Cytidine 5′-Monophosphate (CMP)-Induced Structural Changes in a Multifunctional Sialyltransferase from Pasteurella multocida1,‡

Lisheng Ni,§ Mingchi Sun,§ Hai Yu,§ Harshal Chokhawala,§ Xi Chen,*,§ and Andrew J. Fisher*,§†

Department of Chemistry and the Section of Molecular and Cellular Biology, University of California, One Shields Avenue, Davis, California 95616

Received November 23, 2005; Revised Manuscript Received December 19, 2005

ABSTRACT: Sialyltransferases catalyze reactions that transfer a sialic acid from CMP—sialic acid to an acceptor (a structure terminated with galactose, N-acetylgalactosamine, or sialic acid). They are key enzymes that catalyze the synthesis of sialic acid-containing oligosaccharides, polysaccharides, and glycoconjugates that play pivotal roles in many critical physiological and pathological processes. The structures of a truncated multifunctional Pasteurella multocida sialyltransferase (Δ24PmST1), in the absence and presence of CMP, have been determined by X-ray crystallography at 1.65 and 2.0 Å resolutions, respectively. The Δ24PmST1 exists as a monomer in solution and in crystals. Different from the reported crystal structure of a bifunctional sialyltransferase CstII that has only one Rossmann domain, the overall structure of the Δ24PmST1 consists of two separate Rossmann nucleotide-binding domains. The Δ24PmST1 structure, thus, represents the first sialyltransferase structure that belongs to the glycosyltransferase-B (GT-B) structural group. Unlike all other known GT-B structures, however, there is no C-terminal extension that interacts with the N-terminal domain in the Δ24PmST1 structure. The CMP binding site is located in the deep cleft between the two Rossmann domains. Nevertheless, the CMP only forms interactions with residues in the C-terminal domain. The binding of CMP to the protein causes a large closure movement of the N-terminal Rossmann domain toward the C-terminal nucleotide-binding domain. Ser 143 of the N-terminal domain moves up to hydrogen-bond to Tyr 388 of the C-terminal domain. Both Ser 143 and Tyr 388 form hydrogen bonds to a water molecule, which in turn hydrogen-bonds to the terminal phosphate oxygen of CMP. These interactions may trigger the closure between the two domains. Additionally, a short helix near the active site seen in the apo structure becomes disordered upon binding to CMP. This helix may swing down upon binding to donor CMP—sialic acid to form the binding pocket for an acceptor.

Sialic acids (2-keto-3-deoxy-nonulosonic acids) are negatively charged α-keto acids with a nine-carbon backbone. They are commonly found as terminal carbohydrate residues on surface glycoconjugates (glycoproteins and glycolipids) of higher animals. As the terminal carbohydrate residue, sialic acid is one of the first molecules encountered in cellular interactions and has been found to play important roles in cellular recognition and communication (1, 2). Cell surface sialic acids have also been found in relatively few microorganisms, mainly pathogenic bacteria, and their presence is often associated with virulence (1–4). For example, sialic acid-containing capsular polysaccharides of Neisseria meningitidis, Escherichia coli K1, and group B Streptococci are important virulence factors (5, 6). Sialic acids on lipoooligosaccharides (LOS) of several mucosal pathogens, including N. meningitidis, Neisseria gonorrhoeae, Haemophilus ducreyi, and Haemophilus influenzae strains (7–15) are considered important LOS components for molecular mimicry to allow them to evade the host’s immune system (8, 15–17) or as modulators of microbial interactions with host cells (1, 2, 14, 18).

Sialyltransferases are key enzymes for the biosynthesis of sialic acid-containing structures. They catalyze reactions that transfer a sialic acid from its activated sugar nucleotide donor, cytidine 5′-monophosphate sialic acid (CMP-sialic acid), to its acceptor, usually a structure terminated with a galactose, an N-acetylgalactosamine, or another sialic acid. Sialyltransferases have been isolated and cloned from both

---

1 This work was supported by start-up funds to X.C. from the Regents of the University of California. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

2 Protein coordinates have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb) as entries 2EXO and 2EX1 for the ligand-free apo and the CMP binary complex, respectively.

3 Corresponding authors. A.J.F.: tel, (530) 754-6180; fax, (530) 752-8995; e-mail, fisher@chem.ucdavis.edu. X.C.: tel, (530) 754-6037; fax, (530) 752-8995; e-mail, chen@chem.ucdavis.edu.

4 Department of Chemistry, University of California.

5 Section of Molecular and Cellular Biology, University of California.

6 Abbreviations: ASU, asymmetric unit, CMP, cytidine 5′-monophosphate; DTT, dithiothreitol; GlcNAc, N-acetylgalactosamine; GPTGF, glycogen phosphorylase/glycosyltransferase superfamily; GT, glycosyltransferase; IPTG, isopropyl-1-thio-β-D-galactopyranoside; LOS, lipoooligosaccharide; MAD, multiple wavelength anomalous dispersion; Neu5Ac, N-acetylmuramic acid; rmm, root-mean-square deviation; SeMet, Seleniummethionine; SSRL, Stanford Synchrotron Radiation Laboratory; UDP, uridine 5′-diphosphate.
with an optimal activity at pH 4.5–optimal activity at pH 7.5. The overall structure of this protein to help understand its mechanisms of catalysis. Therefore, it is of extreme interest to elucidate the molecular structure of this protein to help understand its mechanisms of catalysis.

We report herein the crystal structures of the multifunctional sialyltransferase Δ24PmST1 (Scheme 1) (32). It is (a) an α2,3-sialyltransferase with an optimal activity at pH 7.5–9.0, (b) an α2,6-sialyltransferase with an optimal activity at pH 4.5–6.0, (c) an α2,3-sialidase with an optimal pH of 5.0–5.5, and (d) an α2,3-trans-sialidase with an optimal pH of 5.5–6.5. It is unusual that a single enzyme possesses so many different functions. Therefore, it is of extreme interest to elucidate the molecular structure of this protein to help understand its mechanisms of catalysis.

We report herein the crystal structures of the multifunctional sialyltransferase Δ24PmST1 in the absence and the presence of CMP at 1.65 and 2.0 Å resolutions, respectively. The overall structure of Δ24PmST1 consists of two separate Rossmann nucleotide-binding domains that forms a deep nucleotide-binding cleft between the two domains. Comparison of two structures shows a large conformational change of the enzyme upon binding to CMP. This study elucidates a new class of sialyltransferase structure that differs greatly from that of CstII.

**EXPERIMENTAL PROCEDURES**

**Expression and Purification of Pasteurella multocida Multifunctional Sialyltransferase Δ24PmST1.** These procedures were performed as previously reported (32). Briefly, the sialyltransferase gene was cloned excluding the predicted N-terminal membrane anchor comprising the first 24 residues. *E. coli* BL21 (DE3) harboring the recombinant plasmid in a pET23a(+) vector was grown in LB-rich medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) containing 100 μg/mL ampicillin until the OD_{600} reached 0.8–1.0. The expression of the protein was induced by adding 0.1 mM IPTG (isopropyl-1-thio-β-D-galactopyranoside) followed by incubation at 37 °C for 3 h with vigorous shaking at 250 rpm in a C25KC incubator shaker (New Brunswick Scientific, Edison, NJ). The bacterial cells were harvested by centrifugation at 4 °C in a Sorvall Legend RT centrifuge with a swinging bucket rotor at 3690g for 60 min. The cell pellet was resuspended in 20 mL/L cell culture in lysis buffer (pH 8.0, 100 mM Tris-HCl containing 0.1% Triton X-100). Lysozyme (1 mg/L culture) and DNaseI (50 μg/L culture) were then added to the cell suspension. After the mixture was incubated at 37 °C for 50 min with vigorous shaking (250 rpm), the cell lysate was separated from inclusion bodies and other cellular debris by centrifugation (Sorvall RC-5B centrifuge with a S5-34 rotor) at 7000g for 30 min. The enzyme was purified directly from the lysate using a Ni^{2+}–NTA affinity column. All purification procedures were performed on ice or at 4 °C. After loading the cell lysate, the Ni^{2+} column was washed with 6 column vol of binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.5), followed by 8 column vol of washing buffer (20 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.5). The protein was eluted with 8 column vol of elute buffer (200 mM imidazole, 0.5 M NaCl, and 20 mM Tris-HCl, pH 7.5) and detected by a UV/Vis spectrometer at 280 nm using a Smart-Spec 3000 Spectrophotometer (Bio-Rad, Hercules, CA). Fractions containing the purified enzyme were collected and stored at 4 °C.

**Preparation of Selenomethionyl Δ24PmST1.** Selenomethionine-containing *P. multocida* (Pm) sialyltransferase was prepared as reported with modifications (33). Briefly, recombinant *E. coli* BL21 (DE3) were grown in modified M9 minimal media (6.78 g/L Na_{2}HPO_{4}, 3.00 g/L KH_{2}PO_{4}, 0.50 g/L NaCl, 1.24 g/L (NH_{4})_{2}SO_{4}, 2 mM MgSO_{4}, 0.1 mM CaCl_{2}, 4 g/L glucose, trace amount of ZnSO_{4}, MnSO_{4}, H_{3}BO_{3}, NaMoO_{4}, CoCl_{2}, KI, CuSO_{4}, FeSO_{4}, and H_{2}SO_{4}) (34) containing 400 μg/mL ampicillin until the OD_{600}nm reached 0.6–0.8. A sterile mixture of amino acids (100 mg/L of L-lysine, L-phenylalanine, and L-threonine; 50 mg/L of L-isoleucine, L-leucine, and L-valine; and 60 mg/L of L-selenomethionine) was then added. The expression of the protein was induced 15 min later by addition of 0.1 mM IPTG and incubation at 37 °C for 6 h with vigorous shaking at 250 rpm. The procedures for lysate preparation and protein purification were the same as described above except that 0.2 mM of dithiothreitol (DTT) was added to the lysate and all the buffers used for the purification.

**Crystallization of Δ24PmST1.** Purified enzyme was dialyzed in Tris-HCl buffer (20 mM, pH 7.5) and concentrated to 10 mg/mL for crystallization. Native protein was crystallized by hanging drop vapor diffusion using 3-L drops of reservoir buffer (100 mM HEPES, pH 7.5, 100 mM NaCl, and 0.4% Triton X-100). The crystals grew to full size of 0.3 mm × 0.3 mm × 0.05 mm after 2 weeks. The binary CMP crystals were grown under identical conditions except for addition of 2 mM CMP—N-acetylneuraminic acid (CMP—Neu5Ac). Structure determination indicated that the CMP—Neu5Ac was hydrolyzed and the N-acetylneuraminic acid was removed, resulting in only CMP binding to the enzyme (see below). The SeMet-derivatized
protein was crystallized under identical conditions as for the native protein. All crystals were transferred to Paratone-N and frozen in a stream of nitrogen to −173 °C for data collection.

**Data Collection and Phase Determination.** X-ray data from all three crystals (Native, SeMet, and CMP-bound) were collected on beam line 9-2 at Stanford Synchrotron Radiation Laboratory (SSRL). Although all three crystals were grown under identical conditions, the three crystallized in space group *P*2₁ but with different cell parameters and crystal packing (Table 1). The data were indexed and integrated with DENZO and were reduced with SCALEPACK (35). Complete data sets were collected to 1.65, 1.8, and 2.0 Å resolution for the native, SeMet, and CMP complex, respectively. For the native crystals, the Matthews coefficient *V*ₘₐₜₜ = 2.1 Å³/Da and a solvent content of ~41%, while the CMP complex also had one monomer per ASU (*V*ₘₐₜₜ = 2.7 Å³/Da, solvent content ~55%).

Multiwavelength anomalous dispersion (MAD) data were collected on beamline 9-2 at SSRL. Excellent 1.8 Å resolution data sets were collected at three wavelengths corresponding to the Selenium absorption edge (peak and inflection point) as well as a high-energy remote wavelength (Table 1). The three data sets were input into the program SOLVE (37), which was able to find 6 of the 7 selenium sites. The overall Figure of Merit (FOM) from SOLVE was 0.66. The phases were improved by solvent flattening using the program RESOLVE (38, 39), which improved the FOM to 0.74 for all data to 1.8 Å resolution. The SeMet data with phases were then input into the automated protein model-building program ARP/wARP (40), which was able to build the entire protein structure with the exception of residues 98–107 and 145–154, with a final *R*-factor of 20.2%. This 1.8 Å resolution structure was used to solve the structure of native sialyltransferase using the 1.65 Å resolution data by molecular replacement using the program AMoRe (41). The SeMet structure was also used as a phasing model to solve the binary CMP structure by molecular replacement. However, initial molecular replacement attempts were not successful using the complete enzyme structure. A solution, however, was achieved by breaking the apo-enzyme structure into two separate search domains, which corresponded to the two separate Rossmann domains. The solution revealed a significant closure between the two domains upon binding to CMP (see below).

---

**Table 1: Data Collection, Phasing, and Refinement Statistics**

<table>
<thead>
<tr>
<th>x-ray source</th>
<th>Se peak</th>
<th>Se remote</th>
<th>Se inflection</th>
<th>native</th>
<th>CMP complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>wavelength</td>
<td>SSRL</td>
<td>SSRL</td>
<td>SSRL</td>
<td>SSRL</td>
<td>SSRL</td>
</tr>
<tr>
<td>resolution</td>
<td>1.80 (1.86–1.80) Å</td>
<td>1.80 (1.86–1.80) Å</td>
<td>1.80 (1.86–1.80) Å</td>
<td>1.65 (1.69–1.65) Å</td>
<td>2.00 (2.05–2.00) Å</td>
</tr>
<tr>
<td>cell parameters</td>
<td>P₂₁</td>
<td>P₂₁</td>
<td>P₂₁</td>
<td>P₂₁</td>
<td>P₂₁</td>
</tr>
<tr>
<td>α</td>
<td>52.72</td>
<td>63.72</td>
<td>61.25</td>
<td>61.25</td>
<td>61.25</td>
</tr>
<tr>
<td>β</td>
<td>61.54</td>
<td>61.28</td>
<td>64.57</td>
<td>64.57</td>
<td>64.57</td>
</tr>
<tr>
<td>γ</td>
<td>65.05</td>
<td>96.65</td>
<td>64.84</td>
<td>64.84</td>
<td>64.84</td>
</tr>
<tr>
<td>β</td>
<td>111.79°</td>
<td>101.56°</td>
<td>98.62°</td>
<td>98.62°</td>
<td>98.62°</td>
</tr>
<tr>
<td>monomers/ASU</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Vₘₜₜ (Å³/Da); % solvent</td>
<td>2.1; 41%</td>
<td>2.0; 38%</td>
<td>2.7; 55%</td>
<td>2.7; 55%</td>
<td>2.7; 55%</td>
</tr>
<tr>
<td>no. of reflections</td>
<td>241355</td>
<td>253767</td>
<td>251778</td>
<td>280823</td>
<td>126172</td>
</tr>
<tr>
<td>no. unique</td>
<td>35235</td>
<td>35320</td>
<td>35728</td>
<td>86273</td>
<td>33403</td>
</tr>
<tr>
<td>Rmerge(%)</td>
<td>7.2 (20.8)</td>
<td>6.6 (20.2)</td>
<td>6.4 (21.3)</td>
<td>5.2 (30.8)</td>
<td>4.5 (17.0)</td>
</tr>
<tr>
<td>mean(I/σ(I))</td>
<td>24.9 (7.4)</td>
<td>23.5 (7.3)</td>
<td>23.4 (7.2)</td>
<td>15.1 (2.5)</td>
<td>17.8 (5.4)</td>
</tr>
<tr>
<td>completeness (%)</td>
<td>99.6 (99.9)</td>
<td>99.6 (100)</td>
<td>99.5 (99.3)</td>
<td>98.1 (82.1)</td>
<td>98.3 (97.7)</td>
</tr>
<tr>
<td>no. of Se sites</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>figure of merit (solve/resolve)</td>
<td>0.66/0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Refinement Statistics**

| resolution (Å) | 30–1.65 | 64.15–2.00 |
| no. reflections (F ≥ 0) | 86273 | 33403 |
| R-factor(%) | 19.3 | 19.2 |
| R-free(%) | 21.9 | 22.1 |
| rms bond length | 0.005 | 0.006 |
| rms bond angles | 1.22 | 1.24 |

**Ramachandran Plot Statistics**

| residues (%) | 710 | 343 |
| most favorable region (%) | 90.6 | 91.5 |
| allowed region (%) | 9.0 | 7.3 |
| generously allowed region (%) | 0.3 | 0.9 |
| disallowed (%) | 0.1 | 0.3 |

**Asymmetric Unit Content**

| nonhydrogen protein atoms | 6365 | 3110 |
| water | 841 | 398 |
| CMPS (atoms) | 1 (21) |
Model Building and Refinement. Model building was carried out with the molecular graphics program O (42). Energy minimization-coupled coordinate and B-factor refinement were carried out with the program CNS (43) using 95% of the data to the respective maximum resolution. During each stage of refinement, the R-free and R-factor decreased. Noncrystallographic restraints for the native ligand-free structure were released during later stages of refinement resulting in a lower R-free. The quality of the model was checked with the program PROCHECK (44, 45), and results are summarized in Table 1. Only Ala 219 falls in the disallowed region of the Ramachandran plot in all three models (two apo molecules in ASU and one binary structure). The electron density map clearly defines the same conformation in all three models ($\rho \approx 45^\circ$, $\psi \approx 120^\circ$). The reason for the unfavorable main-chain angles is that Ala 219 falls in a tight turn between $\alpha$6b and $\beta$7, which then proceeds into the C-terminal domain. The native and the CPM-bound atomic coordinates have been deposited in the protein data bank, accession numbers 2EX0 and 2EX1, respectively.

RESULTS AND DISCUSSION

Overall Native Structure. The sialyltransferase from P. multocida is likely anchored into the membrane by a predicted membrane-spanning domain consisting of an N-terminal 24 residue stretch of basic and hydrophobic residues (32). Other sialyltransferases are also predicted to be anchored to a membrane, through either an N-terminal or a C-terminal tail (31, 46, 47). A soluble construct was made by deleting the first 24 amino acid residues and replacing Ser 25 with the Met start codon ($\Delta$24PmST1), which was expressed and crystallized. The monomer is made up of two separate nonidentical $\alpha/\beta/\alpha$ domains whose structures and topologies are similar to the Rossmann nucleotide-binding domain (48). There is a deep cleft that forms between the two domains, which comprises the CMP binding site (see below). Thus, the PmST1 structure belongs to the glycosyltransferases B (GT-B) structural group (46, 49–53), which has more diversity in its substrates and products compared to the glycosyltransferases A family. The PmST1 structure, therefore, differs from the only other sialyltransferase structure reported to date, the CstII sialyltransferase from C. jejuni (31). The tetrameric CstII structure belongs to glycosyltransferases-A (GT-A) or GT-A-like structural group (46, 49–54). In the GT-A structures, each monomer consists of a single Rossmann domain and a smaller lid-like structure that folds over the active site (31). However, CstII lacks the DXD motif found in GT-A structures, which is known to coordinate divalent cations involved in binding to the sugar–nucleotide diphosphate moiety (31, 55, 56).

A ribbon diagram of the native ligand-free $\Delta$24PmST1 structure is shown in Figure 1. The N-terminal Rossmann-like domain (residues 25–246) consists of a central seven-stranded parallel $\beta$-sheet ($\beta$3, $\beta$2, $\beta$1, $\beta$4, $\beta$5, $\beta$6, and $\beta$7) flanked by four $\alpha$-helices on the backside (as viewed in Figure 1, $\alpha$1, $\alpha$2, $\alpha$7a, and $\alpha$7b) and seven $\alpha$-helices on the front side ($\alpha$3, $\alpha$4, $\alpha$5a, $\alpha$5b, $\alpha$5c, $\alpha$6a, and $\alpha$6b). The smaller C-terminal domain (residues 247–416) is also a $\alpha/\beta/\alpha$ fold with a central six-stranded parallel $\beta$-sheet ($\beta$10, $\beta$9, $\beta$8, $\beta$11, $\beta$12, and $\beta$13) sandwiched by two $\alpha$-helices on one side (top in Figure 1, $\alpha$8 and $\alpha$9) and five $\alpha$-helices on the other side ($\alpha$7c, $\alpha$10, $\alpha$11, $\alpha$12a, and $\alpha$12b). This domain also has a topology of the Rossmann nucleotide-binding domain. The genomic sequence of wild-type PmST1 ends at Leu 412, but the first four residues of the His-tag linker (413–416) of the recombinant $\Delta$24PmST1 are visible in the electron density map and were modeled in extended conformation over the top of the structure opposite to the N-terminal domain.

The quaternary structure of $\Delta$24PmST1 consists of a monomer with overall dimensions of $\sim$73 Å $\times$ 54 Å $\times$ 45
Even though the native ligand-free enzyme crystallized with two monomers in the asymmetric unit, inspection of crystal packing revealed there is little interaction between the two monomers or with other symmetry-related molecules. Size exclusion chromatography also suggested a monomeric arrangement in solution (data not shown). There is very little structural difference between the two monomers in the crystallographic asymmetric unit. The two monomers superimpose with a root-mean-squared deviation (rmsd) of 0.75 Å for 390 equivalent α-carbons. The largest structural difference is in the loop between α5b and α5c. Residues Thr 174—Lys 178 in the B monomer are pulled away from the N-terminal core slightly to make a lattice contact with a symmetry-related B monomer. The α-carbon of Gly 175 shifts 5.0 Å between the two monomers. Superposition of the individual domains independently resulted in a slightly lower rmsd of 0.64 Å for both respective domains, suggesting a minor difference in the orientation angle between the two domains and mobility between the two domains (see CMP binding below).

**CMP Binary Structure.** To understand the structural requirements for donor substrate binding and obtain clues for catalytic activity, Δ24PmST1 was cocrystallized with the sialyltransferase donor CMP—Neu5Ac. Inspection of the resultant electron density map, however, only revealed the CMP moiety with several well-ordered water molecules binding in the deep cleft between the two nucleotide-binding domains. It is likely that during the time required for crystallization, the enzyme hydrolyzed off the Neu5Ac moiety from CMP—Neu5Ac and only CMP was left bound to the enzyme. This phenomenon is very similar to that reported for CstII, for which crystallization of the CstII sialyltransferase in the presence of CMP—Neu5Ac also resulted in only observing CMP in the active site (31).

The initial CMP structure could not be solved by molecular replacement using the SeMet ligand-free structure as a search model. Nevertheless, a molecular replacement solution was found by dividing the ligand-free structure into two separate search models, each corresponding to an individual Rossmann domain. Indeed, inspection of the solution revealed a significant closure of the two domains over the CMP binding site in the deep cleft (Figure 2A) (see below), explaining the failure of the single-model molecular replacement search. In the final refined structure, residues Met 25—Lys 375 and Asn 385—Leu 412 are well-ordered in the electron density map. The stretch of nine disordered residues, Gln 376—Leu 384, corresponding to helix α12a, could be involved in acceptor binding (see below).

CMP makes extensive hydrogen bonds with the C-terminal nucleotide-binding domain of the enzyme, and no contacts are made with the N-terminal domain. A total of 12 hydrogen bonds develop between the enzyme C-terminal domain and atoms from the base, ribose, and phosphate of CMP (Figures 2B,C). Additionally, three ordered water molecules are in hydrogen-bonding distance to each of the terminal phosphate oxygens (Figures 2B,C). The specificity for the cytosine base comes from two hydrogen bonds between N4 of the base and the main-chain carbonyl oxygens of Gly 266 and Lys 309. In comparison, a uracil (or thymine) base contains a deprotonated carbonyl O4, which is incapable of serving as a hydrogen donor to the main-chain carbonyl oxygens. Additionally, Nζ of Lys 309 donates a hydrogen bond to unprotonated N3 of cytosine, while in uracil (or thymine), N3 is protonated and would be unable to accept this hydrogen bond. The O2 in the cytosine ring accepts hydrogen bonds from Nζ of Lys 309 and main-chain N of Phe 337. There is insufficient space to accommodate a purine base. Additionally, unlike that in the CstII sialyltransferase structure, there is no aromatic stacking of the cytosine ring with any amino acid residues in Δ24PmST1.

Interactions between Δ24PmST1 and the ribose—phosphate moiety of CMP display some resemblance to UDP binding in UDP—GlcNac 2-epimerase (57), its closest structural homologue (see below). In Δ24PmST1, the carbonyl group of Glu 338 makes a parallel bidentate hydrogen-bond pair with both hydroxyl groups (2’ and 3’) of the ribose in CMP, reminiscent of Glu 296 in UDP—GlcNac 2-epimerase. In Δ24PmST1, the ribose ring is in a slightly C2’-exo conformation, but almost planar, while in UDP—GlcNac 2-epimerase, the sugar pucker is more pronounced and in the C2’-endo-C3’-exo conformation. The phosphate moiety of CMP is positioned at the positive dipole moment at the N-terminus of helix α11, similar to the position of the β-phosphate of UDP in UDP—GlcNac 2-epimerase, which lies off the N-terminus of the topologically same helix. In both enzymes, the phosphate makes extensive interactions with both protein and solvent. In Δ24PmST1, each terminal phosphate oxygen hydrogen-bonds to both a solvent molecule and a hydrophilic side chain of the enzyme. CMP phosphate oxygen O1P hydrogen-bonds to Ne2 of His 311 and Wat 29. O2P hydrogen-bonds to both Oy and N from Ser 356 and Wat 1. And finally, O3P hydrogen-bonds to Oy of Ser 355 and Wat 11 (Figure 2B,C). His 311 and Ser 355 are analogous to His 213 and Ser 290 in UDP—GlcNac 2-epimerase, which make similar hydrogen bonds to the UDP β-phosphate. These residues, His 311, Ser 355, and Ser 356 in Δ24PmST1, could play catalytic roles in one or more of the measured activities of this enzyme. In particular, His 311, which stabilizes the phosphate, may serve as a general acid and protonate the departing phosphate oxygen of CMP. This histidine is comparable to His 213 in UDP—GlcNac 2-epimerase, which hydrogen-bonds to a β-phosphate oxygen of UDP and is hypothesized to serve as a general acid (57).

**Apo and Binary Structure Comparison.** Comparison of the ligand-free and the binary CMP—enzyme structures reveals a large N-terminal domain movement upon the enzyme binding to CMP. However, it is possible that the domain shift occurred during or after CMP—Neu5Ac binding or hydrolysis. This domain closure is a result of substrate or CMP binding and not of crystal packing, because comparison of the apo SeMet crystal structure to the apo native enzyme (natural amino acids) structure revealed little structural differences and an rmsd of only 0.486 and 0.345 Å (A and B subunits, respectively) between the two apo forms, which pack in diverse crystal arrangements.

Since the CMP moiety only contacts the C-terminal domain, the C-terminal domain was thus used as a fixed point for superposition and comparison of the CMP-bound and the native apo structures. When the two structures are overlaid at their C-terminal domain, there is a substantial rotation of the N-terminal domain upon binding CMP (Figure 3A). The N-terminal domain rotates ~23° upward toward the CMP binding site to close up the cleft between the two Rossman domains. This results in a movement of >18 Å.
Figure 2: Binary CMP-bound Δ24PmST1 structure. (A) Stereo-ribbon diagram revealing the CMP-bound (stick model with white-colored carbons) in the deep cleft between the two Rossmann domains. Coloring of the protein is the same as in Figure 1. Residues 375 and 385 mark the region of helix α12a, which becomes disordered upon binding CMP. (B) Close-up stereoview of the CMP binding site. The CMP is drawn as a stick model with white-colored carbon atoms (nitrogen, oxygen, and phosphorus atoms are colored blue, red, and orange, respectively). The amino acid side chains of residues that interact with CMP are also drawn as a stick model with tan-colored carbon atoms. Three water molecules that hydrogen-bond to terminal phosphate oxygens in CMP are represented as small red spheres. Many other well-ordered solvent molecules in the active site are not shown for clarity. Potential hydrogen bonds are drawn as green dashes. Electron density corresponding to a simulated-annealing omit map is drawn in gray mesh around the CMP moiety contoured at 5σ. (C) Flatten two-dimensional representation showing the interactions between CMP and protein or solvent atoms. Covalent bonds are drawn as solid purple lines for CMP and tan lines for protein residues. Hydrogen bonds are drawn as green dashed lines with numbers corresponding to the distances in angstroms. van der Waals contacts are represented by hashed red lines around an atom, which point to the direction of the corresponding atom (also with red hashed lines) with which it interacts. Panel C was generated with the program LIGPLOT (68).
for some residues at the distal end of the N-terminal domain (Figure 3A). The pivot point is Asp 249 in the loop linking the two domains.

Comparison of the individual N- or C-terminal domains reveals that significant structural changes have also occurred within the C-terminal domain upon binding CMP, but little difference is seen in the N-terminal domain. The rmsd values for overlaying the C-terminal domains between the CMP-bound and the apo structures are 2.41 and 2.48 Å for A and B apo subunits, respectively (156 equivalent α-carbons). In contrast, the rmsd values for the superposition of the N-terminal domains alone are only 0.55 and 0.72 Å for the A and B apo subunits, respectively (223 equivalent α-carbons).

The significant differences observed when comparing the two C-terminal domain structures center around helix R₁₂a in the C-terminal domain. This helix, which is well-defined in the ligand-free structure, becomes completely disordered in the CMP-bound structure, indicating its mobility in the latter (Figure 3B). The conformational flexibility of this helix suggests that it may swing down and interact with sugar substrates. The R₁₂a helix contains many hydrophilic residues (including three lysines, an aspartate, a glutamate, a serine, an asparagine, and a glutamine), which may be involved in binding to the sialic acid residue in donor CMP–sialic acid and/or the acceptor sugar and possibly contribute residues for catalysis. This is reminiscent of the CstII sialyltransferase structures in which a small loop was disordered in the CMP complex but became ordered upon binding to a CMP–sialic acid analogue (31).

In addition to the disorder of helix R₁₂a, helix R₁₂b rotates up away from the N-terminal domain (Figure 3B). This rotation is due to the flipping of Tyr 388 in helix R₁₂b. In the CMP structure, the phenolic oxygen (O₆) of Tyr 388 hydrogen-bonds to ordered water molecule Wat 1, which in turn hydrogen-bonds to the terminal phosphate oxygen of

**Figure 3:** Conformational changes in Δ24PmST1 upon binding to CMP. (A) Stereo ribbon drawing of the ligand-free apo structure drawn in tan and the CMP-bound binary structure drawn in green with CMP drawn as stick model with white-colored carbons. The structures were superimposed at the C-terminal CMP binding domain to illustrate the large conformational change upon binding to CMP. The C-terminal domains superimpose with an rmsd value of 2.41 Å for 156 equivalent α-carbons. The N-terminal domain rotates ~23° resulting in a movement of >18 Å for some loops in a direction shown by the arrow. (B) Close-up stereoview of the active site showing the movement of Δ24PmST1 upon binding to CMP. The CMP binary structure is drawn in green and the ligand-free apo structure is drawn semitransparent in tan color. Tyr 388 in helix α₁₂b moves up and flips around to hydrogen-bond to Ser 143 and Wat 1 (green sphere). Upon binding to CMP, helix α₁₁ moves slightly away from Tyr 388 to accommodate this large movement. Ser 143 in the N-terminal domain moves 5.2 Å to hydrogen-bond to both Tyr 388 and Wat 1. This interaction could stabilize the conformational change seen in the CMP-bound structure. Residues 375–385 correspond to the disordered helix α₁₂a in the CMP-bound state, which could be involved in sugar binding that would likely occur to the right of the CMP phosphate as shown.
In the apo structure, there would be insufficient room for this flipping because of the position of helix α11. When CMP binds to the enzyme, the phosphate in CMP binds to Ser 355 and Ser 356 at the N-terminal end of helix α11 and pulls it slightly away from helix α12b, thus, creating room for Tyr 388 to flip. Other significant C-terminal domain structural differences include the movement of β13, which can no longer hydrogen-bond with β12 because of the rotation of helix α12b. Additionally, helix α8 also rotates down toward the bound CMP. This movement is likely facilitated by the hydrogen bonding of main-chain carbonyl oxygen of Gly 266 (in the β8–α8 loop) to the N4 of the CMP cytosine base.

One must investigate the nature of the closing of the two Rossmann domains in Δ24PmST1 upon its binding to CMP, since it does not interact with the N-terminal domain (residues 25–246). Therefore, what induces the N-terminal domain to move? Inspection of the CMP-bound structure reveals that Ser 143 of helix α5a moves up 5.2 Å in the CMP-bound structure where the Oy hydrogen-bonds to both Water 1 and the Oy of Tyr 388. Water 1 in turn hydrogen-bonds to the terminal phosphate oxygen in the CMP moiety. The side chain of Tyr 388 flips ~180°, and the Ca moves 5.5 Å causing helix α12b to shift up upon CMP binding (Figure 3B). Therefore, it is likely that this water molecule may occupy the same position of a glycosidic oxygen in CMP–Neu5Ac. A similar conformational change, thus, may be induced in Δ24PmST1 upon its binding to CMP–Neu5Ac.

Comparison to Other GT-B Structures. A search of the PDB database for other structures with similar folds revealed that the closest match was bacterial UDP-N-acetylglucosamine 2-epimerase (57) (PDB ID, 1f6d). The DALI search resulted in a Z score of 10.3 (58). The rmsd between the two enzymes is 3.8 Å based on the structural alignment of 248 α-carbons. Despite the structural similarity, these two enzymes only share 9% sequence identity. Both structures belong to the glycosyltransferase group B (GT-B) structural family and have very similar topologies. Both have an N-terminal Rossmann nucleotide-binding domain consisting of a central seven-stranded parallel β-sheet sandwiched between α-helices and a C-terminal six-stranded nucleotide-binding domain. However, the Δ24PmST1 structure contains additional helices inserted between some of the β-strands. While the UDP–GlcNAc 2-epimerase structure only has a single helix between each β-strand except for the interdomain linkage, the Δ24PmST1 has three helices between β5 and β6 (α5a, α5b, and α5c), two helices between β6 and β7 (α6a and α6b), and two helices between β12 and β13 (α12a and α12b). Another major divergence in topology between these two structures is that the bacterial epimerase contains a C-terminal helix that extends out from the C-terminal nucleotide-binding domain and interacts extensively with the N-terminal domain (57). This interaction, which is uniquely absent in the Δ24PmST1 structure, appears to be common to all other members of the GT-B family whose structures are known (49, 59). In Δ24PmST1, the absence of this C-terminal helix tethering to the N-terminal domain may allow for greater flexibility of the substrate binding pocket between the two Rossmann domains to accommodate a wider variety of substrates and catalytic activities as measured for this enzyme (32).

The second highest structural similarity DALI score (Z score = 9.3) corresponds to the structure of the UDP–glucosyltransferase GtFB that modifies the heptapeptide aglycone in biosynthesis of vancomycin group antibiotics (60) (PDB ID, 1iir). The rmsd between this structure and that of Δ24PmST1 is 4.9 Å for 260 equivalent α-carbons, for which it shares only 6% identity in sequence. GtFB also contains two domains with similar topologies, but GtFB binds UDP–glucose. However, since this structure was determined in the absence of nucleotide, it is difficult to compare the active sites. Given that these enzymes have similar structures and bind sugar nucleotides, an evolutionary link is suggested between this family of sialyltransferases and the bacterial epimerase and glucosyltransferase.

The DALI search did not identify any significant correlation between the P. multocida sialyltransferase reported here and the only other sialyltransferase whose structure is known, CstII from C. jejuni (31). This was expected, since the CstII structure belongs to the GT-A structural family, which consists of a single Rossmann nucleotide-binding domain with a small lid inserted in the center of the Rossmann domain (between β4 and β5) involved in substrate binding (31). It was a surprise, though, that even DALI searches using each individual Rossmann domain of the Δ24PmST1 structure did not result in any homology with a Z score greater than 2.0 to the CstII structure.

In conclusion, the crystal structure of Δ24PmST1 represents the first GT-B type sialyltransferase and a new member of the GT-B glycosyltransferase family. Despite their significant structural similarity, GT-B-type glycosyltransferases have great sequence diversity and dramatic substrate differentiation (49). For example, T4 phage β-glucosyltransferase (BGT) (50, 61) transfers glucose from UDP–glucose (donor substrate) to hydroxymethylcytosine on duplex DNA (acceptor). E. coli MurG, another GT-B glycosyltransferase, transfers GlcNAc from UDP–GlcNAc (donor) to the C4 position of a lipid-linked N-acetylmuramyl peptide (acceptor) (46, 62). Additionally, GtFB transfers glucose from UDP–glucose to heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics (60). Thus, the donor substrates of the GT-B glycosyltransferases diverge from UDP–glucose, UDP–GlcNAc, to CMP–sialic acid. The acceptor substrates vary dramatically from duplex DNA, lipid-linked glycopolypetide, peptide, to oligosaccharide. A common feature shared among these GT-B-type glycosyltransferases is that they belong to microbial or bacteriophage inverting type glycosyltransferases (63). Because of their structural similarity to GT-B glycosyltransferases, both UDP–GlcNAc 2-epimerase (57) and glycogen phosphorylase (64–66) have been categorized together with GT-B glyco- syltransferases into a superfamily named GP GTF (glycogen phosphorylase/glycosyltransferase) superfamily (67). There is excellent structural conservation between protein members in this superfamily, particularly in the C-terminal nucleotide-binding domain. The GP GTF topology, with a large cleft between the two Rossmann domains, appears to be well-adapted to accommodate substrate molecules of varied sizes and diverse structures.

REFERENCES


