THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 11, Issue of March 16, pp. 8190–8196, 2001
Printed in U.S.A.

Structure of a Sialic Acid-activating Synthetase, CMP-acylneuraminate Synthetase in the Presence and Absence of CDP*

Received for publication, August 9, 2000, and in revised form, December 11, 2000
Published, JBC Papers in Press, December 11, 2000, DOI 10.1074/jbc.M007235200

Steven C. Mosimann‡§, Michel Gilbert¶, Dennise Dombroswki‡, Rebecca To¶, Warren Wakarchuk¶, and Natalie C. J. Strynadka‡‖

From the ‡Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver V6T 1Z3, Canada and the ¶National Research Council of Canada, Institute of Biological Sciences, 100 Sussex Dr., Ottawa K1A 0R6, Canada

The x-ray crystallographic structure of selenomethionyl cytosine-5′-monophosphate-acylneuraminate synthetase (CMP-NeuAc synthetase) from *Neisseria meningitidis* has been determined at 2.0-Å resolution using multiple-wavelength anomalous dispersion phasing, and a second structure, in the presence of the substrate analogue CDP, has been determined at 2.2-Å resolution by molecular replacement. This work identifies the active site residues for this class of enzyme for the first time. The detailed interactions between the enzyme and CDP within the mononucleotide-binding pocket are directly observed, and the acylneuraminate-binding pocket has also been identified. A model of acylneuraminate bound to CMP-NeuAc synthetase has been constructed and provides a structural basis for understanding the mechanism of production of "activated" sialic acids. Sialic acids are key saccharide components on the surface of mammalian cells and can be virulence factors in a variety of bacterial species (*e.g. Neisseria*, *Haemophilus*, group B streptococci, etc.). As such, the identification of the bacterial CMP-NeuAc synthetase active site can serve as a starting point for rational drug design strategies.

Cytosine-5′-monophosphate-acylneuraminate synthetase (cytosine-5′-monophosphate-*N*-acetyl neuraminic acid synthetase, CMP-NeuAc synthetase)[1] catalyzes the penultimate step in the addition of sialic acids to the oligosaccharide component of glycoconjugates and is a required component of sialylation pathways. CMP-NeuAc synthetase-deficient mutants do not express sialylated glycoconjugates and can be complemented with functional CMP-NeuAc synthetase in both mammalian (1) and bacterial systems (2). Sialic acids participate in a myriad of signaling, recognition, and cell-cell adhesion phenomena in mammalian cells (3) and are overexpressed in some highly malignant tumors (4). Genetic disorders that lead to altered physiological levels of sialic acids have many consequences in humans and can be fatal (5). In bacterial systems, sialic acids are less common and frequently have roles as virulence factors (6). In *Neisseria gonorrhoeae*, sialylated glycoconjugates protect the organism from phagocytosis and increase serum resistance (7). In *Haemophilus ducreyi* the presence of 2-keto-3-deoxy-manno-octulosonate-containing glycoconjugates correlate with the organism's pathogenicity (8). Likewise, the sialylated capsule of *Neisseria meningitidis*, *Escherichia coli* K1, and group B streptococci are virulence factors (9–11). It has been suggested that bacterial species produce sialylated glycoconjugates to mimic the host and escape detection by the immune system (12). Clearly, the mechanism of production of sialic acid-containing glycoconjugates is of clinical interest. Although bacterial and eucaryotic CMP-NeuAc synthetase enzymes share many catalytic properties, several important differences have been reported, including substrate specificity, tertiary structure, inhibitor sensitivity, and cellular localization (13). As a result, bacterial CMP-NeuAc synthetase enzymes can be targeted by rational drug design strategies.

CMP-NeuAc synthetase is also of considerable interest in the field of biotechnology. In the presence of Mg^{2+}, CMP-NeuAc synthetase utilizes CTP and acylneuraminate (NeuAc, a sialic acid) as substrates and produces pyrophosphate and cytosine-5′-monophosphate acylneuraminate (CMP-NeuAc or CMP-NANA), an "activated" sugar nucleotide monophosphate (Fig. 1) (14, 15). The reaction catalyzed by CMP-NeuAc synthetase (and homologous enzymes) is unique in that sialic acid is directly coupled to a cytosine monophosphate. All other activated sugar nucleotide synthetases use a phosphorylated sugar substrate and produce a sugar nucleotide diphosphate (16). Given the expense of chemically synthesizing CMP-NeuAc and its instability, CMP-NeuAc synthetase together with various sialyltransferases are valuable for the preparative enzymatic synthesis of biologically relevant, sialylated oligosaccharides (17). In turn, these sialylated oligosaccharides are used in clinical medicine and are invaluable tools for the study of the many processes that require sialic acids.

In this work, the refined x-ray crystallographic structures of selenomethionyl CMP-NeuAc synthetase and CMP-NeuAc synthetase in complex with the substrate analogue CDP (CMP-NeuAc synthetase-CDP) are reported. The active site and substrate binding sites for this class of enzyme are described in detail for the first time, and a comparison of the two structures indicates substrate binding is accompanied by significant conformational rearrangements within the active site.

EXPERIMENTAL PROCEDURES

Protein Purification—Recombinant CMP-NeuAc synthetase was overexpressed in *E. coli* and purified using published protocols (18) with modifications. Briefly we used a combination of anion-exchange on a MonoQ column (Amersham Pharmacia Biotech) and gel filtration on a Superose 12 column (Amersham Pharmacia Biotech). Selenomethionyl

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Present address: Dept. of Chemistry and Biochemistry, University of Lethbridge, 4401 University Dr., Lethbridge T1K 3M4, Canada.

‖ To whom correspondence should be addressed: Tel.: 604-221-0789; Fax: 604-221-5227; E-mail: natalie@byron.biochem.ubc.ca.

[1] The abbreviations used are: CMP-NeuAc synthetase, CMP-acylneuraminate synthetase or cytosine-5′-monophosphate-*N*-acetyl neuraminic acid synthetase; CMP-NeuAc, cytosine-5′-monophosphate-*N*-acetyl neuraminic acid; MAD, multiple wavelength anomalous dispersion; NeuAc, acylneuraminate; CMPKDOS, cytosine-5′-monophosphate-2-keto-3-deoxy-manno-octonate synthetase; KDO, 2-keto-3-deoxy-manno-octonate; r.m.s., root mean square.

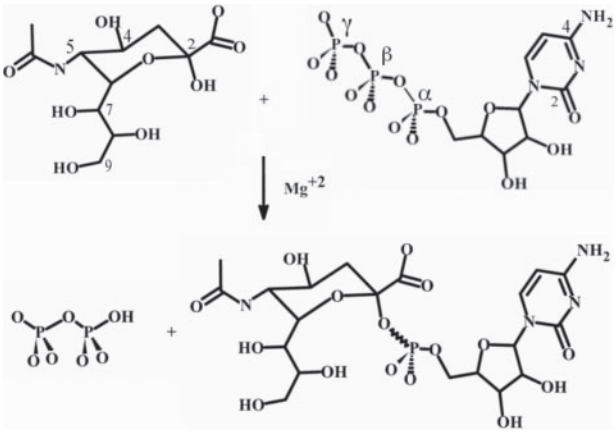

FIG. 1. **The reaction catalyzed by CMP-NeuAc synthetase in the presence of Mg^{2+}.** CMP-NeuAc synthetase specifically utilizes the β anomer of NeuAc (a sialic acid) and CTP as substrates and produces pyrophosphate and CMP-NeuAc, an activated sugar monophosphate nucleotide. In turn, CMP-NeuAc is substrate for sialyltransferases that add sialic acids to cellular glycoconjugates. Selected NeuAc and CTP atoms are labeled with their *corresponding number* or *Greek letter*.

CMP-NeuAc synthetase was overexpressed in a methionine auxotroph of *E. coli* (WA-834) grown in a defined medium containing 40 mg $liter^{-1}$ selenomethionine (19) and purified using the same protocol. All protein samples were purified to homogeneity and were of the expected molecular mass as judged by mass spectrometry and SDS-polyacrylamide gel electrophoresis.

Crystallization and Data Collection—All crystals used in this work were grown by vapor diffusion using hanging drops consisting of 3 μl of protein stock (5.0 mg ml^{-1} in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM dithiothreitol, and 0.1 mM EDTA) and 3 μl of reservoir solution. Crystals of CMP-NeuAc synthetase and CMP-NeuAc synthetase-CDP were both grown over reservoir solutions composed of 16% polyethylene glycol 2000, 100 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$ and 1 mM dithiothreitol. The CMP-NeuAc synthetase crystals grow in 2–3 days, are orthorhombic (space group, $P2_12_12_1$; unit cell, a = 42.99, b = 59.52, c = 157.56 Å; $\alpha = \beta = \gamma = 90°$), and a 2.0-Å resolution data set has been collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamline X8-C in the National Synchrotron Light Source. Orthorhombic crystals of CMP-NeuAc synthetase-CDP were grown in the presence of 10 mM CDP (unit cell, a = 43.01, b = 69.89, c = 157.78 Å; $\alpha = \beta = \gamma = 90°$), appeared after 7 days, and diffracted to 2.2 Å on a local Rigaku RU200 rotating anode (50 kV, 100 mA) equipped with OSMIC mirrors. All CMP-NeuAc synthetase crystals in this work were rod-shaped and the largest crystals grew to ~0.80 × 0.15 × 0.15 mm in size. All intensity data sets were collected at 100 K after flash-freezing in mother liquor containing 22% glycerol. Intensity data collection and processing statistics for each data set used in this work are summarized in Table I.

Structure Solution—The structure of the CMP-NeuAc synthetase homodimer was determined by MAD phasing based on 10 selenium atom positions. The Se atom positions were identified using both SHAKE'N'BAKE (20) and SOLVE (21) and subjected to positional, isotropic *B*-factor, and occupancy refinement within SOLVE prior to the calculation of phases (Table I). The initial phase estimates were improved by density modification using DM (22) prior to calculation of a 2.0-Å MAD phased electron density map. The initial electron density map is easily interpretable and an initial trace of the CMP-NeuAc synthetase main chain was generated within XTALVIEW (23). The CMP-NeuAc synthetase-CDP structure was determined by molecular replacement using the CMP-NeuAc synthetase model and the program AMORE (24). Both, CMP-NeuAc synthetase[2] and CMP-NeuAc synthetase-CDP[3] were refined using CNS 1.0 (25), and all interactive fitting of the models made use of XTALVIEW. Table II lists common statistical indicators for both refined models.

Structure Analysis and Figure Preparation—PROMOTIF (26) and PROCHECK (27) were used to identify secondary structures and to check the stereochemical quality of both models throughout the refinement. SURFACE (22) and GRASP (28) were utilized to examine the molecular surface of the models and to examine the local environment of specific residues. XTALVIEW was used for various least-squares superpositions and DOCKVISION (29) was used for the NeuAc modeling. Docking made use of an energy-minimized acylneuraminate structure derived from Protein Data Bank accession code 1WIA and the refined CMP-NeuAc synthetase-CDP model. The starting NeuAc model was used to generate all common NeuAc conformations, and all conformations were docked (10,000–50,000 trials) against a static CMP-NeuAc synthetase-CDP model using either an energy term or conformational energy terms as a scoring function. The best docking results were obtained using CMP-NeuAc synthetase-CDP and CDP conformation I. The figures were prepared using GRASP (28) and MOLSCRIPT (30).

RESULTS AND DISCUSSION

Overall Fold—The overall fold of the CMP-NeuAc synthetase-CDP homodimer is presented in Fig. 2. The CMP-NeuAc synthetase monomer is composed of an ~190 residue, globular α/β domain (~50 × 50 × 40 Å) that is structurally classified as an $\alpha\beta\alpha$, three-layered sandwich (31) and a small 35-residue domain that is directed away from the rest of the monomer and forms an extended dimerization interface. The dimerization domains of each monomer wrap around one another in a tight interaction and make contacts with the globular α/β domain of the opposite monomer. The resulting functional homodimer is highly asymmetric with dimensions of ~85 × 45 × 40 Å.

A twisted, central β-sheet forms the hydrophobic core of the α/β domain of the monomer and is flanked by helices giving rise to the $\alpha\beta\alpha$, three-layered sandwich structure. The β-sheet contains six parallel β strands (β1, β4, β5, β6, β7, β11) and a single antiparallel β strand (β10) connected with a −1x, −1x, +3x, +2x, −1x, +2x topology (32). The connectivity of consecutive parallel β strands is right-handed except for a relatively rare left-handed linkage between strand β6 and strand β7. Alignments of known CMP-NeuAc synthetase sequences (33) indicate that residues between strands β6 and β7 (102–110) are among the most highly conserved in the comparison (Fig. 3). Three α helices (B, C, F) pack against and bury the face of the central β-sheet furthest from the dimerization interface. The remaining face of the β-sheet is covered by two α helices (E, H) and contacts with the dimerization domain of the opposite monomer. The 3_{10} helix A, a short β ribbon (β2, β3), 3_{10} helix D, and α helix I are all located on the C-terminal side of the mostly parallel, central β-sheet. These secondary structures and their associated loops form a deep pocket at the C-terminal ends of the β-sheet that is the mononucleotide-binding pocket. The presence of a substrate-binding pocket at the C-terminal ends of predominantly parallel β-sheets is a recurring theme in structural biology (34).

The dimerization domain contains a 3_{10} helix (helix G) and a β ribbon (strand β8 and strand β9) that is continuous with the central β-sheet of the opposite monomer. The dimerization domain is located between strands β7 and β10 and is directed away from the rest of the α/β domain by a β-bulge at position 175 of strand β10. In the homodimer, a β-bulge in strand β10 facilitates an antiparallel hydrogen-bonding network between the N terminus of strand β10 and strand β8 of the opposite monomer. The dimerization interface buries a large surface area despite its modest size. Using a probe radius of 1.4 Å, a single monomer buries ~1450 $Å^2$ in the dimer (35).

A search of representative structures in the Protein Data Bank using the DALI server (36) indicates the CMP-NeuAc synthetase overall fold has only limited homology with other known folds in the Protein Data Bank. In general, enzymes containing a mononucleotide-binding (Rossman) fold (37) resemble the N-terminal 100 residues of the α/β domain, up to but not encompassing the left-handed linkage between strands

[2] The atomic coordinates for CMP-NeuAc synthetase are available in the Protein Data Bank under PDB 1EYR.

[3] The atomic coordinates for CMP-NeuAc synthetase-CDP are available in the Protein Data bank under PDB 1EZI.

TABLE I
Summary of data collection and MAD phasing statistics for CMP-NeuAc synthetase-CDP and CMP-NeuAc synthetase

	CMP-NeuAc synthetase		Selenomethionyl CMP-NeuAc synthetase[a]	
	CDP	Peak	Inflection	Remote
Resolution (Å)	20.0–2.2	20.0–2.0	20.0–2.0	20.0–2.0
Wavelength (Å)	1.5418	0.9790	0.9786	0.9500
Ref. unique	21889	26150	25681	25606
Redundancy	6.3	9.1	8.2	8.1
Complete (%)	87.9	99.0	99.0	99.0
$I/\sigma(I)$	13.1	21.0	21.2	20.7
Rmerge[b]	0.061	0.074	0.071	0.073
Unit cell				
a (Å)	43.01		42.99	
b (Å)	69.89		59.52	
c (Å)	157.48		157.56	
$\alpha = \beta = \gamma$ (°)	90.0		90.0	
Number of Se sites			10	
SOLVE Z-score[c]			56.6	
Refined Se occupancy			0.61–1.46	
Refined Se B_{iso}[d]			20.6–60.0	
Figure of merit			0.44	

[a] Data were collected at the NSLS DataCol'99 Workshop on rapid structure determination techniques (April 17–22, Brookhaven, NY). The time between the collection of three-wavelength MAD data and the calculation of an interpretable electron density map was 16 h.
[b] $R_{merge} = \Sigma|I_{hkl,o} - \langle I_{hkl,o}\rangle|$, where $\langle I_{hkl,o}\rangle$ is the mean intensity of reflection o and $I_{hkl,o}$ represents the individual measurements of reflection o and its symmetry equivalent reflections.
[c] Z-score = corrected sum of standard deviations above mean for four separate SOLVE criteria.
[d] $B_{iso,i} = 8\pi ui2$, where $ui2$ is the mean vibrational amplitude of atom i.

TABLE II
Summary of refinement statistics for selenomethionyl CMP-NeuAc synthetase and CMP-NeuAc synthetase in complex with CDP

	CMP-NeuAc synthetase-CDP	Se CMP-NeuAc synthetase
Resolution (Å)	20.0–2.2	20.0–2.0
R^a/R_{free}[b]	0.185/0.247	0.212/0.267
r.m.s.[c] bonds (Å)	0.008	0.008
r.m.s. angles (°)	1.3	1.5
Model		
A	1–225	1–73, 80–225
B	1–148, 151–223	1–14, 18–74, 81–222
CDP	2	N/A
Solvent	446	404
$B_{iso,ave}$[d], monomer A	31.9	31.6
$B_{iso,ave}$, monomer B	32.8	31.5
$B_{iso,ave}$, CDP	36.6	N/A
$B_{iso,ave}$, solvent	41.6	40.8

[a] $R = \Sigma(|F_{obs} - F_{calc}|)/\Sigma(|F_{obs}|)$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[b] R_{free} is defined as for R, except F_{obs} is a set of test reflections (5%) that have been excluded from the refinement.
[c] r.m.s. = $(\Sigma(X_r - X_{ideal})^2/N)^{0.5)}$, where X_x and X_{ideal} are observed and target values, respectively and N is the number of observations.
[d] $B_{iso,ave}$ is the average of the mean vibrational amplitude (B-factor) for a given set of atoms.

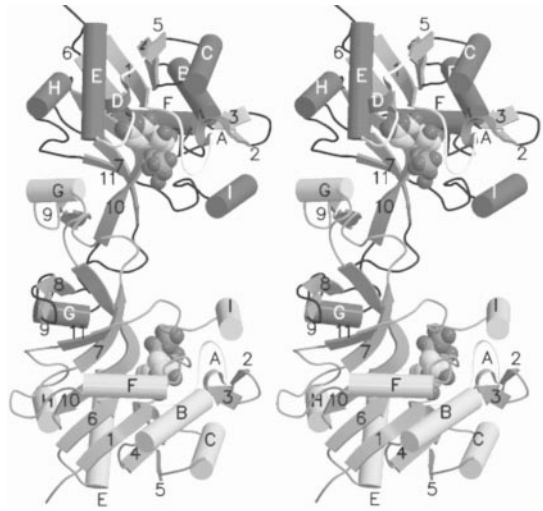

FIG. 2. **The CMP-NevAc synthetase homodimer in complex with CDP.** CMP-NevAc synthetase is shown as a barrel (*helices*) and arrow (*strands*) drawing, and the CDP substrate analogue is shown as a van der Waals surface. Monomer A is indicated with *dark gray helices* and *strands*, while monomer B is indicated with *light gray helices* and *strands*. CMPANS has a single α/β domain with an $\alpha\beta\alpha$ three-layered sandwich fold and 35 residue protrusion that forms the dimerization interface. The 11 β strands (*arrows*) are arranged in a 7-stranded, mostly parallel β-sheet and two β ribbons. The eight helices pack against either side of the central β-sheet or form part of the dimerization interface (*helix G*).

β6 and β7. The similarities between the N terminus of CMP-NeuAc synthetase and Rossman fold containing proteins are restricted to the order and connectivity of secondary structures, because CMP-NeuAc synthetase and related enzymes have unique mononucleotide-binding pockets. The CMP-NeuAc synthetase α/β domain fold is similar to that reported for cytosine-5′-monophosphate-2-keto-3-deoxy-manno-octonate synthetase (CMPKDOS (38)), a homologous, dimeric enzyme that synthesizes a related, activated sugar compound, cytosine-5′-monophosphate 2-keto-3-deoxy-manno-octonate (CMP-KDO). These two enzymes are similar in length but share limited amino acid sequence identity (18%, Fig. 3), and the fold of the CMPKDOS dimerization domain is significantly different from the equivalent structure in this work. In the CMPKDOS monomer, the dimerization domain is globular and compact and does not make contacts with the α/β domain of the opposite monomer. Despite the differences in the dimerization interface, both enzymes share a highly asymmetric overall shape. Because the CMPKDOS coordinates are yet to be released and were not made available to us, we cannot present a detailed comparison of the CMP-NeuAc synthetase and CMPKDOS overall folds.

Comparison of CMP-NeuAc Synthetase and CMP-NeuAc Synthetase-CDP—The two monomers of CMP-NeuAc synthetase-CDP in the asymmetric unit have nearly identical main-chain conformations. The monomers superpose with an r.m.s. deviation of 0.60 Å when considering all common main-chain atoms (888 of 912 atoms). Similarly, there are no large-scale conformational differences in a comparison of the CMP-NeuAc synthetase- and CMP-NeuAc synthetase-CDP-refined models and the two homodimers superpose with an r.m.s. deviation of 0.71 Å for 1728 equivalent main-chain atoms (Fig. 4). The most prominent differences in a comparison of the two refined models occur in the mononucleotide-binding pocket. In the unligan-

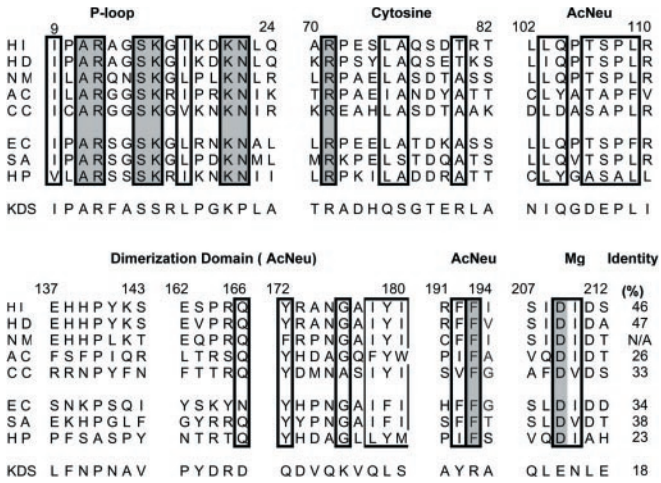

FIG. 3. **Amino acid sequence alignment of the substrate binding regions of known bacterial CMP-NeuAc synthetase enzymes and a representative bacterial CMPKDOS.** *Shaded boxes* indicate amino acid identity, *unshaded boxes* indicate conservative amino acid replacements and *labels* indicate the role of each region in substrate binding. The equivalent residues from a representative CMPKDOS sequence are shown *below* and the residue numbers are from the *N. meningitidis* enzyme. HI (*H. influenza*), HD (*H. ducreyi)*, NM (*N. meningitidis*), AC (*Aeromonas caviae*), CC (*Campylobacter coli*), EC (*E. coli*), SA (*Streptococcus agalactiae*), HP (*Helicobacter pylori*), KDS (CMPKDOS from *H. influenzae*).

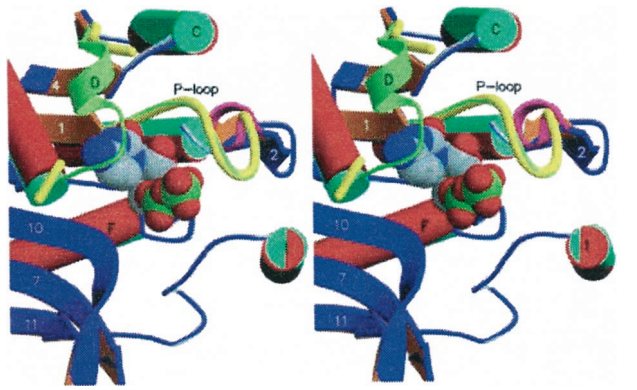

FIG. 4. **CMP-NeuAc synthetase active site in the presence and absence of CDP.** The CMP-NeuAc synthetase-CDP (*blue strands* and *green helices*) and CMP-NeuAc synthetase (*dark orange strands* and *red helices*) structures superpose with a r.m.s. deviation of 0.71 Å for all common main-chain atoms, and the only relatively, large conformation differences are associated with residues implicated in mononucleotide binding. In the presence of CDP, the P-loop and helix D (*light green ribbon*) are well ordered and directly contact the bound substrate analogue. In the absence of CDP, the equivalent residues are poorly ordered (*P-loop*) or disordered (*helix D*). This suggests substrate binding is accompanied by conformational rearrangements that orders residues involved in mononucleotide binding.

ded CMP-NeuAc synthetase structure, the P-loop (residues 10–22) and residues following strand β5 (71–80) have relatively weak electron density or are fully disordered in the refined maps. In the CMP-NeuAc synthetase-CDP structure, the equivalent residues make specific contacts with the bound substrate analogue (Table III and see Fig. 6) either directly or through ordered solvent molecules, and have well ordered electron density in the refined maps. The average main-chain temperature factors for residues 10–22 and 71–80 of CMP-NeuAc synthetase-CDP and CMP-NeuAc synthetase are 33.1 $Å^2$ (184 atoms) and 50.3 $Å^2$ (128 atoms), respectively. In CMP-NeuAc synthetase-CDP, the P-loop primarily interacts with the ribose and phosphate moieties of the substrate analogue, whereas residues 71–80 are responsible for base recognition. At the same time, these residues make relatively few specific

TABLE III
Hydrogen bonds and van der Waals contacts between CMP-NeuAc synthetase and CDP

	CDP conformation I		CDP conformation II	
		Å		Å
Hydrogen bonds				
O2	Arg-71 $N^{\eta 2}$	2.9	Arg-71 $N^{\eta 2}$	3.0
	Arg-12 N	2.9	Arg-12 N	3.0
N3	Arg-71 $N^{\eta 1}$	2.9	Arg-71 $N^{\eta 1}$	3.0
N4	Ala-80 O	3.0	Ala-80 O	3.0
	Wat-94	3.0	Wat-94	3.1
O2′	Asn-22 $N^{\delta 2}$	2.9	Asn-22 $N^{\delta 2}$	3.2
	Wat-160	3.1	Wat-160	2.8
O3′	Asn-22 $O^{\delta 1}$	2.6	Asn-22 $O^{\delta 1}$	2.7
O1A	Lys-21 N^{ζ}	3.3		
O3A	Arg-12 $N^{\eta 1}$	3.3		
O1B	Lys-21 N^{ζ}	3.2		
O2B	Arg-12 $N^{\eta 1}$	2.8	Asp-211 $O^{\delta 1}$	2.9
			Asp-209 $O^{\delta 1}$	2.9
O3B	Ser-15 O^{η}	3.2		
	Arg-12 $N^{\eta 1}$	3.0		
	Wat-160	2.8		
Protein contacts				
Base	Leu-10 C^{β}	3.3	Leu-10 C^{β}	3.3
	Ala-11 C	2.9	Ala-11 C	2.9
	Arg-12 C^{γ}	3.1	Arg-12 C^{γ}	3.2
	Leu-75 O	3.9	Leu-75 O	3.8
	Ser-81 O	3.9	Ser-81 O	4.0
Ribose	Ser-82 $O^{\gamma 1}$	3.7	Ser-82 O^{η}	3.7
	Gln-104 $O^{\delta 1}$	3.8	Gln-104 $O^{\delta 1}$	3.8
	Pro-105 C^{δ}	3.6	Pro-105 C^{δ}	3.6
	Thr-106 $O^{\gamma 1}$	3.4	Thr-106 $O^{\gamma 1}$	3.4
Phosphate	Asp-211 $O^{\delta 1}$	3.8	Ile-210 O	3.6

contacts with the remainder of the enzyme. Apparently, CDP itself provides specific contacts that allow these residues to fold about the mononucleotide substrate analogue and become relatively ordered. Consequently, we propose mononucleotide binding is accompanied by a conformational rearrangement that orders both the P-loop and residues 71–80. This model of substrate binding can account for the observed differences in the refined electron density of the two structures and the observed depth of the mononucleotide-binding pocket. It also explains the ordered bi bi kinetic mechanism (CTP binds first and CMP-NeuAc dissociates last) observed in the homologous enzyme from *H. ducreyi* (39). The P-loop and residues 71–80 contain 7 of the 13 invariant residues in an alignment of known bacterial CMP-NeuAc synthetase enzymes (Fig. 3). In agreement with this work, amino acid sequence analysis, site-directed mutagenesis, and chemical modification studies indicate residues 10–22 (phosphate binding loop or P-loop) following strand β1 are integral components of the mononucleotide-binding pocket (33, 40). Likewise, in the CMPKDOS structure (38), residues of the P-loop were observed to bind a heavy atom derivative ($IrCl_6^{3-}$), and the authors speculated that these residues play a role in phosphate binding.

Mononucleotide Binding Pocket—The molecular surface of CMP-NeuAc synthetase in the presence of CDP is shown in Fig. 5. The large cleft at the C-terminal end of the central β-sheet is the mononucleotide-binding pocket. The cleft is formed by residues following strand β1 (P-loop), strand β5 (71–80), and strand β11 (208–212). Polar residues line the cleft and are responsible for binding the nucleotide substrate (Fig. 6 and Table III). As shown in Fig. 7, the substrate analogue, CDP, is bound in the mononucleotide-binding pocket in two distinct conformations. The cytosine base and ribose moiety are equivalent in the two conformations, whereas the α and β phosphoryl groups adopt different conformations (I, II) as a result of differences in the C4′-C5′-O5′-P1A and C5′-O5′-P1A-O3A torsion angles.

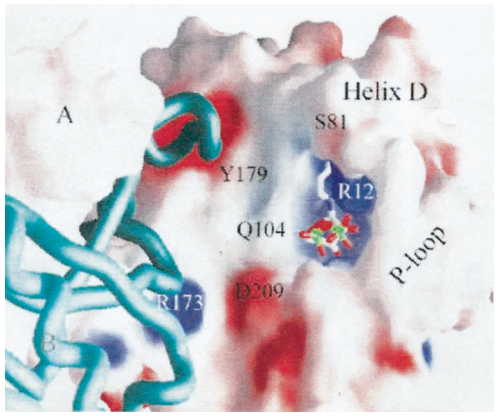

FIG. 5. **Molecular surface of the CMP-NeuAc synthetase active site.** Monomer A is shown as a GRASP surface colored according to its calculated electrostatic potential (*red* for negative and *blue* for positive). Monomer B is shown as C^{α} trace (*light blue tube*). The CDP-binding pocket is deep and lined with polar residues from the P-loop (residues 10–22) and loops on either side of helix D (residues 70–72 and 76–82). Adjacent to the CDP-binding pocket is the NeuAc-binding pocket, a second deep cleft comprising residues from both monomers of the CMP-NeuAc synthetase homodimer. In particular, Ser-82, Gln-104, Thr-106, and Tyr-179 of monomer A and Lys-142 and Arg-165 of monomer B line the NeuAc-binding pocket.

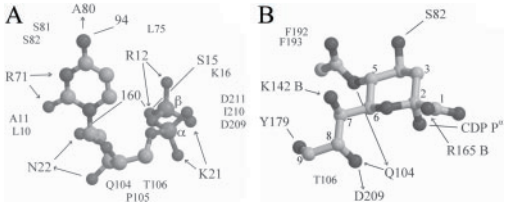

FIG. 6. **Observed and modeled contacts between CMP-NeuAc synthetase and CDP and CMP-NeuAc synthetase and NeuAc, respectively.** *A*, observed hydrogen bonds and contacts between CMP-NeuAc synthetase and CDP. Hydrogen bonds are indicated with *solid lines* and residues providing contacts are listed adjacent to the atoms they contact. *Arrows* indicate donated hydrogen bonds as inferred from the structure. Two ordered, solvent molecules form hydrogen bonds with CDP and are indicated as *circles*. *B*, modeled hydrogen bonds (*dashed lines*) and contacts from the docking (29) of NeuAc with CMP-NeuAc synthetase-CDP. The model suggests a number of plausible interactions. Residues at positions 82, 104, 106, 142, 165, 179, 193, and 209 are conserved or conservatively replaced in CMP-NeuAc synthetase sequences from *Neisseria*, *Haemophilus*, *Aeromonas*, and *Campylobacter* species (~230 residues). The equivalent residues, excepting 142, are conserved in the larger CMP-NeuAc synthetase enzymes of *Escherichia*, *Helicobacter*, and *Streptococcus* species.

The cytosine base is completely buried (521 $Å^2$) in the deep cleft, and its position is fixed by a series of hydrogen bonds. The $N^{\eta 1}$ and $N^{\eta 2}$ atoms of the invariant Arg-71 donate hydrogen bonds to O2 and N3 of the base and give rise to the enzymes specificity for cytosine nucleotides. Arg-12 N also donates a hydrogen bond to O2 while Ala-80 O and an ordered solvent molecule accept hydrogen bonds from N4. These interactions completely satisfy the hydrogen bonding potential of the base. The cytosine base orientation is further fixed by van der Waals contacts with the Leu-10 side chain and a stacking interaction with the Arg-12 guanidinium. Together, the Leu-10 and Arg-12 side chains form a narrow pocket that contacts both faces of the base. The residues that bind or contact the cytosine base are all derived from the P-loop and residues following strand $\beta 5$ (71–80). These are the same regions that are largely disordered in the CMP-NeuAc synthetase structure determined in the absence of a substrate analogue. The hydrogen bonding potential of the ribose O2′ and O3′ atoms are satisfied by Asn-22 $N^{\delta 2}$ and $O^{\delta 1}$, respectively. Residues Gln-104, Pro-105, and Thr-106 form the remainder of the ribose subsite of the mononucleotide-

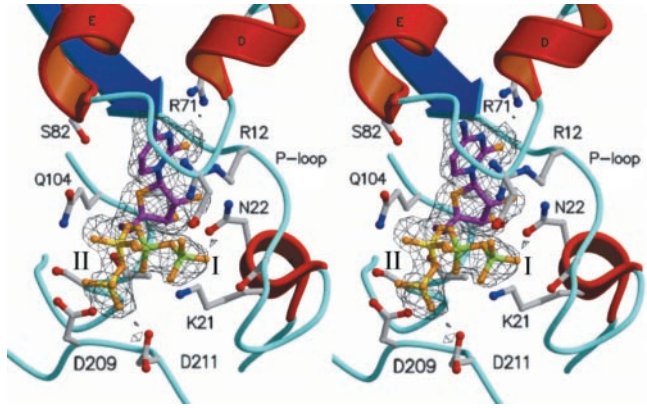

FIG. 7. **The refined $2F_o - F_c$ electron density associated with the bound CDP substrate analogue.** The CDP-binding site of CMP-NeuAc synthetase-CDP is shown as a barrel and strand drawing while residues directly contacting the bound CDP and the two observed CDP conformations are indicated in *ball-and-stick*. The two diphosphate conformations of CDP are labeled *I* (*green phosphorous atoms*) and *II* (*yellow phosphorous atoms*). The biologically significant conformation of the α and β phosphoryl groups is conformation I (*green phosphorous atoms*). The refined $2F_o - F_c$ electron density for the bound CDP is contoured at 1.2 σ.

binding pocket and are part of the conserved left-handed linkage between strand $\beta 6$ and strand $\beta 7$. Leu-10, Arg-12, and several ordered solvent molecules also contribute to the ribose subsite. Residues of the P-loop are the most highly conserved in sequence alignments of bacterial CMP-NeuAc synthetase enzymes and residues Arg-12, Ser-15, Lys-16, Lys-21, and Asn-22 are invariant in these alignments (Fig. 3).

Notably, Lys-16 and Asn-22 are not conserved in CMPKDOS sequences. Whereas it has been demonstrated that Lys-16 is not essential for CMP-NeuAc synthetase activity (40), catalysis is more efficient in the Lys-16-containing holoenzyme. The lack of conservation of Asn-22 is more difficult to understand. In CMP-NeuAc synthetase, Asn-22 hydrogen bonds to the O2′ and O3′ atoms of the ribose moiety and apparently discriminates between CTP and dCTP. CMPKDOS has a proline residue at the equivalent position in amino acid sequence alignments. Clearly, proline is unable to specifically interact with the ribose in a manner similar to that observed in CMP-NeuAc synthetase. In the absence of a description of the CMPKDOS active site it is unclear how CMPKDOS specifically binds the ribose moiety and discriminates between CTP and dCTP.

The two conformations of the α and β phosphoryl groups of the substrate analogue CDP are a result of differences in the torsion angles about the C5′-O5′ and O5′-P1A bonds (Fig. 7), respectively. They have been modeled as equivalent conformations during the refinement and have comparable average isotropic *B*-factors. Conformation I of the α and β phosphoryl group has a C4′-C5′-O5′-P1A torsion angle of $-115°$ and a C5′-O5′-P1A-O3A torsion angle 155°, whereas conformation II has torsion angles of $-160°$ and 50°, respectively. Several observations strongly suggest that conformation I closely mimics that of the substrate, CTP: (*a*) the presence of an oxyanion-hole-like structure formed by Ser-15 N and Gly-17 N that is adjacent to the β phosphate and is suitably positioned to interact with a γ phosphate, (*b*) the presence of a potential Mg^{2+}-binding site bridging the phosphoryl groups of the substrate and including the relatively conserved Asp-209 and Asp-211, and (*c*) the number and nature of the contacts between CMP-NeuAc synthetase and the substrate analogue phosphoryl groups (Table III and Fig. 6). Additionally, in conformation I a single, ordered water molecule (Wat-160) hydrogen bonding to Asn-14 O and the β phosphate, is suitably positioned to donate a hydrogen to the pyrophosphate leaving group. In CDP con-

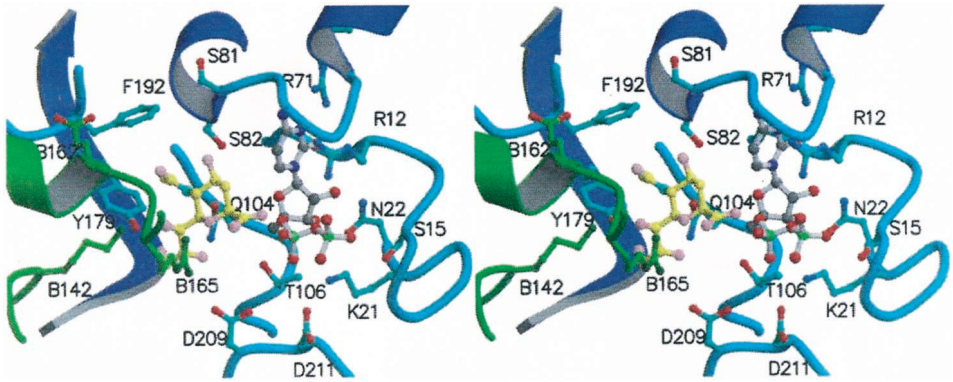

FIG. 8. **A model of NeuAc (*yellow carbons*) docked in the NeuAc binding cleft and its relation to the bound CDP (*black carbons*).** In the lowest energy docking, the NeuAc carboxylate is directed away from the cleft and the O4 and *N*-acetyl functional groups are directed into the cleft. The hydroxyl groups, O7, O8, and O9 of NeuAc are within hydrogen-bonding distance of Lys-142 (monomer B), Asp-209 and Tyr-179, respectively. The view has been chosen to show the contacts between the modeled NeuAc and CMP-NeuAc synthetase. The nucleophilic O2 atom (hidden under the NeuAc carboxylate) is directed at the α phosphate of CDP and is positioned for an S_N2 attack. There are no protein atoms suitably positioned to act as a general base, and an ordered solvent molecule likely serves as the general base in this mechanism.

formation II, there are no apparent γ phosphate- or Mg^{2+}-binding sites and relatively fewer contacts between the enzyme and phosphoryl groups. Why then do we observe conformation II in this work? The C4′-C5′-O5′-P1A dihedral angle (conformation II) adopts a staggered conformation and relieves the stereochemical strain arising from the eclipsed C4′-C5′-O5′-P1A dihedral angle present in conformation I. The absence of a binding site for the γ phosphate in conformation II suggests CTP will not bind in two distinct, equally populated conformations as observed with CDP.

The bulk of the interactions with phosphate moieties, in conformation I, are from residues of the P-loop. This is similar to the case for many Rossman fold proteins, which also have a phosphate binding or P-loop immediately following their N-terminal β strand. However, the details of the interaction between the P-loop and the base and phosphate are very different in these two groups of enzymes. This is not surprising because CMP-NeuAc synthetase releases pyrophosphate as a product, whereas kinases, G-proteins, and other Rossman fold-containing proteins release a single phosphate, the γ phosphoryl group (37). In conformation II, there are fewer contacts between the phosphates and the protein.

NeuAc Binding Pocket—As seen in Fig. 5, the active site cleft of CMP-NeuAc synthetase extends beyond the mononucleotide-binding pocket toward the dimerization domain and forms the NeuAc-binding pocket. The pocket is lined with a number of polar side chains (Fig. 5, 6) and can easily accommodate NeuAc. Atoms O4 and O9 and the NeuAc carboxylate are solvent-exposed and may interact with ordered solvent molecules in addition to CMP-NeuAc synthetase. Alternatively, substrate binding may involve a small conformational change in CMP-NeuAc synthetase that decreases the volume of the binding pocket. The polar side chains of residues Ser-82, Gln-104, Thr-106, and Asp-209 line the NeuAc pocket nearest the bound substrate analogue (CDP), whereas Tyr-179 and residues Glu-137, Lys-142, Glu-162, and Arg-165 from the dimerization domain of the opposite monomer form the edges of the binding pocket furthest from the bound CDP. Interestingly, 4 of the 9 polar residues within the proposed NeuAc-binding pocket are derived from residues of the dimerization domain. This implies that the sequence and structure of the dimerization domain is an important determinant for the discrimination of various sialic acids. Thus, the dimerization domains of homologous enzymes with specificities for related sialic acid derivatives are likely to have limited amino acid sequence identity and adopt a conformation different from that observed in CMP-NeuAc synthetase. As pointed out earlier, the dimerization interfaces of CMP-NeuAc synthetase and CMPKDOS are very different, and sequence alignments indicate the dimerization domains of CMP-NeuAc synthetase enzymes are not highly conserved (35, Fig. 3).

NeuAc Modeling—A model of an energy-minimized NeuAc that has been docked in the NeuAc-binding pocket is presented in Fig. 8. The docked NeuAc model produces the lowest energy using either the DockVision energy function (−31.04 kJ) or conformational energy function (−9.40 kJ). In the docked model, the NeuAc functional groups make a number of plausible interactions with invariant or conserved residues (Figs. 6 and 8). The axial, O2 atom of NeuAc is directed at the α phosphate of the biologically significant CDP conformer (conformation I) in accordance with the known specificity of CMP-NeuAc synthetase (41). The NeuAc C2 carboxylate (a strong acid) is directed away from the cleft and forms a salt bridge with the Arg-165 guanidinium in the model. As seen in Fig. 3, a positively charged amino acid able to counter the negative charge of the C2 carboxylate is present at positions 164 or 165 of each amino acid sequence. The *N*-acetyl group at C5 of NeuAc docks with the acetyl oxygen exposed to bulk solvent, the nitrogen atom hydrogen bonding with Gln-104 O^{ϵ} and the methyl group packing against the aromatic residues Tyr-179, Phe-192, and Phe-193. Although, these residues are all conserved among CMP-NeuAc synthetase enzymes (Fig. 3), only Gln-104 is conserved between CMP-NeuAc synthetase and CMPKDOS enzymes. This is reasonable because KDO, the CMPKDOS substrate, has a hydroxyl at position C5 (equivalent to the nitrogen of NeuAc) but has no counterpart to the carbonyl oxygen and methyl group of NeuAc. Consequently, only the Gln-104 interaction with the C5 substituent is expected to be conserved between these enzymes.

The O4 atom of NeuAc is paired with the Ser-82 OH in this model and may also interact with the Ala-80 N. At the base of the binding pocket, the Gln-104 side chain contacts both the O8 and N5 atoms of NeuAc model and is suitably positioned to help CMP-NeuAc synthetase discriminate between the sialic acids with variable C6 functional groups. The O8 atom is also within contact distance of the Asp-209 carboxylate. The O7 atom and O9 atoms of NeuAc are directed at the dimerization domain of the opposite monomer and are within hydrogen-bonding distances of the Lys-142 and Tyr-179 side chains. Additional residues may interact with the NeuAc model via bridging solvent molecules or may directly interact with NeuAc if CMP-NeuAc synthetase undergoes a small conformational change upon substrate binding.

In the CMP-NeuAc synthetase reaction mechanism, the O2

atom of NeuAc is the nucleophile in an S_N2 attack on the α phosphate (41). Our structure and docking results support this mechanism and we predict the leaving group pyrophosphate is stabilized by the absolutely conserved Arg-12, Lys-21, one or more bridging Mg^{2+} cations, and possibly the conserved Lys-16 (Fig. 3). As previously mentioned, an ordered solvent molecule hydrogen bonding to Asn-14 O is suitably positioned to donate a hydrogen bond to the leaving group pyrophosphate. In the model, the nucleophilic NeuAc O2 atom is not within hydrogen-bonding distance of any atom of the enzyme and, consequently, there is no readily apparent general base. An ordered solvent molecule may serve as the general base or, alternatively, a conformational rearrangement may bring a general base within hydrogen-bonding distance of the O2 atom upon NeuAc binding. We favor a mechanism in which an ordered solvent molecule serves as the general base, because it seems unlikely that a conformational rearrangement could bring a protein atom within hydrogen-bonding distance of the NeuAc O2 atom without disrupting the specific interactions detailed above.

Acknowledgments—We thank M.-F. Karwaski for technical help in the purification of the CMP-NeuAc synthetase and the selenomethionyl CMP-NeuAc synthetase, D. Krajcarski for mass spectrometry measurements, and the instructors and staff at the National Synchrotron Light Source Data Col'99 course for their generous assistance. This work was supported by an NSERC strategic grant (to W. W. and N. S.). S. M. is a NSERC post-doctoral fellow. N. S. is a Canadian Institute of Health Research Scholar, a Howard Hughes Medical Institute International Scholar, and a Burroughs Wellcome New Investigator.

REFERENCES

1. Munster, A. K., Eckhardt, M., Potvin, B., Muhlenhoff, M., Stanley, P., and Gerardy-Schahn, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9140–9145
2. Haft, R. F., Wessels, M. R., Mebane, M. F., Conaty, N., and Rubens, C. E. (1996) *Mol Microbiol.* **19,** 555–563
3. Kelm, S., and Schauer, R. (1997) *Int. Rev. Cytol.* **175,** 137–240
4. Kern, W. F., Spier, C. M., Miller, T. P., and Grogan, T. M. (1993) *Leuk. Lymphoma* **12,** 1–10
5. Rosenberg, A. (1995) *The Biology of Sialic Acids*, Plenum Press, New York
6. Jann, B., and Jann, K. (1990) *Curr. Top. Micorbiol. Immunol.* **150,** 19–42
7. Rest, R. F., and Frangipane, J. V. (1992) *Infect. Immun.* **60,** 989–997
8. Lagergard, T. (1992) *Microb. Pathog.* **13,** 203–217
9. Bitter-Suermann, D. (1993) *Polysialic Acid: From Microbe to Man*, Birkhauser Verlag, Basel
10. Jann, K., and Jann, B. (1985) *The Virulence of Escherichia coli: Reviews and Methods*, Academic Press, Orlando, Florida
11. Timmis, K. N., Boulnois, G. J., Bitter, S. D., and Cabello, F. C. (1985) *Curr. Top. Microbiol. Immunol.* **118,** 197–218
12. Mandrell, R. E., and Apicella, M. A. (1993) *Immunobiology* **187,** 382–402
13. Kean, E. L. (1991) *Glycobiology* **5,** 441–447
14. Warren, L., and Blacklow, R. S. (1962) *J. Biol. Chem.* **237,** 3527–3534
15. Roseman, S. (1962) *Proc. Natl. Acad. Sci. U. S. A.* **48,** 437–441
16. Kapitonov, D., and Yu, R. K. (1999) *Glycobiology* **9,** 961–978
17. Gilbert, M., Bayer, R., Cunningham, A. M., DeFrees, S., Gao, Y., Watson, D. C., Young, N. M., and Wakarchuk, W. W. (1998) *Nat. Biotechnol.* **16,** 769–772
18. Gilbert, M., Watson, D. C., and Wakarchuk, W. (1997) *Biotech. Lett.* **19,** 417–420
19. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and. Sweet, R. M. (1993) *Nature* **362,** 219–223
20. Miller, R., Gallo, S. M., Khalak, H. G., and Weeks, C. M. (1994) *J. Appl. Cryst.* **27,** 613–621
21. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
22. Collaborative Computational Project, Number 4. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
23. McRee, D. E. (1993) *Practical Protein Crystallography*, Academic Press, San Diego, CA
24. Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50,** 157–163
25. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
26. Chan, A. W. E., Hutchinson, E. G., Harris, D., and Thornton, J. M. (1993) *Protein Sci.* **2,** 1574–1590
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26,** 283–291
28. Nicholls, A., Sharp, K., and Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11,** 281–296
29. Hart, T. N., Ness, S. R., and Read, R. J. (1997) *Proteins Struct. Funct. Genet. Suppl.* **1,** 205–209
30. Kraulis, P. K. (1991) *J. Appl. Cryst.* **24,** 946–950
31. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) *Structure* **5,** 1093–1108
32. Richardson, J. S., and Richardson, D. C. (1989) *Principles and Patterns of Protein Conformation*, Plenum Press, New York
33. Tullius, M. V., Vann, W. F., and Gibson, B. W. (1999) *Protein Sci.* **8,** 666–675
34. Petsko, G. A., Kenyon, G. L., Gerlt, J. A., Ringe, D., and Kozarich, J. W. (1993) *Trends Biochem. Sci.* **18,** 372–376
35. Schwabe, C. (1996) *FASEB. J.* **10,** 184–185
36. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
37. Schulz, G. E. (1992) *Faraday Discuss.* **93,** 85–93
38. Jelakovic, S., Jann, K., and Schulz, G. E. (1996) *FEBS Lett.* **391,** 157–161
39. Samuels, N. M., Gibson, B. W., and Miller, S. M. (1999) *Biochemistry* **38,** 6195–6203
40. Stoughton, D. M., Zapata, G., Picone, R., and Vann, W. F. (1999) *Biochem. J.* **343,** 397–402
41. Ambrose, M. E., Freese, S. J., Reinhold, M. S., Warner, T. G., and Vann, W. F. (1992) *Biochemistry* **31,** 775–780

Structure of a Sialic Acid-activating Synthetase, CMP-acylneuraminate Synthetase in the Presence and Absence of CDP

Steven C. Mosimann, Michel Gilbert, Dennise Dombroswki, Rebecca To, Warren Wakarchuk and Natalie C. J. Strynadka

J. Biol. Chem. 2001, 276:8190-8196.
doi: 10.1074/jbc.M007235200 originally published online December 11, 2000
