# Molecular Basis of 3,6-Dideoxyhexose Biosynthesis: Elucidation of CDP-Ascarylose Biosynthetic Genes and Their Relationship to Other 3,6-Dideoxyhexose Pathways[1]

Jon S. Thorson, Stanley F. Lo, and Hung-wen Liu*

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

*Received March 12, 1993*

Among the vast number of monosaccharides identified as components of *O*-specific polysaccharides, the 3,6-dideoxyhexoses have attracted particular attention due to their high immunogenicity.[2] Although five of the eight possible 3,6-dideoxyhexoses (abequose, colitose, paratose, tyvelose, and ascarylose) have been found in various strains of *Enterobacteriaceae*,[3] only the formation of ascarylose (**1**, *asc*), present in the *O*-antigen of *Yersinia pseudotuberculosis* VA, has been studied at the enzymatic level.[4] The proposed biosynthetic pathway for ascarylose (Scheme I) begins with the conversion of D-glucose-1-phosphate (**2**) to CDP-D-glucose (**3**), catalyzed by D-glucose-1-phosphate cytidylyltransferase ($E_p$).[5] Subsequent transformation to CDP-6-deoxy-D-*glycero*-L-*threo*-4-hexulose (**4**) is carried out by a $NAD^+$-dependent CDP-D-glucose 4,6-dehydratase ($E_{od}$).[6] The C-3 deoxygenation to 3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose (**5**) is accomplished in two consecutive reactions[4] catalyzed by CDP-6-deoxy-D-*glycero*-L-*threo*-4-hexulose-3-dehydrase ($E_1$), a PMP-dependent enzyme,[7] and CDP-6-deoxy-$\Delta^{3,4}$-glucoseen reductase ($E_3$), a [2Fe–2S]-containing flavoprotein.[8] The final steps from **5** to CDP-ascarylose (**1**) include C-5 epimerization and stereospecific C-4 reduction, which are presumably two separate enzymatic processes. However, since the last two catalysts are yet to be isolated, one cannot exclude the possibility that these transformations are mediated by a single protein. We have recently cloned and characterized the gene cluster encoding the proteins of this important biosynthetic pathway. This study represents the first example of the complete characterization–correlation of gene products with their catalytic activities for an entire 3,6-dideoxyhexose pathway,[9] providing distinct molecular evidence supporting the postulated sequence of events in the biosynthesis of not only ascarylose but also other 3,6-dideoxyhexoses.

The isolation of three overlapping clones[10] has resulted in the identification of a fragment of *Yersinia* DNA (>6.5 kb) containing a cluster of six complete genes (Scheme I). The first four genes have been overexpressed in *Escherichia coli* and the recombinant proteins purified to homogeneity. Their identities have been elucidated by specific activity assay[11] and *N*-terminal analysis and are assigned as (in 5′ to 3′ order) *asc*D/$E_3$, *asc*A/$E_p$, *asc*B/$E_{od}$, and *asc*C/$E_1$. Translation of the fifth orf (open reading frame, 3′ from *asc*C) results in a protein (21 663 Da subunit size) 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 **5** is not expected to require any cofactor on the basis of an enolization–reprotonation mechanism, this orf (*asc*E) most likely encodes CDP-3,6-dideoxy-D-*glycero*-D-*glycero*-4-hexulose-5-epimerase ($E_{ep}$). The last enzyme in the proposed ascarylose scheme is a NAD(P)H-dependent C-4 reductase ($E_{red}$). Examination of the final complete orf (3′ from the assigned *asc*E gene) reveals a gene that, when translated, contains the necessary nicotinamide binding motif[6b,12] in a protein with a subunit size of 33 220 Da. Hence, we have assigned this gene *asc*F, which could code for CDP-3,6-dideoxy-L-*glycero*-D-*glycero*-4-hexulose-4-reductase ($E_{red}$).

Support for the above assignments is provided by a GC-MS assay (Scheme II)[7a] in which the use of $NaBD_4$ to reduce the resulting ascarylose (**1**) allows indisputable differentiation among the enzymatic products. For this assay, homogeneous $E_p$, $E_{od}$, $E_1$, and $E_3$ were used to generate **5**. To this mixture was added NADPH and NADH, along with the crude *asc*E and *asc*F products via expression from pUC18-based plasmids in *E. coli* HB101. Two blanks were run in parallel, the first with crude enzymes boiled prior to use and the second simply utilizing crude extracts of pUC18 in HB101. Results obtained from GC-MS (EI and CI) analysis are consistent with the production of ascarylitol tetraacetate (**A**), the only possible monodeuterated 3,6-dideoxyhexitol derivative from this mixture, thereby strongly supporting the existence of both the epimerase and the reductase in the expressed extracts. In the presence of denatured $E_{ep}$ and $E_{red}$ or in the pUC18-HB101 blank, only the dideuterated species **B** was formed.[13] While in separate assays the formation of **B** as the final product when crude *asc*F was used independently in the incubation is expected, identification of **B** after incubation with crude *asc*E alone is a surprise, since C-5 epimerization is expected to generate a 3,6-dideoxyhexitol product having a different GC retention time but an identical MS fragmentation pattern of **B**.[14] Thus, the apparent inability of the *asc*E and *asc*F gene products to function separately suggests that these gene products must participate together, perhaps in an $\alpha\beta$ system.[15] Although it is commonly believed that two enzymes are required for epimerase–reductase processes,[16] there exists one example in which a single protein

(1) This work was supported in part by the National Institutes of Health Grant GM 35906. H.-w.L. is the recipient of a National Institutes of Health Research Career Development Award (GM 00559).

(2) (a) Bishop, C. T.; Jennings, H. J. In *The Polysaccharides*; Aspinall, G. O., Ed.; Academic Press: New York, 1982; Vol. 1, p 291. (b) Raetz, C. R. H. *Annu. Rev. Biochem.* **1990**, *59*, 129.

(3) Williams, N. R.; Wander, J. D. In *The Carbohydrates: Chemistry and Biochemistry*; Pigman, W., Horton, D., Eds.; Academic Press: New York, 1980; Vol. 1B, p 761.

(4) Gonzalez-Porque, P. In *Vitamin $B_6$ Pyridoxal Phosphate, Chemical, Biochemical, and Medical Aspects*; Dolphin, D., Poulson, R., Avramovic, O., Eds.; Wiley-Interscience: New York, 1986; Part B, p 392.

(5) Rubenstein, P. A.; Strominger, J. L. *J. Biol. Chem.* **1974**, *249*, 3789.

(6) (a) Yu, Y.; Russell, R. N.; Thorson, J. S.; Liu, L.-d.; Liu, H.-w. *J. Biol. Chem.* **1992**, *267*, 5868. (b) Thorson, J. S.; Oh, E.; Liu, H.-w. *J. Am. Chem. Soc.* **1992**, *114*, 6941.

(7) (a) Weigel, T. M.; Liu, L.-d.; Liu, H.-w. *Biochemistry* **1992**, *31*, 2129. (b) Weigel, T. M.; Miller, V. P.; Liu, H.-w. *Biochemistry* **1992**, *31*, 2140.

(8) (a) Han, O.; Miller, V. P.; Liu, H.-w. *J. Biol. Chem.* **1990**, *265*, 8033. (b) Miller, V. P.; Liu, H.-w. *J. Am. Chem. Soc.* **1992**, *114*, 1890.

(9) Although a few *O*-antigen biosynthetic genes have been cloned, the actual functions of a number of the key 3,6-dideoxyhexose genes have never been established: (a) Liu, D.; Verma, N. K.; Romana, L. K.; Reeves, P. R. *J. Bacteriol.* **1991**, *173*, 4814. (b) Kessler, A. C.; Brown, P. K.; Romana, L. K.; Reeves, P. R. *J. Gen. Microbiol.* **1991**, *137*, 2689.

(10) The experimental details will be published elsewhere (Thorson, J. S.; Lo, S. F.; Liu, H.-w.).

(11) The activity of $E_p$ is based upon a coupled assay with $E_{od}$.[5] The product of the $E_{od}$ reaction, which may be derived directly from CDP-D-glucose for the $E_{od}$ assay, can be observed at 320 nm under basic conditions.[6a] The activity of $E_1$ is determined on the basis of tritium release in the presence of [$^3$H]PMP or by coupling with $E_3$ and observing product formation via GC-MS.[7a] Enzyme $E_3$ is assayed on the basis of its NADH:DCPIP oxidoreductase activity[8a] or by analysis of product formation by GC-MS.[7a]

(12) Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing: New York, 1991; p 141.

(13) The byproducts of this assay include **D** and **C**. However, since their GC-MS behaviors are distinct, their existence in the analytical sample imposed no complication on this assay.

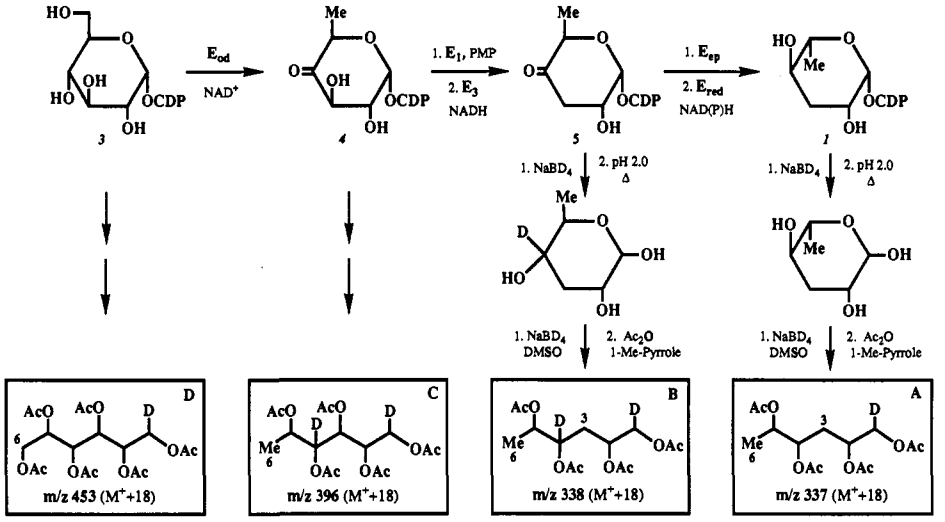
(14) Since the GC retention times of **A** and **B** are distinct, epimerization alone is expected to give a compound structurally identical to **A** with the labeling patterns of **B**. Thus, the lack of a GC peak corresponding to **A** in the presence of *asc*E product alone supports our assignment. However, low expression of the individual gene products from the corresponding constructs could also lead to a similar observation.

(15) However, the fact that these gene products must associate for activity, in conjunction with the lack of a cofactor motif, does not definitively prove that *asc*E carries the epimerase activity.

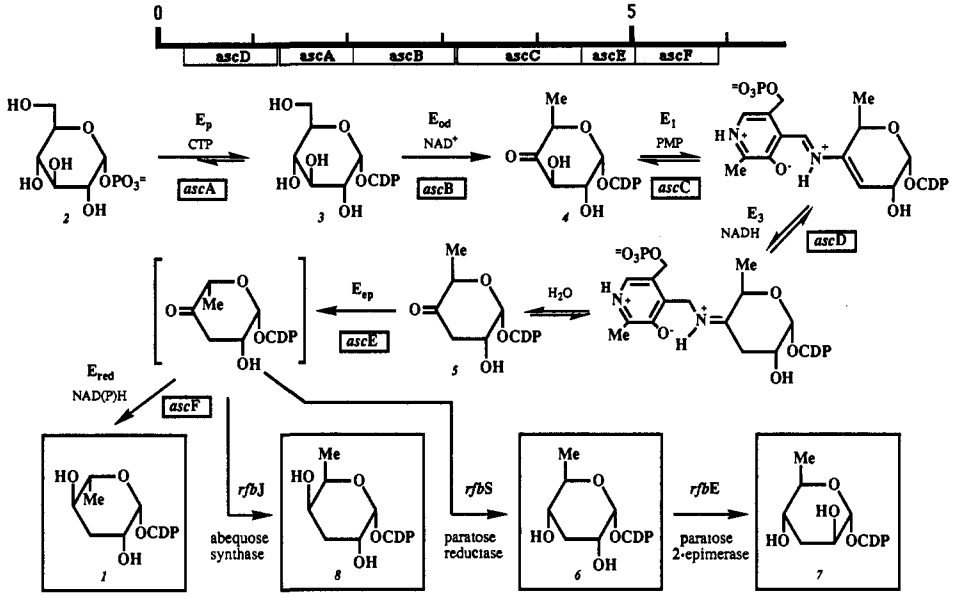
(16) Gaugler, R. W.; Gabriel, O. *J. Biol. Chem.* **1973**, *248*, 6041. (b) Melo, A.; Glaser, L. *J. Biol. Chem.* **1967**, *243*, 1475. (c) Marumo, K.; Lindquist, L.; Verma, N.; Weintraub, A.; Reeves, P. R.; Lindberg, A. A. *Eur. J. Biochem.* **1992**, *204*, 539.

0002-7863/93/1515-5827$04.00/0 

**Scheme I**




**Scheme II**




appears to possess both epimerase and reductase capabilities.[17] The present ascarylose pathway implies that general 3,6-dideoxyhexose biosynthesis in *Enterobacteriaceae* can utilize either two separate entities, as in the production of tyvelose, or a single protein consisting of two different polypeptides, which is now revealed in the ascarylose pathway, to catalyze the final epimerization and reduction.

As shown in Scheme I, all 3,6-dideoxyhexose biosynthetic pathways (colitose the exception) share the common intermediate **5** and diverge beyond. For paratose (**6**), C-4 reduction is catalyzed by paratose reductase to the corresponding ribose configuration. Subsequently, the CDP-paratose-2-epimerase-catalyzed C-2 epimerization results in the formation of tyvelose (**7**). In the biosynthesis of abequose (**8**), C-4 reduction is catalyzed by abequose synthase. Although the genes for these enzymes have been recently cloned from *Salmonella* and sequenced,[18] due to the lack of facile assays several of the cloned genes essential for the early steps of their biosynthesis have continually evaded designation. Interestingly, data bank searches using the translated sequences of each of the newly elucidated *asc* gene products reveal extensive residue identity[19] with a translated region of the recently released *S. typhimurium rfb* (*O*-antigen) cluster.[18b] Since the partially labeled gene order of the *rfb* cluster was found to be identical to that of the ascarylose cluster, we can now also assign the *S. typhimurium rfb* orf 10.4 as coding for the *Salmonella* $E_1$ equivalent and *rfb* orf 7.6 as coding for the *Salmonella* $E_3$ equivalent in the abequose, paratose, and tyvelose biosynthetic pathways.[9a,18b,20] Thus, although studies of the principles underlying 3,6-dideoxyhexose polymorphism in *Y. pseudotuberculosis* are still in their infancy,[9b] the unambiguous assignment of these genes through our current study has laid a strong foundation for the study of 3,6-dideoxyhexose polymorphism in *Salmonella* and *Yersinia* in particular and *O*-antigen variation in general.[21]

(17) Chang, S.; Duerr, B.; Serif, G. *J. Biol. Chem.* **1988**, *263*, 1693.
(18) (a) Wyk, P.; Reeves, P. *J. Bacteriol.* **1989**, *171*, 5687. (b) Jiang, X.-M.; Neal, B.; Santiago, F.; Lee, S. J.; Romana, L. K.; Reeves, P. R. *Mol. Microbiol.* **1991**, *5*, 695.
(19) The *asc*C-derived amino acid sequence was found to be 73% homologous to that of the *Salmonella rfb* orf 10.4, and the *asc*D-derived sequence displayed 53% homology with the translated sequence of *rfb* orf 7.6 (previously thought not be part of the abequose biosynthetic gene cluster).[18b] In addition, 78% and 71% residue identities were found with the *asc*A-*rfb*F and *asc*B-*rfb*G amino acid sequences, respectively.
(20) (a) Brown, P. K.; Romana, L. K.; Reeves, P. R. *Mol. Microbiol.* **1991**, *5*, 1873. (b) Brown, P. K.; Romana, L. K.; Reeves, P. R. *Mol. Microbiol.* **1992**, *6*, 1385. (c) Wang, L.; Romana, L. K.; Reeves, P. R. *Genetics* **1992**, *130*, 429.
(21) Reeves, P. R. *FEMS Microbiol. Lett.* **1992**, *79*, 509.