Bifunctional Epimerase/Reducate Enzymes Facilitate the Modulation of 6-Deoxy-Heptoses Found in the Capsular Polysaccharides of Campylobacter jejuni

Dao Feng Xiang, Manas K. Ghosh, Alexander S. Riegert, James B. Thoden, Hazel M. Holden,* and Frank M. Rauschel*

ABSTRACT: Campylobacter jejuni is a human pathogen and the leading cause of food poisoning in the United States and Europe. Surrounding the exterior surface of this bacterium is a capsular polysaccharide (CPS) that consists of a repeating sequence of common and unusual carbohydrate segments. At least 10 different heptose sugars have thus far been identified in the various strains of C. jejuni. The accepted biosynthetic pathway for the construction of the 6-deoxy-heptoses begins with the 4,6-dehydration of GDP-D-glycero-D-manno-heptose by a dehydratase, followed by an epimerase that racemizes C3 and/or C5 of the product GDP-6-deoxy-4-keto-D-lyxo-heptose. In the final step, a C4-reductase catalyzes the NADPH reduction of the resulting 4-keto product. However, in some strains and serotypes of C. jejuni, there are two separate C4-reductases with different product specificities in the gene cluster for CPS formation. Five pairs of these tandem C4-reductases were isolated, and the catalytic properties were ascertained. In four out of five cases, one of the two C4-reductases is able to catalyze the isomerization of C3 and C5 of GDP-6-deoxy-4-keto-D-lyxo-heptose, in addition to the catalysis of the reduction of C4, thus bypassing the requirement for a separate C3/C5-isomerase. In each case, the 3′-end of the gene for the first C4-reductase contains a poly-G tract of 8–10 guanine residues that may be used to control the expression and/or catalytic activity of either C4-reductase. The three-dimensional structure of the C4-reductase from serotype HS:15, which only does a reduction of C4, was determined to 1.45 Å resolution in the presence of NADPH and GDP.

INTRODUCTION

Campylobacter jejuni is a human pathogen that is the leading cause of food poisoning in the United States and Europe. The external surface of this bacterium is coated with a capsular polysaccharide (CPS) that helps it to evade the host immune system. The various bacterial strains of C. jejuni have capsular polysaccharides with different repeating sugar sequences. At least 33 different serotypes of C. jejuni have been isolated, and the chemical structures of the capsular polysaccharides for 12 of these strains have been reported. The repeating polysaccharides in the HS:4 and HS:15 serotypes are illustrated in Figure 1. In addition to a carbohydrate backbone, some CPS structures are further modified by methylations, methyl phosphorylamidations, and amidations. Among the most common monosaccharide residues that are found in the structurally characterized capsular polysaccharides are heptoses. These relatively rare sugars include those that differ in stereochemistry at C2 through C6, in addition to deoxy versions at C3 and C6.

It is currently thought that all of the heptoses found in the capsular polysaccharides in C. jejuni originate from GDP-D-glycero-D-manno-heptose (1). For the biosynthesis of the 6-deoxy-heptoses, the generally accepted biosynthetic pathway involves the successive activities of a 4,6-dehydratase, a C3- or C3/C5-epimerase, and a C4-reductase, as shown in Figure 2. From this biosynthetic pathway, eight different 6-deoxyheptoses are possible (three stereocenters at C3, C4, and C5), and we and Creuzenet’s group have characterized the C4-reductases from multiple serotypes that can synthesize six of the eight possible products (Figure 3). A bioinformatic
analysis of the DNA-sequence serotypes of *C. jejuni* has demonstrated that there is a single common 4,6-dehydratase,9 but two types of epimerases are utilized for the biosynthesis of the 6-deoxy-heptoses.10 We have demonstrated that the largest class of epimerases (from serotypes HS:3, HS:4, HS:8, HS:10, HS:12, HS:23/36, HS:29, HS:33, and HS:41) is limited to the epimerization of C3, whereas the second-most populated class of epimerases (from serotypes HS:2, HS:15, HS:32, HS:42, and HS:63) catalyzes the racemization of C3 and C5.10

For the eight C4-reductases that have been functionally characterized to date (serotypes HS:2, HS:3, HS:4, HS:15, HS:23/36, HS:33, HS:42, and HS:53), the epimerase in the gene cluster of that serotype is a C3/C5-epimerase if C5 is epimerized but is a C3-epimerase if C5 is not epimerized in the final 6-deoxy heptose product.11 However, for some serotypes (namely, HS:8, HS:10, HS:29, HS:41, and HS:63), there are two nonredundant C4-reductases in the gene cluster.12 In each of these cases, the C4-reductases are not expected to synthesize the same product. For example, in the HS:10 serotype, the two C4-reductases (dubbed HS:10A and HS:10B) are predicted to synthesize GDP-6-deoxy-D-altro-heptose (4) and GDP-6-deoxy-L-galacto-heptose (6), respectively.11 In the chemically characterized CPS from serotype HS:10, the heptose has been identified as 6-deoxy-L-galacto-heptose (6), and thus, the C4-reductase dubbed HS:10B has been used to catalyze the final product.19 This observation is curious since the single epimerase found in the gene cluster used for the biosynthesis of the CPS in serotype HS:10 is a C3-epimerase.12 Yet, the formation of 6-deoxy-L-galacto-heptose (6) requires the epimerization of C5.

In this paper, we have explored three overlapping issues regarding the biosynthesis of 6-deoxy-heptoses within the CPS of *C. jejuni*. How can C5 be epimerized in those serotypes that possess a single epimerase but can apparently only racemize C3? Why do some strains of *C. jejuni* possess two distinct C4-reductases with different product specificities? Finally, what dictates whether one or the other of these C4-reductases is expressed for the biosynthesis of a specific 6-deoxy-heptose product? Here, we demonstrate that a subset of C4-reductases in *C. jejuni* can catalyze the epimerization of C3/C5 in addition to the reduction of C4. We also demonstrate that in those serotypes, which possess two distinct C4-reductases with different product outcomes, the gene for one of them has a poly-G tract near the 3′-end of the gene. This poly-G tract may be utilized to regulate the expression or catalytic activity of one or the other of the two C4-reductases in that particular gene cluster via slipped-strand mispairing phase variation.18,21

### MATERIALS AND METHODS

**Materials.** Unless otherwise noted, all chemicals used in this study were obtained from Sigma-Aldrich, Carbosynth, or GE Healthcare Biosciences. The bacterial growth medium, agarose, and isopropyl-β-D-thiogalactopyranoside (IPTG) were purchased from Research Products International (RPI). *Escherichia coli* strain BL21-Gold (DE3) and the restriction enzyme DpnI were obtained from New England Biolabs. PhT Turbo DNA polymerase was purchased from Agilent. DNA sequencing was conducted by Eton Biosciences, Inc. GDP-D-glycero-β-D-manno-heptose (1) was enzymatically synthesized and purified as described previously.19

**Plasmid Construction.** The genes encoding the C4-reductases from different serotypes of *C. jejuni* were codon-optimized for expression in *E. coli* and ordered from either Twist Biosciences (San Francisco, CA) or GenScript (Piscataway, NJ) in a pET28a (+) expression vector with an N-terminal hexa-histidine tag. The UniProt codes for the two C4-

---

**Figure 1.** Structures of the repeating capsular polysaccharides in the HS:4 and HS:15 serotypes. The HS:4 serotype contains N-acetyl-D-glucosamine and 6-deoxy-D-id-o-heptose (or L-glycero-D-id-o-heptose), whereas the HS:15 serotype contains L-arabinose and 6-deoxy-L-gulo-heptose.12,13 The CPS from serotype HS:4 has also been shown to be decorated with an O-methyl phosphoramidate moiety on the ido-heptose at C2 or C7.12

**Figure 2.** Biosynthetic pathway for the formation of 6-deoxy-heptoses.

**Figure 3.** Structures of six GDP-6-deoxy-heptoses that have been isolated from the catalytic activity of eight C4-reductases from various strains of *C. jejuni*.12,16
reductases from serotypes HS:8, HS:10, HS:29, HS:41, and HS:63 are as follows: F2X762 (dubbed HS:8A), F2X761 (dubbed HS:8B), F2X783 (dubbed HS:10A), F2X782 (dubbed HS:10B), A0A0U3CA27 (dubbed HS:29A), A0A0U3CA33 (dubbed HS:29B), Q5M6T8 (dubbed HS:41A), Q5M6T9 (dubbed HS:41B), A0A0S2CGF3 (dubbed HS:63A), and A0A0S2CGF0 (dubbed HS:63B). The UniProt codes for the single C4-reductases from serotypes HS:15 and HS:42 are A0A3Z9HSX9 and F2X7E5, respectively.

The sequences of the expressed proteins are presented in Figure S1. We also prepared proteins with a truncation at either the N- or C-terminal ends of specific C4-reductases. A total of 28 amino acids from the N-terminal ends of the HS:8B and HS:10B C4-reductases was removed via the synthesis of the appropriate expression plasmid. For the HS:8A, HS:29A, and HS:63A C4-reductases, we removed 44, 37, and 36 amino acids, respectively, from the C-terminal ends of these proteins via reconstruction of the expression plasmid. The sequences of these truncated proteins are found in Figure S2.

**Protein Expression and Purification.** The plasmid DNA for the production of the HS:41B C4-reductase was used to transform BL21 (DE3) cells, and a single colony was used to inoculate 5 mL of sterile LB medium, which was then cultured for 15 h at 37 °C. The 5 mL culture was subsequently used to inoculate 1 L of LB medium supplemented with 50 μg/mL kanamycin and then grown at 37 °C until the OD<sub>600</sub> reached ~0.6. At this point, 1.0 mM IPTG was added to the culture, the temperature was decreased to 20 °C, and growth was continued for an additional 18 h. The cells were harvested by centrifugation at 7000g at 4 °C, frozen in liquid N₂, and stored at −80 °C prior to purification of the protein. For purification of the HS:41B C4-reductase, the cell pellets from 3 L (~10 g) were resuspended in 100 mL of binding buffer (50 mM HEPES, pH 8.0, 0.25 M KCl, and 10 mM imidazole) and then disrupted by sonication using a W-380 ultrasonic processor (Heat Systems-Ultrasonics, Inc.). The cell lysate was clarified by centrifugation before being passed through a 0.2 μm syringe filter (VWR) and then loaded onto a 5 mL HisTrap HP column (GE Healthcare) attached to an NGC liquid chromatography system (Bio-Rad) previously equilibrated with binding buffer. The His-tagged protein was eluted with a 0 to 50% gradient of elution buffer (50 mM HEPES, pH 8.0, 0.25 M KCl, and 0.5 M imidazole). The fractions were pooled based on the SDS-PAGE gel electrophoresis results and dialyzed against 50 mM HEPES containing 0.25 M KCl to remove the imidazole. Protein purity was judged to be >95% based on SDS–PAGE. The yield of purified HS:41B C4-reductase was ~30 mg per liter of cell culture. The other C4-reductases were expressed and purified using the same procedures as that for the HS:41B C4-reductase. The yield of the proteins varied from 7 to 40 mg per liter of cell culture. The GDP-d-glycerol-3-phosphate dehydratase from serotypes HS:23/36 and the C3- and C3/C5-epimerases from different serotypes of C. jejuni were purified as described previously.

**Mutagenesis.** The C110S variant of the HS:41B C4-reductase was constructed using the QuickChange site-directed mutagenesis method from Agilent Technologies. The entire gene was subsequently sequenced to ensure that there were no unwanted mutations, deletions, or insertions occurring during the process of mutagenesis. The gene used for the production of the S242F variant of the HS:10A C4-reductase was synthesized by Twist Biosciences (San Francisco, CA) and inserted into a pET28a (+) vector with a N-terminal hexahistidine purification tag. These two proteins were expressed and purified to homogeneity using the same procedure that was used for the related wild-type proteins. The protein purity was judged to be >95% based on SDS–PAGE. The protein sequences of the two variants are provided in Figure S3.

**Isolation of the C4-Reductase Reaction Products.** To identify the reaction product from each C4-reductase, a 1 mL reaction mixture containing 4.0 mM GDP-d-glycerol-3-phosphate (1), 4.0 μM GDP-d-glycerol-3-phosphate-4,6-dehydratase, 4.0 μM C3 or C3/C5 epimerase (if added) and 4.0 μM C4-reductase, 10 mM acetaldehyde, 0.15 mM NADPH, 2.3 U aldehyde dehydrogenase, and 50 mM potassium phosphate buffer, pH 7.5 (or pH 7) was incubated at 22 °C for 18 h. For the HS:29A and HS:63A C4-reductases, the incubation time was increased to 24 h. The proteins from each reaction were removed by filtration using PALL 10K membrane filters, and the filtrate was loaded onto a 1 mL HiTrap Q HP anion exchange column, which was connected to an NGC HPLC (Bio-Rad) system. The product was eluted using a linear gradient of 500 mM NH₄HCO₃ and the products pooled based on the absorbance at 255 nm. The isolated samples were lyophilized using a freeze dryer (LABCONCO) and subsequently redissolved in 50 mM potassium phosphate buffer in D₂O, pD 7.5, for ¹H NMR and ESI-MS analysis.

**Determination of Protein Concentrations.** Concentrations of the proteins were determined spectrophotometrically using computationally derived molar absorption coefficients at 280 nm. The values of ε<sub>280</sub> (M<sup>−1</sup> cm<sup>−1</sup>) used for the C4-reductase dúbbed HS:8A, HS:8B, HS:10A, HS:10B, HS:29A, HS:29B, HS:41A, HS:41B, HS:63A, and HS:63B are 40,340, 41,661, 40,340, 44,810, 44,350, 46,300, 41,830, 47,790, 41,830, and 47,790, respectively.

**Determination of Kinetic Constants.** The kinetic parameters for the reactions catalyzed by the various C4-reductases were determined by coupling the reaction with a C3- or C3/C5-epimerase in the presence of NADPH. The reaction mixtures contained 1.0 mM GDP-6-deoxy-4-keto-d-lyxo-heptose (2), 1.0 μM C3- or C3/C5-epimerase, 0.3 mM NADPH, and 50 mM HEPES/K<sup>+</sup> buffer at pH 7.5. Absorbance decreases at 340 nm from the conversion of NADPH to NADP<sup>+</sup> were monitored spectrophotometrically at 25 °C with a SpectraMax<sub>340</sub> UV–visible spectrophotometer. The rate constants are reported as the average of three experiments.

**Sequence Similarity Network for the 25 C4-Reduc- tases.** The sequence similarity network (SSN) for 25 C4-reductases was generated using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST, http://efi-igb.illinois.edu/efi-est/). The FASTA protein sequences of the 25 C4-reductases from 20 different serotyped strains of C. jejuni were obtained from UniProt and submitted to the EFI-EST webtool. All network layouts were created and visualized using Cytoscape 3.8.2.

**Crystallization and Structure Determination.** For crystallization, the N-terminally tagged C4-reductase from HS:15 was dialyzed against 10 mM TRIS and 200 mM NaCl at pH 8.0, followed by concentration to 22 mg/mL. Crystallization conditions were screened at room temperature via the hanging drop method of vapor diffusion using protein that had been preincubated with 5.0 mM GDP and 5.0 mM NADPH. Diffraction quality crystals were grown from 18–22% poly-
(ethylene glycol)-3350, 2% isopropanol, and 100 mM HEPES (pH 7.5). The crystals belonged to the triclinic space group P1 with unit cell dimensions of $a = 51.8$ Å, $b = 56.6$ Å, $c = 67.4$ Å, $\alpha = 88.0^\circ$, $\beta = 82.0^\circ$, and $\gamma = 70.1^\circ$. The asymmetric unit contained one dimer. Crystals were serially transferred to a cryoprotectant solution composed of 30% poly(ethylene glycol)-3350, 250 mM NaCl, 2% isopropanol, 5.0 mM GDP, 5.0 mM NADPH, and 18% ethylene glycol for X-ray data collection. X-ray data were collected to 1.45 Å resolution using a BRUKER D8-VENTURE sealed tube system equipped with HELIOS optics and a PHOTON II detector, processed with SAINT and scaled with SADABS (Bruker AXS).

The structure was solved via molecular replacement with PHASER using as a search model the coordinates from PDB entry 7M13. Iterative cycles of model building with COOT and refinement with REFMAC led to a final X-ray model with an overall $R$-factor of 19.0% at 1.45 Å resolution.

Determinant structures of the capsular polysaccharides. The final enzyme in the biosynthetic pathway, the C4-reductase, determines the ultimate stereochemical outcome of the 6-deoxy-heptoses in the CPS via the selection of the specific diastereomeric GDP-6-deoxy-4-keto-heptose intermediate to be reduced. We have previously identified 25 different C4-reductases contained within the gene clusters for the biosynthesis of the CPS in various strains of C. jejuni. The sequence similarity network at a sequence identity cutoff of 89% segregates these enzymes into nine different clusters (Figure 4).

![Figure 4](https://example.com/fi4.png)

**Figure 4.** Sequence similarity network for 25 C4-reductases identified within various strains of C. jejuni at a sequence identity cutoff of 89%. The green-colored nodes denote C4-reductases that have previously been isolated and the products characterized. The yellow-colored nodes indicate C4-reductases that are characterized in this investigation. The gray-colored nodes are for reductases likely involved in the biosynthesis of 3,6-dideoxy heptoses, and those colored blue are additional C4-reductases that have not been purified. The product specificities for the numbered clusters are as follows: Group-1 (GDP-6-deoxy-altro-heptose (4)); Group-2 (GDP-6-deoxy-gulo-heptose (6)); Group-3 (GDP-6-deoxy-ido-heptose (5)); Group-4 (GDP-6-deoxy-l-gulo-heptose (8)); Group-5 (GDP-6-deoxy-n-manno-heptose (3)); Group-6 (GDP-6-deoxy-l-gluco-heptose (7)); and Group-7 (GDP-6-deoxy-o-ido-heptose (5)).

### RESULTS AND DISCUSSION

**The C4-Reductases from C. jejuni.** The 6-deoxy-heptose moieties are an integral part of the capsular polysaccharides in C. jejuni. The generally accepted biosynthetic pathway for the construction of these sugars is initiated by the dehydrogenation of GDP-3-deoxy-glycerol-6-phosphate (1), followed by an epimerization of C3 and/or C5, and finalized by the reduction of the carbonyl group at C4 (Figure 2). Thus far, six different 6-deoxy-heptoses have been identified in the chemically proteins, each from a different cluster, have been purified to homogeneity, and the product specificity has been mapped to the SSN. Two of the clusters (labeled as Group-8 and Group-9) are likely involved in the biosynthesis of 3,6-dideoxy heptoses, while the remaining seven clusters specify proteins for six of the eight possible 6-deoxy heptoses. For the eight previously purified C4-reductases, the corresponding C3- or C5/C5'-epimerase matches the requirement of the final product in the sense that when C5 is required to be epimerized, the corresponding gene cluster contained a C3/C5'-epimerase, and when C5 is not required to be epimerized, it contained a C3-epimerase.

The product specificities for 14 of the remaining C4-reductases can be predicted based on the product outcomes for the eight C4-reductases that have been functionally characterized, and these predictions are indicated within the sequence similarity network in Figure 4. These assignments raise two important issues. Five of the serotypes (HS:8, HS:10, HS:29, HS:41, and HS:63) possess two different C4-reductases. For example, in serotype HS:10, the enzyme

---

### Table 1. X-ray Data Collection and Model Refinement Statistics

<table>
<thead>
<tr>
<th>PDB code</th>
<th>8EWU</th>
</tr>
</thead>
<tbody>
<tr>
<td>resolution limits (Å)</td>
<td>50.0−1.45 (1.55−1.45)</td>
</tr>
<tr>
<td>number of independent reflections</td>
<td>121498 (20129)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>96.2 (87.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.8 (3.2)</td>
</tr>
<tr>
<td>avg I/avg I (t)</td>
<td>18.4 (3.2)</td>
</tr>
<tr>
<td>$R_{	ext{sym}}$ (%)</td>
<td>4.9 (33.8)</td>
</tr>
<tr>
<td>$R$-factor (overall)/no. of reflections</td>
<td>190/121498</td>
</tr>
<tr>
<td>$R$-factor (working)/no. of reflections</td>
<td>18.8/115400</td>
</tr>
<tr>
<td>$R$-factor (free)/no. of reflections</td>
<td>21.9/6098</td>
</tr>
<tr>
<td>number of protein atoms</td>
<td>5845</td>
</tr>
<tr>
<td>number of heteroatoms</td>
<td>860</td>
</tr>
<tr>
<td>average B values (Å²)</td>
<td>16.7</td>
</tr>
<tr>
<td>protein atoms (Å²)</td>
<td>11.0</td>
</tr>
<tr>
<td>ligand (Å²)</td>
<td>26.1</td>
</tr>
<tr>
<td>solvent (Å²)</td>
<td>0.012</td>
</tr>
<tr>
<td>bond lengths (Å)</td>
<td>1.75</td>
</tr>
<tr>
<td>planar groups (Å)</td>
<td>0.009</td>
</tr>
<tr>
<td>Ramachandran regions (%)</td>
<td>97.9</td>
</tr>
<tr>
<td>most favored</td>
<td>97.9</td>
</tr>
<tr>
<td>additionally allowed</td>
<td>2.1</td>
</tr>
<tr>
<td>generously allowed</td>
<td>0.0</td>
</tr>
</tbody>
</table>

$a R_{	ext{sym}} = (\sum|I| - \overline{I}/\sum I) \times 100$. $b$Statistics for the highest resolution bin. $c$-factor = (|$\sum F_o$| - |$\sum F_c$|)/|($\sum F_o^2$| + |$\sum F_c^2$|)/2 $\times 100$, where $F_o$ is the observed structure-factor amplitude and $F_c$ is the calculated structure-factor amplitude.

---

$^{27,28}$ Relevant X-ray data collection and refinement statistics are provided in Table 1.
dubbed HS:10A is expected to produce GDP-6-deoxy-D-altro-heptose (4), whereas enzyme HS:10B is expected to produce GDP-6-deoxy-L-galacto-heptose (6). In the structurally determined CPS from serotype HS:10, it is 6-deoxy-L-galacto-heptose (6) that is actually found within the CPS. Moreover, in the gene cluster for serotype HS:10, the enzyme will only catalyze the racemization of C3, yet in the structurally characterized product, C5 is epimerized relative to GDP-D-glycero-D-manno-heptose (1). A different situation is apparent in serotype HS:41 where the C4-reductase dubbed HS:41A is expected to produce GDP-6-deoxy-D-altro-heptose (4), whereas enzyme HS:41B is expected to produce GDP-6-deoxy-L-galacto-heptose (6). The chemically characterized 6-deoxyheptose from serotype HS:41 was found to be GDP-6-deoxy-D-altro-heptose (4), and thus, the product is formed by the catalytic activity of the C4-reductase dubbed HS:41A.

**Expression and Purification.** The 10 C4-reductases from serotypes HS:8, HS:10, HS:29, HS:41, and HS:63 were purified to homogeneity. However, the original plasmid constructions for the expression of the C4-reductases dubbed HS:8B and HS:10B led to proteins that were essentially insoluble. Closer inspection of the amino acid sequence alignment for the 25 C4-reductases (Figure S4) clearly shows that the C4-reductases dubbed HS:8B and HS:10B are postulated to initiate gene expression at an alternate upstream codon relative to the other 23 C4-reductases. To test whether the postulated 28-amino acid extension at the N-terminus of the protein was problematic, we resequenced the expression plasmids to exclude this region and initiated gene expression at an alternate upstream alignment for the 25 C4-reductases (Figure S4) clearly shows C4-reductase activity of the C4-reductase dubbed HS:41A.

**Product Specificities.** The 10 C4-reductases from serotypes HS:8, HS:10, HS:29, HS:41, and HS:63 were purified to homogeneity, and the product outcomes were ascertained using GDP-6-deoxy-4-keto-4-L-idoo-heptose (2) as the initial substrate at a concentration of 4.0 mM under conditions in which we added either a C3-epimerase or a C3/C5-epimerase to the reaction mixture in the presence of an appropriate reducing system. We conducted similar experiments in the absence of any additional epimerase to the reaction mixture. The product outcomes were determined by NMR spectroscopy, and the results of these experiments are presented in Table 2. The 1D and 2D ¹H NMR spectra of the isolated products are provided in Figures S5–S55.

For the A-type C4-reductases (HS:8A, HS:10A, HS:29, HS:41A, and HS:63A), no products were obtained in the absence of an added epimerase. However, in the presence of the appropriate epimerase, the expected products were produced in high yield with all five enzymes (Table 2). For the B-type C4-reductases (HS:8B, HS:10B, HS:29B, HS:41B, and HS:63B), the expected products were obtained in the presence or absence of an added epimerase, except for HS:8B where the product, GDP-6-deoxy-L-idoo-heptose (5), was only obtained in the presence of an added C3-epimerase.

### Table 2. Product Analysis for C4-Reductases Formed in the Presence of Either a C3 or C3/C5 Epimerase or the Absence of an Added Epimerase

<table>
<thead>
<tr>
<th>enzyme</th>
<th>predicted product</th>
<th>epimerase</th>
<th>observed product</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS:8A</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
</tr>
<tr>
<td>HS:8B</td>
<td>GDP-6-deoxy-L-idoo-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-idoo-heptose</td>
</tr>
<tr>
<td>HS:10A</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
</tr>
<tr>
<td>HS:10A (S242F)</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
<td>C3</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
</tr>
<tr>
<td>HS:10B</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
<tr>
<td>HS:15</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
<tr>
<td>HS:29A</td>
<td>GDP-6-deoxy-L-idoo-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-idoo-heptose</td>
</tr>
<tr>
<td>HS:29B</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
<tr>
<td>HS:41A</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-D-altro-heptose</td>
</tr>
<tr>
<td>HS:41B</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
<tr>
<td>HS:41B (C110S)</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
<tr>
<td>HS:42</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
<tr>
<td>HS:63A</td>
<td>GDP-6-deoxy-L-gluco-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-gluco-heptose</td>
</tr>
<tr>
<td>HS:63B</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
<td>none</td>
<td>GDP-6-deoxy-L-galacto-heptose</td>
</tr>
</tbody>
</table>

*Initial substrate concentration of 4.0 mM GDP-6-deoxy-4-keto-D-lyxo-heptose (2) at pH 7.5 and 22 °C in the presence or absence of a C3- or C3/C5-isomerase.*

**Epimerase Activities of C4-Reductases.** With the B-type C4-reductases from serotypes HS:10, HS:29, HS:41, and
HS:63, the enzymes must have the ability to catalyze the epimerization at both C3 and C5, since the stereochemistry at C3 and C5 of the product, GDP-6-deoxy-4-keto-d-altro-heptose (6), is opposite to that of the substrate, GDP-6-deoxy-4-keto-d-lyxo-heptose (2). This is clearly shown by the fact that HS:10B, HS:29B, HS:41B, and HS:63B can all catalyze the NADPH-dependent formation of GDP-6-deoxy-l-galacto-heptose (6) directly from compound 2 in the absence of any additional epimerase. The epimerase activity of the C4-reductase HS:10B was tested directly by the addition of 10 μM enzyme to a solution of 4.0 mM GDP-6-deoxy-4-keto-d-lyxo-heptose (2) in the presence of 150 μM NADP+, and the NMR spectrum was taken as a function of time. After 15 h, the formation of the C3, C5, and C3/C5 isomerized products is clearly shown for the hydrogen at C1 (Figure 5).

![Figure 5. Demonstration of the ability of the HS:10B C4-reductase to catalyze the isomerization of C3 and C5 of GDP-6-deoxy-4-keto-d-altro-heptose (2). The enzyme was incubated with compound 2 in the presence of NADPH at pH 7.5. (A) Control experiment showing a portion of the 1H NMR spectrum of compound 2 prior to the addition of an enzyme. The doublet at ~5.48 ppm is from the hydrogen attached to C1. (B) Portion of the 1H NMR spectrum of compound 2 after the addition of 4 μM HS:10B C4-reductase in the presence of NADPH. The unresolved doublet of doublets at ~5.38 ppm is for the C3-isomerized product, and the triplet at ~5.31 ppm is for the C5-isomerized product. The triplet at ~5.02 ppm is from the C3/C5-isomerized product. Additional details are provided in the text.](https://pubs.acs.org/doi/10.1021/acs.biochem.2c00633)

It has previously been shown that certain members of the SDR family of enzymes have the ability to catalyze the epimerization of the stereochemistry at C3 and/or C5 in addition to the oxidation/reduction at C4. For example, GDP-1-fucose synthase from E. coli catalyzes the NADPH-dependent formation of GDP-1-fucose from GDP-6-deoxy-4-keto-d-mannose via a reaction mechanism that requires the epimerization of C5 and C5 prior to the reduction of the carbonyl group at C4. The epimerization of C3 and C5 has been shown to require the acid/base functionalities of Cys109 and His179. Examination of the amino acid sequences of the C4-reductases from C. jejuni shows that all five members of Group-5 (HS:10B, HS:29B, HS:41B, HS:63B, and HS:42) in the SSN for the C4-reductases (Figure 4) have a conserved cysteine and histidine at residue positions 110 and 180, respectively (Figure S4).

To determine whether the cysteine/histidine dyad in the C4-reductase from HS:41B was required for the isomerization of C3 and C5, we mutated Cys110 to serine. When the C110S variant was tested with GDP-6-deoxy-4-keto-d-altro-heptose (2) in the presence of a C3/C5-epimerase, we identified GDP-6-deoxy-l-galacto-heptose (6) as the product in the presence of NADPH (Figure S37 and Table 2). This result demonstrates that the C4-reductase activity of this variant has not been perturbed (Table 2). However, in the absence of an added epimerase, no product was produced, showing that the cysteine residue was required for the isomerization of either C3 and/or C5 of the GDP-6-deoxy-4-keto-d-altro-heptose (2) substrate. Curiously, when we added the C3-isomerase, the C110S variant was able to weakly catalyze the formation of GDP-6-deoxy-d-altro-heptose (4) as illustrated in Figure S36. Therefore, this variant is able to slowly catalyze the reduction of the “wrong” substrate that differs in stereochemistry from the “normal” substrate for this enzyme at C5.

The lack of a cysteine at residue position 110 explains, in part, why the HS:8B C4-reductase is unable to catalyze product formation in the absence of an added epimerase (Table 2). However, it is curious to note that enzymes from Group-9 and Group-4 also possess a conserved cysteine and histidine at residue positions 110 and 180. However, the HS:15 C4-reductase is unable to catalyze the formation of GDP-6-deoxy-l-gulo-heptose (8) in the absence of an added epimerase (Table 2). In the C4-reductase from the HS:2 serotype (DdahC), the residues at positions 110 and 180 are tyrosine and histidine, respectively. Creuzenet’s group has mutated Tyr110 of this C4-reductase to cysteine but found no evidence for the ability of this variant to catalyze an epimerization of either C3 or C5 with heptose substrates. Therefore, there must be additional factors that control whether or not these SDR reductases are functionally able to catalyze the epimerization of C3 and/or C5 of potential substrates.

**Rates of Product Formation.** The rates of product formation for the 10 C4-reductases at an initial substrate concentration of 1.0 mM are presented in Table 3. In the presence of the appropriate epimerase, the A-type C4-reductases catalyze the formation of the expected product with rate constants that vary from 12 to 102 min⁻¹. The apparent slowness of the C4-reductase HS:10A was initially concerning (12 min⁻¹). However, we noticed that in the published amino acid sequence of this reductase, residue position 242 was a serine. However, in all of the other published sequences of the C4-reductases from C. jejuni (see Figure S4), residue position 242 is conserved as a phenylalanine. It is unclear whether or not this mutation was the result of a sequencing error or a natural variation found in that specific strain of C. jejuni. Nevertheless, we changed Ser242 to a phenylalanine and the resulting enzyme variant was found to be greater than 16-fold more active (200 min⁻¹).

**Poly-Guanine Tracts in the Genes for the A-Type C4-Reductases.** Poly-guanine (poly-G) tracts have been identified within the gene clusters used for the expression of the enzymes required for capsular polysaccharides in C. jejuni. These poly-G tracts have been shown to enable the bacterium to alter the expression of specific genes that can change the chemical nature of the outer polysaccharide via slipped-strand mispairing phase variation during DNA replication. The corresponding change in the subsequent coding sequence via the addition or deletion of guanine residues within the poly-G tracts of the DNA can result in premature truncation of the expressed protein via the
occurrence of a new stop codon.\textsuperscript{21} Interestingly, all of the A-type C4-reductases have a poly-G tract of 8–11 consecutive guanine residues near the 3′-end of the genes as illustrated in Figures S38–S42. These poly-G tracts are thus likely to be involved in the control of the catalytic properties of either the A-type or B-type C4-reductases or in the level of protein production. Such changes would change the specific 6-deoxy-heptose that was used to coat the external surface of the bacterial capsular polysaccharide, although the specific mechanism for this alteration is unclear to us. Nevertheless, it is highly probable that in the case of the HS:10 and HS:41 serotypes, only one of the two C4-reductases was functional in the specific strain of \textit{C. jejuni} that was used to establish the chemical structure of the CPS. In the case of the HS:10 serotype, it is clear that the HS:10B C4-reductase was functional, whereas in the HS:41 serotype, the HS:41A C4-reductase governed the final product outcome.

To address how the insertion (or deletion) of additional guanine residues within the 3′-ends of the genes for the A-type C4-reductases may have an effect on the expression/activity of these enzymes, we show the published sequences for the DNA that resides between the end of the A-type C4-reductases and the putative start of the B-type C4-reductases. These DNA sequences are shown in Figures S38–S42. For the A-type C4-reductase from serotype HS:8, there is a string of 10 guanine residues near the 3′-end of the gene (Figure S38). Shortly thereafter, there is a TTG codon (highlighted in yellow) that has been proposed (by others) as the putative start site for initiation of the HS:8B C4-reductase. A canonical ATG initiation site for the B-type C4-reductase is approximately 84 residues downstream from the TTG codon. Attempts to express soluble protein using the TTG start site has proven problematic, but production of soluble protein from the canonical AUG start site proceeded smoothly. It thus appears that the expression of protein with this “leader” peptide does not lead to the production of soluble protein in our hands.

In the DNA sequence for the HS:8 serotype, the stop codon (TAA, highlighted in red) for the HS:8A C4-reductase is shortly after the ATG codon used for the initiation of the HS:8B C4-reductase. The close localization of the stop codon for the HS:8A C4 reductase and the putative start codon for the HS:8B C4 reductase may enhance the expression of the latter via translational coupling in which the close proximity of the stop codon for the upstream gene to the start codon of the downstream gene may allow the ribosome to remain bound to the RNA long enough for the initiation of the expression of the downstream gene.\textsuperscript{21} It is interesting to note that this intervening DNA sequence has no stop codons in two separate reading frames over 84 bases. However, when one or two additional guanine residues are added to the poly-G tract, a new stop codon for the HS:8A C4-reductase is formed shortly thereafter (see Figure S38). For these two situations, the new stop codon for the HS:8A C4-reductase is now very close to the putative TTG start site for the HS:8B C4-reductase. Again, it is not so clear as to how this change in the DNA sequence will affect the initiation of expression for the HS:8B C4-reductase. Very similar situations are also observed for the HS:10, HS:29, and HS:63 serotypes. The insertion (or deletion) of additional guanine residues within the poly-G tracts changes the location of the stop codon for the A-type C4-reductase. The apparent exception is that for the HS:41 serotype where the addition of a single extra guanine results in the loss of a stop codon for the HS:41A C4-reductase.

The product specificity for the A- and B-type C4-reductases is different, and in two of the serotypes where the chemical structure of the CPS has been determined (HS:10 and HS:41), only one of the two possible 6-deoxy-heptoses has been identified. This result indicates that only one of the two C4-reductases was catalytically active. The poly-G tracts near the 3′-ends of the genes for the A-type C4-reductases are the likely determinants for the functional expression of the tandem genes via the change in the position of the stop codon for the gene of the A-type C4-reductase, but the actual mechanistic details have not been ascertained at this time.

\textbf{Structure of the C4-Reductase from Serotype HS:15.}\n
The crystals of the C4-reductase from serotype HS:15 used in this investigation belonged to the space group \textit{P1} with a dimer in the asymmetric unit. They were grown in the presence of NADPH and GDP. The model was refined at 1.45 Å resolution and to an overall \textit{R}\textsubscript{factor} of 19.0%. The backbone electron densities for both polypeptide chains were continuous from the N- to C-termini (352 amino acids per subunit). Likewise, the corresponding electron densities for the GDP and NADPH ligands were unambiguous. Shown in Figure 6 is a ribbon representation of the dimer. Each subunit folds into

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
enzyme & observed product & epimerase & observed rate (min\textsuperscript{-1}) \\
\hline
HS:8A & GDP-6-deoxy-\alpha-ido-heptose & C3 & 37 ± 1 \\
HS:8B & GDP-6-deoxy-\alpha-altro-heptose & C3 & 50 ± 4 \\
HS:10A & GDP-6-deoxy-\alpha-altro-heptose & C3 & 12.4 ± 0.1 \\
HS:10A S242F & GDP-6-deoxy-\alpha-altro-heptose & C3 & 201 ± 2 \\
HS:10B & GDP-6-deoxy-\alpha-galacto-heptose & none & 5.6 ± 0.1 \\
 & & C3 & 41 ± 1 \\
 & & C3/C5 & 56 ± 3 \\
HS:15 & GDP-6-deoxy-\alpha-galacto-heptose & C3/C5 & 16 ± 3 \\
HS:29A & GDP-6-deoxy-\alpha-altro-heptose & C3 & 102 ± 3 \\
HS:29B & GDP-6-deoxy-\alpha-galacto-heptose & none & 5.7 ± 0.4 \\
 & & C3 & 24 ± 1 \\
 & & C3/C5 & 25 ± 2 \\
HS:41A & GDP-6-deoxy-\alpha-altro-heptose & C3 & 76 ± 2 \\
HS:41B & GDP-6-deoxy-\alpha-galacto-heptose & none & 6.9 ± 0.1 \\
 & & C3 & 18.6 ± 0.2 \\
 & & C3/C5 & 20 ± 1 \\
HS:41B (C110S) & GDP-6-deoxy-\alpha-altro-heptose & C3 & <0.2 \\
 & GDP-6-deoxy-\alpha-galacto-heptose & C3/C5 & 2.1 ± 0.1 \\
HS:42 & GDP-6-deoxy-\alpha-galacto-heptose & none & 7.5 ± 0.1 \\
 & & C3 & 38 ± 1 \\
 & & C3/C5 & 58 ± 1 \\
HS:63A & GDP-6-deoxy-\alpha-gluco-heptose & C3/C5 & 20 ± 1 \\
HS:63B & GDP-6-deoxy-\alpha-galacto-heptose & none & 4.8 ± 0.1 \\
 & & C3 & 18 ± 1 \\
 & & C3/C5 & 26 ± 1 \\
\hline
\end{tabular}
\caption{Relative Rate Constants for C4-Reduction in the Presence of Different Epimerases\textsuperscript{a}}
\end{table}

\textsuperscript{a}Initial substrate concentration of 1.0 mM GDP-6-deoxy-4-keto-\alpha-D-
lyxo-heptose (2) at pH 7.5 and 25 °C in the presence or absence of a C3- or C3/C5-isomerase.

\textsuperscript{a}
two domains with the active sites wedged between the domains. The N-terminal domain, responsible for NADPH binding, consists of a six-stranded parallel β-sheet with α-helices flanking both sides. NADPH lies in an extended conformation across the C-terminal edge of the β-sheet with the nicotinamide ring projecting into the interface formed by the two domains. The C-terminal region, which anchors the GDP moiety, contains a β-hairpin motif and a two-stranded parallel β-sheet. A total of 13 α-helices surround these three β-sheet regions. The overall subunit architecture of the C4-reductase places it into the well-defined short chain dehydrogenase/reductase (SDR) superfamily.

A structural search against the Protein Data Bank utilizing SSM revealed that the overall fold of the C4-reductase from serotype HS:15 is similar to that of human GDP-\(\text{L-fucose}\) synthase with bound NADP and the product GDP-\(\text{L-fucose}\) (PDB code 4BL5, to be published). Indeed, the \(\alpha\)-carbons for the two models superimpose with a root-mean-square deviation of 1.2 Å. Provided in Figure 7 is the superposition of the ribbon drawings for these two enzymes. Other than the extended loop in the C4-reductase near Gln198 to Asp223, the overall three-dimensional models for the two enzymes are strikingly similar.

As noted below, the GDP-\(\text{L-fucose}\) synthases catalyze both C3 and C5 epimerizations as well as 4-keto moiety reductions and these reactions require the presence of a cysteine/histidine dyad. In addition to containing the canonical Tyr-X-X-Lys pair characteristic of SDR superfamily members, the C4-reductase from serotype HS:15 also contains the cysteine/histidine pair, which is in the proper position for an epimerization as indicated in Figure 8. Intriguingly, this enzyme functions only as a C4-reductase. It is thus not possible to predict, based solely on the presence of this cysteine/histidine couple, whether a C4-reductase will have epimerization as well as reduction activities.

Predicted Structural Changes to the A-Type and B-Type C4-Reductases. It is clear that the A-type and B-type C4-reductases have the potential to possess C-terminal and N-terminal polypeptide extensions to the prototypical C4-reductases described previously. To further help understand the structural consequences for these putative extensions, we employed AlphaFold to predict the structural features of these additions. In Figure 9a is the predicted structure of the HS:29A C4-reductase that includes the C-terminal extension. The monomer is aligned with the previously determined structure of the C4-reductase (Cj1428) from serotype HS:2 (PDB ID: 7M13). It appears that the C-terminal extension forms a predominantly \(\alpha\)-helical extension that is not predicted to interfere with dimer formation or the active site residues. Shown in Figure 9b is the predicted structure for the N-terminal extension for the HS:8B C4-reductase. The N-terminal extension is also predicted to be predominantly \(\alpha\)-
regulate catalytic activity, or they may be used as recognition actions. N-terminal or C-terminal extension failed because of limited extensions from the other C4-reductases examined in this C5 of GDP-6-deoxy-4-keto-C4-reductases was able to catalyze the epimerization of C3 and ultimate step during the biosynthesis of stereochemical HS:63 serotypes of polysaccharides in the HS:8, HS:10, HS:29, HS:41, and HS:63 (A0A0S2CGZ7 and A0A0S2CGF0).

CONCLUSIONS

We identified coding regions for two different C4-reductases in the gene clusters for the biosynthesis of the capsular polysaccharides in the HS:8, HS:10, HS:29, HS:41, and HS:63 serotypes of C. jejuni. These C4-reductases catalyze the ultimate step during the biosynthesis of stereochemical variations of 6-deoxy-heptoses used for the CPS. The 10 C4-reductases were purified to homogeneity, and their substrate and product specificities were determined. We demonstrated that in serotypes HS:10, HS:29, HS:41, and HS:63, one of the C4-reductases was able to catalyze the epimerization of C3 and C5 of GDP-6-deoxy-4-keto-β-D-lyxo-heptose (2) prior to the NADPH-dependent reduction of C4. A poly-G tract of 8–11 consecutive guanine residues is found near the 3’-end of the coding region for the first C4-reductase. Expansion or contraction of this poly-G tract may lead to changes in the expression levels of the two C4-reductases and subsequent changes in the carbohydrate structure of the capsular polysaccharide. These changes may enable C. jejuni to further evade the host immune system. The three-dimensional structure of the C4-reductase from serotype HS:15 was determined to high resolution and demonstrated that while it does not catalyze an epimerization reaction, nonetheless, it contains the required cysteine/histidine dyad.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.2c00633.

Amino acid sequences of the purified proteins; enzyme modifications; NMR spectra of carbohydrate substrates and products (PDF)

Accession Codes

C4-reductases from serotypes HS:8 (F2X762 and F2X761), HS:10 (F2X783 and F2X782), HS:29 (A0A0U3C2A7 and A0A03AL93), HS:41 (Q5M6T8 and Q5M6T9), and HS:63 (A0A0S2CGZ7 and A0A0S2CGF0).

AUTHOR INFORMATION

Corresponding Authors

Hazel M. Holden — Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States; orcid.org/0000-0001-6214-3638; Email: hazel_holden@biochem.wisc.edu

Frank M. Raushel — Department of Chemistry and Biophysics, Texas A&M University, College Station, Texas 77843, United States; orcid.org/0000-0002-5918-3089; Email: raushel@tamu.edu

Authors

Dao Feng Xiang — Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

Manas K. Ghosh — Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

Alexander S. Riegert — Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas 77843, United States

James B. Thoden — Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.2c00633

Funding

This work was funded by the National Institutes of Health (GM 139428 and GM122825 to F.M.R. and GM 134643 to H.M.H.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Professors Jennifer Hermann (TAMU), Michael Manson (TAMU), and Reinhard Sterner (Regensburg University) for many helpful discussions.

REFERENCES


