of each gene and the codon bias of the host organism. It has been well documented that *E. coli* and *Salmonella* spp. show a preference in favor of the codons recognized by the more prevalent RNA species, especially for genes expressed at high levels (34). Interestingly, in the *asc* genes with lower levels of expression, there is a particularly noticeable bias in favor of the Ile codon ATA (used 7 times in *ascD*, 8 times in *ascE*, and 17 times in *ascF*), which is rarely used in *E. coli* or *Salmonella* genes. Thus, the introduction of simple point mutations, designed to change the rare Ile ATA codon to the highly utilized ATC, may result in enhanced expression of these desired gene products. Since expression of certain genes at lower levels may be essential for the regulation of the overall biosynthetic pathway, further exploration of this variance in expression, in conjunction with study of the postulated feedback inhibition mechanisms among the various *asc* products, will certainly enhance our understanding of the formation of this class of unusual sugars.

In summary, the cloning and sequencing of the *asc* cluster has allowed the integration of the CDP-ascarylose biosynthetic pathway into an established evolutionary model for the construction of 3,6-dideoxyhexoses. Our extension of the evolutionary model is consistent with interspecific gene transfer events, as proposed in a recent review (27). However, the extreme variance, as a result of the seemingly rational insertion of regions A and C, does not necessarily reflect random genetic drift rather than adaptive change as previously implied (27). This study has broadened our knowledge of the variation of the O antigen in members of the family *Enterobacteriaceae* and serves to strengthen the foundation for studies of O-antigen polymorphism. In addition, this effort has yielded large amounts of the desired proteins, previously unattainable from the wild-type strain, which will prove invaluable for our future mechanistic studies of these intriguing pathways.

REFERENCES

1. **Ashwell, G., and J. Hackman.** 1971. The chemistry of the unique carbohydrates of bacterial lipopolysaccharides, p. 235–266. *In* G. Weinbaum, S. Kadis, and S. A. Ajl (ed.), Microbial toxins, vol. IV. Bacterial endotoxins. Academic Press, New York.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1989. Current protocols in molecular biology. John Wiley & Sons, New York.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
4. **Bishop, C. T., and H. J. Jennings.** 1982. Immunology of polysaccharides, p. 291–330. *In* G. O. Aspinall (ed.), The polysaccharides, vol. 1. Academic Press, New York.
5. **Blakeney, A. B., P. J. Harris, R. J. Henry, and B. A. Stone.** 1983. A simple and rapid preparation of alditol acetates for monosaccharide analysis. Carbohydr. Res. **113:**291.
6. **Brahmbhatt, H. N., P. Wyk, N. B. Quigley, and P. R. Reeves.** 1988. Complete physical map of the *rfb* gene cluster encoding biosynthetic enzymes for the O antigen of *Salmonella typhimurium* LT2. J. Bacteriol. **170:**98–102.
7. **Brenner, D. J.** 1978. Characterization and identification of Enterobacteriaceae by DNA hybridization, p. 71–117. *In* M. Stefanini (ed.), Progress in clinical pathology. Grunc and Stratton, New York.
8. **Brutlag, D. L., J.-P. Dautricourt, S. Maulik, and J. Relph.** 1990. Improved sensitivity of biological sequence database searches. Comput. Appl. Biosci. **6:**237–245.
9. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
10. **Glaser, L., and R. Zarkowsky.** 1971. Dehydration in nucleotide-linked deoxy sugar synthesis, p. 465–480. *In* P. Boyer (ed.), The enzymes, 3rd ed., vol. 5. Academic Press, New York.
11. **Gonzalez-Porque, P.** 1986. Cytidine diphosphate 4-keto-6-deoxy-D-glucose-3-dehydrogenase, p. 391–419. *In* D. Dolphin, R. Poulson, and O. Avramovic (ed.), Vitamin B_6 pyridoxal phosphate: chemical, biochemical, and medical aspects, part B. Wiley Interscience, New York.

12. **Grisebach, H.** 1978. Biosynthesis of sugar components of antibiotic substances. Adv. Carbohydr. Chem. Biochem. **35**:81–126.
13. **Hanessian, S.** 1966. Deoxy sugars. Adv. Carbohydr. Chem. Biochem. **21**:143–207.
14. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. Gene **28**:351–359.
15. **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.
16. **Kenne, L., and B. Lindberg.** 1983. Bacterial polysaccharides, p. 287–363. *In* G. O. Aspinall (ed.), The polysaccharides, vol. 2. Academic Press, New York.
17. **Kessler, A. C., P. K. Brown, L. K. Romana, and P. R. Reeves.** 1991. Molecular cloning and genetic characterization of the *rfb* region from *Yersinia pseudotuberculosis* serogroup IIA, which determines the formation of the 3,6-dideoxyhexose abequose. J. Gen. Microbiol. **137**:2689–2695.
18. **Kessler, A. C., A. Haase, and P. R. Reeves.** 1993. Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serogroup IIA. J. Bacteriol. **175**:1412–1422.
19. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
20. **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.
21. **Lo, S. F., V. P. Miller, Y. Lei, J. S. Thorson, H.-W. Liu, and J. L. Schottel.** 1994. CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase from *Yersinia pseudotuberculosis*: enzyme purification and characterization of the cloned gene. J. Bacteriol. **176**:460–468.
22. **Lüderitz, O., A. M. Staub, and O. Westphal.** 1966. Immunochemistry of O and R antigens of *Salmonella* and related *Enterobacteriaceae*. Bacteriol. Rev. **30**:192–248.
23. **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. Ajl (ed.), Microbial toxins, vol. IV. Bacterial endotoxins. Academic Press, New York.
24. **Miller, V. P., and H.-w. Liu.** 1992. Biosynthesis of 3,6-dideoxyhexoses: new evidence supporting a radical mechanism for C-3 deoxygenation. J. Am. Chem. Soc. **114**:1880–1881.
25. **Miller, V. P., J. S. Thorson, O. Ploux, S. F. Lo, and H.-w. Liu.** 1993. Cofactor characterization and mechanistic studies of CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase: exploration into a novel enzymatic C-O bond cleavage event. Biochemistry **32**:11934–11942.
26. **Raetz, C. R. H.** 1990. Biochemistry of endotoxins. Annu. Rev. Biochem. **59**:129–170.
27. **Reeves, P. R.** 1993. Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9**:17–22.
28. **Roantree, R. J.** 1967. Salmonella O antigen and virulence. Annu. Rev. Microbiol. **21**:443–466.
29. **Rubenstein, P. A., and J. L. Strominger.** 1974. Enzymatic synthesis of cytidine diphosphate 3,6-dideoxyhexoses. IX. Purification and properties of the cytidine diphosphate-D-glucose pyrophosphorylase from *Pasteurella pseudotuberculosis* type V. J. Biol. Chem. **249**:3789–3796.
30. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
31. **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.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
33. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential application. Nucleic Acids Res. **15**:1281–1295.
34. **Sueoka, N.** 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85**:2653–2657.
35. **Thorson, J. S., T. M. Kelly, and H.-W. Liu.** 1994. Cloning, sequencing, and overexpression in *Escherichia coli* of the α-D-glucose-1-phosphate cytidylyltransferase gene isolated from *Yersinia pseudotuberculosis*. J. Bacteriol. **176**:1840–1849.
36. **Thorson, J. S., and H.-w. Liu.** 1993. Characterization of the first PMP-dependent iron-sulfur containing enzyme which is essential for the biosynthesis of 3,6-dideoxyhexoses. J. Am. Chem. Soc. **115**:7539–7540.
37. **Thorson, J. S., and H.-w. Liu.** 1993. Coenzyme B_6 as a redox cofactor—a new role for an old coenzyme? J. Am. Chem. Soc. **115**:12177–12178.
38. **Thorson, J. S., S. F. Lo, and H.-w. Liu.** 1993. Molecular basis of 3,6-dideoxyhexose biosynthesis: elucidation of CDP-ascarylose biosynthetic genes and their relationship to other 3,6-dideoxyhexose pathways. J. Am. Chem. Soc. **115**:5827–5828.
39. **Thorson, J. S., S. F. Lo, H.-w. Liu, and C. R. Hutchinson.** 1993. Biosynthesis of 3,6-dideoxyhexoses: new mechanistic reflections upon 2,6-dideoxy, 4,6-dideoxy, and amino sugar construction. J. Am. Chem. Soc. **115**:6993–6994.
40. **Thorson, J. S., E. Oh, and H.-w. Liu.** 1992. Studies of complex enzymatic pyridine nucleotide-dependent transformations: structure and cofactor binding correlation in CDP-D-glucose oxidoreductase. J. Am. Chem. Soc. **114**:6941–6942.
41. **Vesterberg, O.** 1971. Staining of protein zones after isoelectric focusing in polyacrylamide gels. Biochim. Biophys. Acta **243**:345–348.

41a. **Weigel, T. M.** 1991. Ph.D. thesis. University of Minnesota, Minneapolis.

42. **Weigel, T. M., L.-d. Liu, and H.-w. 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.
43. **Westphal, O., and O. Lüderitz.** 1960. 3,6-Didesoxy-hexosen—chemie und biologie. Angew. Chem. **72**:881–891.
44. **Wierenga, R. K., P. Terpstra, and W. G. J. Hol.** 1986. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins using an amino acid sequence fingerprint. J. Mol. Biol. **187**:101–107.
45. **Williams, N. R., and J. D. Wander.** 1980. Deoxy and branched-chain sugars, p. 761–798. *In* W. Pigman and D. Horton (ed.), The carbohydrates: chemistry and biochemistry, vol. 1B. Academic Press, New York.
46. **Wyk, P., and P. R. Reeves.** 1989. Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B *Salmonella*: homology with galactose epimerase. J. Bacteriol. **171**:5687–5693.
47. **Yates, R. W.** 1991. *In* Handbook of biochemistry, p. 708. Plenum Press, New York.
48. **Yu, Y., R. N. Russell, J. S. Thorson, L.-d. Liu, and H.-w. Liu.** 1992. Mechanistic studies of the biosynthesis of 3,6-dideoxyhexoses in *Yersinia pseudotuberculosis*: purification and stereochemical analysis of CDP-D-glucose oxidoreductase. J. Biol. Chem. **267**:5868–5875.