THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 42, Issue of October 15, pp. 44023–44029, 2004
Printed in U.S.A.

Molecular Structure of α-D-Glucose-1-phosphate Cytidylyltransferase from *Salmonella typhi**

Received for publication, July 9, 2004, and in revised form, August 3, 2004
Published, JBC Papers in Press, August 3, 2004, DOI 10.1074/jbc.M407755200

Nicole M. Koropatkin and Hazel M. Holden‡

From the Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

Dideoxysugars, which display biological activities ranging from mediating cell-cell interactions to serving as components in some antibiotics, are synthesized in various organisms via complex biochemical pathways that begin with the attachment of α-D-glucose 1-phosphate to either CTP or dTTP. Here we describe the three-dimensional structure of the α-D-glucose-1-phosphate cytidylyltransferase from *Salmonella typhi*, which catalyzes the first step in the production of CDP-tyvelose. For this investigation, the enzyme was crystallized in the presence of its product, CDP-glucose. In contrast to previous reports, the enzyme exists as a fully integrated hexamer with 32-point group symmetry. Each subunit displays a "bird-like" appearance with the "body" composed predominantly of a seven-stranded mixed β-sheet and the two "wings" formed by β-hairpin motifs. These two wings mediate subunit-subunit interactions along the 3-fold and 2-fold rotational axes, respectively. The six active sites of the hexamer are situated between the subunits related by the 2-fold rotational axes. CDP-glucose is anchored to the protein primarily by hydrogen bonds with backbone carbonyl oxygens and peptidic NH groups. The side chains of Arg^{111} and Asn^{188} from one subunit and Glu^{178} and Lys^{179} from the second subunit are also involved in hydrogen bonding with the ligand. The topology of the main core domain bears striking similarity to that observed for glucose-1-phosphate thymidylyltransferase and 4-diphosphocytidyl-2-*C*-methylerythritol synthetase.

Deoxysugars are ubiquitous in nature where they function in a variety of biological processes, including but not limited to cell adhesion, the immune response, the determination of ABO blood groups, fertilization, antibiotic function, and microbial pathogenicity (1–3). They differ from the more common carbohydrates such as glucose or mannose by the substitution of at least one hydroxyl group with hydrogen. As a consequence, these sugars have altered properties with respect to their more common counterparts such as enhanced chemical stability and hydrophobicity. Of special interest are the 3,6-dideoxyhexoses, which are found in the O-antigens of surface lipopolysaccharides of some Gram-negative bacteria such as *Yersinia pseudotuberculosis* and *Salmonella typhi* (4–7). Many of these deoxysugars, such as tyvelose, paratose, and abequose, have been demonstrated to be the main antigenic determinants and to contribute to the serological specificity of these bacteria (4, 6).

Recent years have, indeed, witnessed a renewed interest in the role of the 3,6-dideoxyhexoses in the establishment of infection in host organisms. In the pathogenic species of both *Yersinia* and *Salmonella,* the pathway leading to the production of 3,6-deoxysugars is initiated by glucose-1-phosphate cytidylyltransferase, which catalyzes the formation of CDP-glucose from α-D-glucose-1-phosphate and CTP (Scheme 1). The cytidylyltransferases from *Salmonella*, *Yersinia*, and *Azotobacter* species have been purified and shown to have typical subunit molecular weights of ~30,000 (8–14). Initial studies have suggested that the cytidylyltransferases function as tetramers (14).

Based on the study of the enzyme from *Salmonella enterica* strain LT2, the catalytic reaction is believed to proceed via a ping-pong mechanism whereby CTP binds, a CMP moiety is transferred to the enzyme, and pyrophosphate is released. Subsequently, CMP and glucose 1-phosphate are condensed to the final product, CDP-glucose (13). This proposed mechanism is in sharp contrast to that of the well studied α-D-glucose-1-phosphate thymidylyltransferases, which also function in the production of deoxysugars in organisms such as *Escherichia coli* and *Pseudomonas aeruginosa* (15, 16). In the case of these enzymes, catalysis is thought to proceed via a sequential ordered Bi Bi mechanism (17, 18). Numerous investigations have demonstrated that the cytidylyltransferases are subject to feedback inhibition by both the diphosphocytidyl deoxyhexose product and other intermediates generated along the biosynthetic pathway (8–10, 12).

Whereas several classes of enzymes catalyze the formation of diphosphonucleotidyl hexoses, the cytidylyltransferases are thought to be most closely related to the glucose-1-phosphate thymidylyltransferases. Both types of enzymes catalyze nearly identical reactions, albeit with different nucleotides, and both are regulated via feedback inhibition. Strikingly, however, the amino acid sequence similarities between the cytidylyltransferases and the thymidylyltransferases are quite low at ~15%. Here we describe the three-dimensional structure of the cytidylyltransferase from the pathogen *S. typhi* complexed with its product CDP-glucose and determined to a nominal resolution of 2.1 Å. This enzyme catalyzes the first step in the overall pathway leading to the production of CDP-D-tyvelose, which is found in the O-antigens of *S. typhi*. In contrast to the tetrameric quaternary structures of the glucose-1-phosphate thymidylyltransferases (17–20), the *S. typhi* cytidylyltransferase is a fully integrated hexamer displaying 32-point group symmetry with its active sites formed between the subunits related by the 2-fold rotational axes.

* This work was supported in part by National Institutes of Health Grant DK47814 (to H. M. H.). 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.

The atomic coordinates and structure factors (code 1TZF) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡ To whom correspondence should be addressed. Tel.: 608-262-4988; Fax: 608-262-1319; E-mail: Hazel_Holden@biochem.wisc.edu.

EXPERIMENTAL PROCEDURES

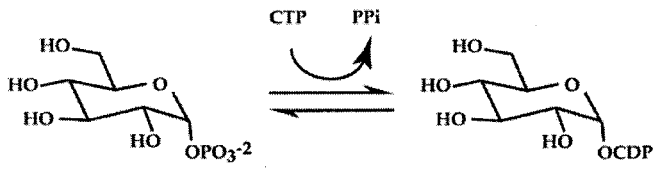
Cloning of the Glucose-1-phosphate Cytidylyltransferase Gene—The sequence for the gene encoding glucose-1-phosphate cytidylyltransferase, *rfbF*, has been reported and allowed for the design of primers for gene amplification (21). Genomic DNA for *S. typhi* isolates CDC numbers 87–2059 and 88–2009 was obtained as a generous gift from the laboratory of Dr. Stanley Maloy at the University of Illinois at Urbana-Champaign. The *rfbF* gene was PCR amplified from genomic DNA such that the forward primer 5′-CTA**GCTAGC**AAAGCGGTCATCCTGGCT-GGT-3′ included an NheI site, whereas the reverse primer 5′-CCC**AA-GCTTT**TTACTCCCAGGTTTTCCACGGAGC-3′ included a HindIII site for cloning into the expression vector pET-28a (Novagen). (The bold letters in the primer nucleotide sequence indicate the location of the restriction sites NheI and HindIII in the forward and reverse primers, respectively.) To create an NheI recognition site in the forward primer, the codons GCT (Ala) and AGC (Ser) were inserted after the starting methionine of the enzyme. By cloning into the NheI site of the pET-28a vector, a thrombin-cleavable His_6 tag was added to the N terminus of the protein. The *rfbF* gene was amplified by PCR with *Pfx* Platinum DNA polymerase (Invitrogen) according to the manufacturer's instructions and standard cycling conditions. The PCR product was purified with the QIAquick PCR purification kit (Qiagen, Inc.) and then digested for 16 h at 37 °C with both NheI and HindIII. The resulting gene fragment was gel purified with a QIAquick gel purification kit (Qiagen, Inc.) to remove digestion by-products. The purified fragment was subsequently ligated into the *E. coli* expression plasmid pET-28a (Novagen) that had been previously cut with the same restriction enzymes. The ligation mixture was used to transform *E. coli* DH5α cells, which were then plated onto LB media supplemented with 30 μg/ml of kanamycin. Individual colonies were selected, cultured overnight, and the plasmid DNA was extracted with a Qiaprep spin miniprep kit (Qiagen, Inc.). Plasmids were tested for incorporation of the *rfbF* gene by digestion with both NheI and HindIII. Positive clones were sequenced with the ABI Prism™ Big Dye Primer Cycle sequencing kit (Applied Biosystems, Inc.) to ensure that no mutations had been introduced during PCR amplification.

Protein Expression—E. coli Rosetta(DE3)pLysS cells were transformed with the pET28a-*rfbF* plasmid. Transformants were plated onto LB media supplemented with 30 μg/ml chloramphenicol and 30 μg/ml kanamycin. After ~16 h growth at 37 °C, the plates were scraped and the cells re-suspended in LB media to use for the inoculation of eight 2-liter baffled flasks containing 500 ml of TB media supplemented with kanamycin and chloramphenicol. Cultures were grown at 26 °C for 8 h, at which point the cultures had reached an A_{600} of ~0.4. At this time, the cultures were induced with a final concentration of 60 μM isopropyl-1-thio-β-D-galactopyranoside. The cells were allowed to grow overnight (16 h) at 26 °C before harvesting by centrifugation at 6000 × *g* for 8 min. The cell paste was frozen in liquid nitrogen and stored at −80 °C.

Expression of the Selenomethionine-labeled Protein—E. coli Rosetta(DE3)pLysS cells were transformed with the pET28a-*rfbF* plasmid and plated onto LB media supplemented with kanamycin as described above. The transformants were grown overnight (~16 h) at 37 °C after which several colonies were selected and used to inoculate 100 ml of M9 minimal media supplemented with 30 μg/ml of chloramphenicol and 30 μg/ml of kanamycin for growth overnight at 37 °C. Subsequently, 10 ml of this culture were used to inoculate six 2-liter baffled flasks each containing 500 ml of M9 minimal media supplemented with 5 μg/ml thiamine, 30 μg/ml chloramphenicol, and 30 μg/ml kanamycin. Cultures were grown at 37 °C to an A_{600} of ~0.45 before adjusting the temperature to 30 °C for the remainder of the growth period. At this time, each flask was supplemented with 50 mg each of L-lysine, L-threonine, L-phenylalanine, and 25 mg each of L-leucine, L-isoleucine, L-valine, and L-selenomethionine (22). After 20 min the cells were induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside and allowed to grow for 7 h. The cultures were then harvested by centrifugation at 6000 × *g* for 8 min, and the cell paste was frozen in liquid nitrogen.

Protein Purification—All protein purification steps were conducted at 4 °C unless otherwise noted. The purification scheme was the same for both the native and selenomethionine-substituted protein. Approximately 22 g of cell paste was thawed in 60 ml of Ni-NTA[1] lysis buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, and 10 mM imidazole, pH 8.0. Cells were lysed on ice by four cycles of sonication (30 s) separated by 3 min of cooling. The lysate was centrifuged at 4 °C for 40 min at 20,000 × *g* to remove cellular debris. The clarified lysate was loaded onto a 10-ml Ni-NTA-agarose column (Qiagen, Inc.) equilibrated previously with Ni-NTA lysis buffer. After loading, the column was washed with Ni-NTA wash buffer (50 mM NaH_2PO_4, 500 mM NaCl, and 20 mM imidazole, pH 8.0) until the OD_{280} of the flow through decreased to a stable baseline. The protein was eluted from the column with a gradient of 20 mM to 300 mM imidazole in Ni-NTA lysis buffer. Protein-containing fractions were pooled based on SDS-PAGE and dialyzed against 20 mM Tris-Cl, 500 mM NaCl, pH 8.4. The His_6 tag was removed by thrombin according to the manufacturer's instructions. The cleaved glucose-1-phosphate cytidylyltransferase was dialyzed overnight at 4 °C against buffer containing 50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, pH 8.0, and further purified from both thrombin and digestion products via Ni-NTA-agarose column chromatography (Qiagen, Inc). Fractions containing the thrombin-cleaved enzyme were pooled based on SDS-PAGE and dialyzed against 20 mM HEPPS, 500 mM NaCl, pH 8.0. The protein was concentrated to 9.5 mg/ml, and aliquots were quickly frozen in liquid nitrogen for storage at −80 °C. Analytical ultracentrifugation experiments confirmed the hexameric nature of the enzyme.[2]

Crystallization of Native and Selenomethionine-labeled Glucose-1-phosphate Cytidylyltransferase—Crystallization trials on the purified protein were conducted via hanging drop and utilized both an "in-house" designed sparse matrix screen of 144 conditions and the Crystal Screen (Hampton Research) at both 4 °C and room temperature. Prior to crystallization trials, CDP-glucose and $MgCl_2$ were added to the protein solution to final concentrations of 10 mM each. Single crystals were observed growing at room temperature from 25% poly(ethylene glycol) 4000, 200 mM ammonium sulfate, and 100 mM sodium acetate, pH 4.6. Refinement of the crystallization conditions led to large single crystals (both wild-type and selenomethionine-substituted protein) grown at 4 °C against precipitant solutions of 4–7% poly(ethylene glycol) 4000, 175 mM ammonium sulfate, and 100 mM sodium acetate, pH 5.0. The crystals achieved typical dimensions of ~0.5 × 0.5 × 0.1



SCHEME 1

[1] The abbreviations used are: Ni-NTA, nickel-nitrilotriacetic acid; HEPPS, 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid; CDP-ME, 4-diphosphocytidyl-2-*C*-methylerythritol.

[2] J. Kim and F. Raushel, personal communication.

TABLE I
X-ray data collection statistics

	Wavelength	Resolution	Number of independent reflections	Completeness	Redundancy	Avg I/Avg $\sigma(I)$	R_{sym}[a]
	Å	Å		%			%
Peak	0.97981	50–2.1	22262	97.9	9.8	46.8	8.8
		2.18–2.1[b]	2166	98.8	7.5	8.9	20.1
Inflection	0.97992	50–2.1	22212	97.5	9.4	42.4	8.0
		2.18–2.1	2156	97.6	6.9	6.9	23.0
Remote	0.96411	50–2.1	22186	97.4	8.9	37.2	6.8
		2.18–2.1	2149	97.4	6.1	5.5	26.9
Native	0.96410	50–2.1	22401	98.5	14.1	65.3	6.7
		2.18–2.1	2163	98.1	5.4	9.9	17.6

[a] $R_{sym} = (\Sigma|I - \bar{I}|\Sigma I) \times 100$.

[b] Statistics for the highest resolution bin.

mm in 2 weeks and belonged to the space group $P6_322$ with unit cell dimensions of $a = b = 88.4$ Å and $c = 162.8$ Å and one molecule per asymmetric unit.

High Resolution X-ray Data Collection—Both native and selenomethionine-substituted protein crystals were handled in similar manners. All synthetic mother liquor and cryoprotectant solutions contained 10 mM CDP-glucose and 10 mM $MgCl_2$. Crystals were harvested from the hanging drop experiments and soaked for 1–2 h in a synthetic mother liquor containing 8% poly(ethylene glycol) 4000, 175 mM ammonium sulfate, 300 mM NaCl, and 100 mM sodium acetate, pH 5.0. The crystals were then gradually transferred to a final cryoprotectant solution containing 25% poly(ethylene glycol) 4000, 175 mM ammonium sulfate, 500 mM NaCl, 12% ethylene glycol, and 100 mM sodium acetate, pH 5.0, and flash-cooled to −150 °C in a stream of liquid nitrogen for storage until synchrotron beam time became available. X-ray data sets from both the selenomethionine-substituted and the wild-type protein crystals were collected on a 3 × 3 tiled "SBC3" CCD detector at the Structural Biology Center 19-ID beam line (Advanced Photon Source, Argonne National Laboratory, Argonne, IL) to 2.1 Å resolution. The x-ray data were processed with HKL2000 and scaled with SCALEPACK (23). Relevant x-ray data collection statistics are presented in Table I.

X-ray Structural Analyses—The structure of the *S. typhi* glucose-1-phosphate cytidylyltransferase was solved via multiwavelength anomalous dispersion. The software package SOLVE (24) was utilized to determine and refine the positions of the 11 selenium atoms in the asymmetric unit. Solvent flattening with the program RESOLVE resulted in an interpretable electron density map calculated to 2.1 Å resolution (25). A model was built with TURBO (26) and subjected to alternate cycles of least squares refinement with the software package TNT (27) and manual model building. The *S. typhi* cytidylyltransferase contains 259 amino acid residues. The final model includes Met^1-Gly^{13}, Val^{22}-Glu^{259}, one CDP-glucose molecule, one Mg^{2+} ion (refined at 0.5 occupancy), and 161 waters. Refinement statistics are given in Table II. Electron density corresponding to the CDP-glucose is displayed in Fig. 1. The only significant outlier in the Ramachandran plot was Gln^{211}, which adopted ϕ, ψ values of −102° and −118°, respectively. The electron density for this residue was unambiguous. Gln^{211} lies at the start of an α-helix with its carboxamide side chain lying at the interface formed with a symmetry related molecule.

TABLE II
Least squares refinement statistics

Resolution limits (Å)	50–2.1
R-factor (overall) %/number of reflections[a]	19.9/22358
R-factor (working) %/number of reflections	19.7/20120
R-factor (free) %/number of reflections	24.8/2238
Number of Protein Atoms	2005
Number of Hetero-atoms	198
Average *B* values (Å^2)	
Protein atoms[b]	38.9
CDP-glucose	32.2
Solvents	50.2
Weighted root mean square deviations from ideality	
Bond lengths (Å)	0.010
Bond angles (deg)	2.2
Trigonal planes (Å)	0.004
General planes (Å)	0.009
Torsional angles (deg)[c]	18.2

[a] *R*-factor = $(\Sigma|F_o - F_c|/\Sigma|F_o|)\times 100$ where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.
[b] These include multiple conformations for Glu^{93}, His^{148}, and Gln^{169}.
[c] The torsional angles were not restrained during the refinement.

RESULTS AND DISCUSSION

Tertiary Structure of the Monomer—The glucose-1-phosphate cytidylyltransferase from *S. typhi* crystallized in the hexagonal space group $P6_322$ with one monomer per asymmetric unit. The subunit has overall molecular dimensions of ~60 Å × 50 Å × 60 Å and contains 14 strands of β-sheet and seven major α-helices. As displayed in Fig. 2*a*, the protein has a quite unusual and almost bird-like appearance with the "main body" dominated by a seven-stranded mixed β-sheet and the wings formed by β-hairpin motifs. The seven-stranded mixed sheet is formed by β1, β2, β5, β6, β8, β12, and β13 (delineated by Lys^4-Ala^9, Lys^{49}-Cys^{55}, Asn^{96}-Asp^{101}, Phe^{125}-Tyr^{129}, Ala^{152}-Thr^{157}, Phe^{192}-Asn^{195}, and Leu^{223}-His^{228}) with β12 running anti-parallel to the other strands. The first β-hairpin motif or wing is defined by strands β3 (Asp^{75}-His^{79}) and β4 (Arg^{84}-His^{88}), which are connected by a Type I turn (Met^{80}-Asn^{83}). In the second wing, the two strands β9 (Ala^{165}-Ile^{168}) and β10 (Val^{173}-Glu^{178}) are connected by a Type I′ turn (Gln^{169}-Gln^{172}).

Quaternary Structure—Previous gel filtration analyses of the *Y. pseudotuberculosis* cytidylyltransferase, which is 80% identical in amino acid sequence to the *Salmonella* enzyme, suggested that the protein exists as a tetramer, which is similar to the thymidylyltransferases (14). In this study, however, analytical ultracentrifugation experiments clearly demonstrated that the enzyme migrates as a hexamer in solution. Indeed, examination of the crystalline packing reveals that the hexamer with overall dimensions of ~100 Å × 90 Å × 80 Å packs along mutually perpendicular 2-fold and 3-fold crystallographic axes such that the protein exhibits 32-point group symmetry. Subunit-subunit interactions along the 3-fold axis are mediated primarily by the first β-hairpin motif (Fig. 2*b*). As a result of this interaction, the seven-stranded β-sheet in the main body of the protein is extended by two anti-parallel strands with β3 of one subunit running anti-parallel to β5 in a 3-fold related monomer. The side chain of Met^{80}, which initiates the Type I turn connecting β-strands 3 and 4, forms a stacking interaction with Tyr^{69} in the neighboring subunit and projects into a hydrophobic pocket formed by Met^{40}, Phe^{66}, Phe^{70}, Trp^{95}, and Val^{97} (second subunit). The total surface area buried between two adjacent monomers in the 3-fold interface is ~1800 Å^2 based on a probe radius of 1.4 Å (28). The second symmetrical interface in the hexamer occurs along the crystallographic 2-fold axes as shown in Fig. 2*c* for one pair. Here the interface is formed by the exchange of the second β-hairpin motif (β9 and β10) across the dyad. This subunit-subunit in-

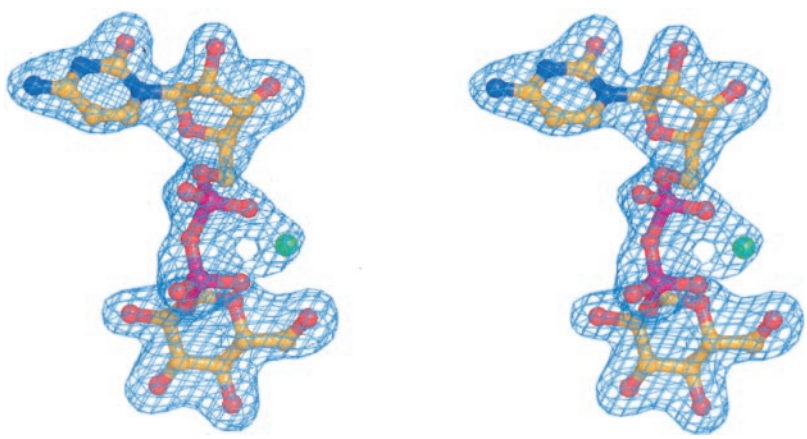


FIG. 1. **Electron density corresponding to the CDP-glucose moiety.** The map shown was calculated with coefficients of the form $(F_o - F_c)$, where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude from the model lacking the coordinates for the ligand. The map was contoured at 4 σ.

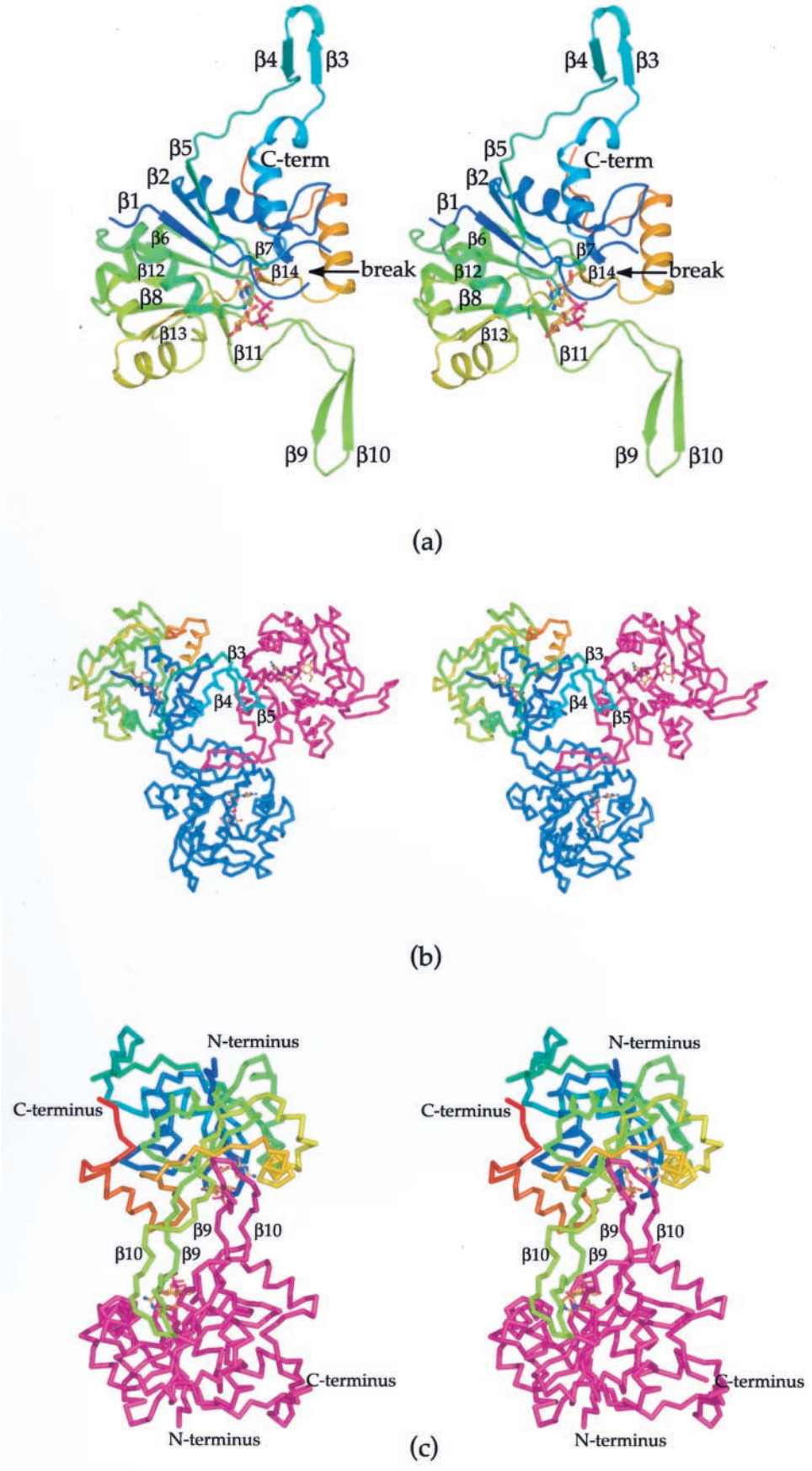


FIG. 2. **The structure of glucose-1-phosphate cytidylyltransferase.** The cytidylyltransferase from *S. typhi* is a hexamer that displays 32-point group symmetry. Shown in *a* is a ribbon representation of one subunit of the hexamer color ramped (*blue* to *yellow*) from the N terminus to the C terminus. The CDP-glucose ligand is displayed in a *ball-and-stick* representation. A view of the hexamer down the 3-fold rotational axis is given in *b*. For clarity, only three of the six subunits are shown. One subunit is color coded as in *a*, whereas the second and third are depicted in *magenta* and *blue*. The β-strands involved in subunit-subunit interactions are labeled β3, β4, and β5. The view presented in *c* is down one of the 2-fold rotational axes. For clarity, only two subunits are shown.

terface is much more extensive with a total buried surface area of ~4000 Å^2 between two monomers.

The Active Site—Close-up views of the active site are depicted in Fig. 3, *a* and *b*. The ligand lies across the mixed β-sheet of the core near β-strands 1 and 6. The cytosine base adopts the *anti*-conformation whereas the ribose assumes the

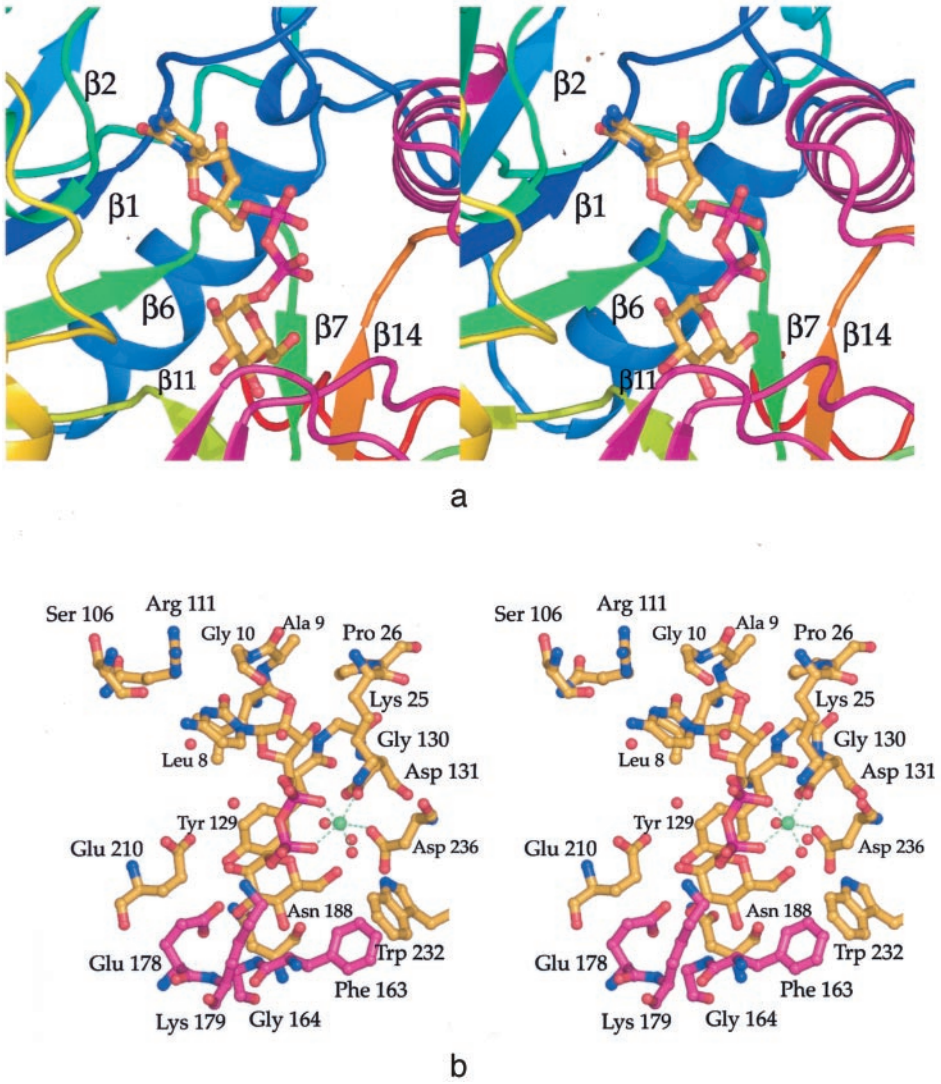


FIG. 3. **The active site of glucose-1-phosphate cytidylyltransferase.** Shown in *a* is the location of the bound CDP-glucose ligand in context with the tertiary structure of the protein. The color coding for the ribbon representation is as described in Fig. 2. Note that the ribbon representation depicted in *magenta* corresponds to the second subunit forming the active site. Those amino acid residues located within ~3.6 Å of the CDP-glucose are displayed in *b*. For clarity, Gly^{11} was omitted from the figure. Residues corresponding to the second subunit forming the active site are displayed in *magenta*.

$C_{2'}$-*endo* pucker. In sharp contrast to the active sites of the glucose-1-phosphate thymidylyltransferases, in cytidylyltransferase the active site is formed by residues contributed from two subunits, specifically those related by the crystallographic dyads. The amino group of the cytosine ring is hydrogen bonded to the carbonyl oxygen of Ser^{106} (3.1 Å) and a well ordered water molecule (2.9 Å), whereas the carbonyl group of the ring lies within 2.8 Å of the backbone peptidic nitrogen of Gly^{10}. The side chain N^{ϵ} of Arg^{111} is positioned at 3.0 Å from N-3 of the base. With respect to the ribosyl group of the ligand, O-2 lies within 2.7 Å of the carbonyl oxygen of Gly^{10} whereas O-3 sits within 2.9 Å and 3.0 Å of the backbone peptidic nitrogen of Gly^{130} and the backbone carbonyl oxygen of Leu^{8}, respectively. There is a partially occupied Mg^{2+} ion that is octahedrally coordinated by two solvents, $O^{\delta 2}$ of Asp^{131}, $O^{\delta 1}$ of Asp^{236}, and two phosphoryl oxygens. Ligand-metal bond distances range from 1.9 Å to 2.4 Å. An electrostatic interaction occurs between one of the β-phosphoryl oxygens of the ligand and the ϵ-amino group of Lys^{179} in the second subunit (3.1 Å).

Whereas the interactions that anchor the nucleotide base and the ribose of CDP-glucose to the protein are provided by the first subunit, the hexose moiety of the ligand is wedged into the active site such that O-2 and O-3 are bridged by the carboxylate group of Glu^{178} in the second subunit (3.1 Å and 3.0 Å, respectively). Additionally, $N^{\delta 2}$ of Asn^{188} (first subunit) sits at 2.9 Å from O-3. The C-4 hydroxyl oxygen is positioned at 2.6 Å from the carbonyl group of Asn^{188} (first subunit) and 3.1 Å from the backbone peptidic nitrogen of Gly^{164} (second subunit). Finally, the C-6 hydroxyl group points toward the magnesium ion and sits at 2.7 Å from a well ordered water molecule. As shown in Fig. 3*b*, the aromatic side chain of Tyr^{129} abuts one side of the hexose moiety.

Comparison with Glucose-1-phosphate Thymidylyltransferase—Given the similar biochemical roles and reactions catalyzed by the cytidylyltransferases and thymidylyltransferases, it was anticipated that the molecular architectures of these enzymes might be similar. Strikingly, however, an initial alignment of their primary sequences in the absence of the structural data presented here revealed little significant homology beyond the first 60 residues. Shown in Fig. 4*a* is a three-dimensional alignment of the individual subunits of the cytidylyltransferase and the *E. coli* thymidylyltransferase (Protein Data Bank accession code 1H5T) (17). The models superimpose with a root mean square deviation of 2.1 Å for 185 structurally equivalent α-carbons. Both proteins have a similar core structure of a seven-stranded mixed β-sheet. The first major deviation between the models for these two enzymes occurs at Phe^{66} in the cytidylyltransferase and Leu^{68} in the thymidylyltransferase. Rather than folding into a β-hairpin wing, the thymidylyltransferase contains a simple loop of five amino acid residues. The two polypeptide chains correspond again at Glu^{93} and Trp^{74} in the cytidylyltransferase and thymidylyltransferase, respectively. The second major change occurs at Gly^{161} in the cytidylyltransferase and Asn^{141} in the

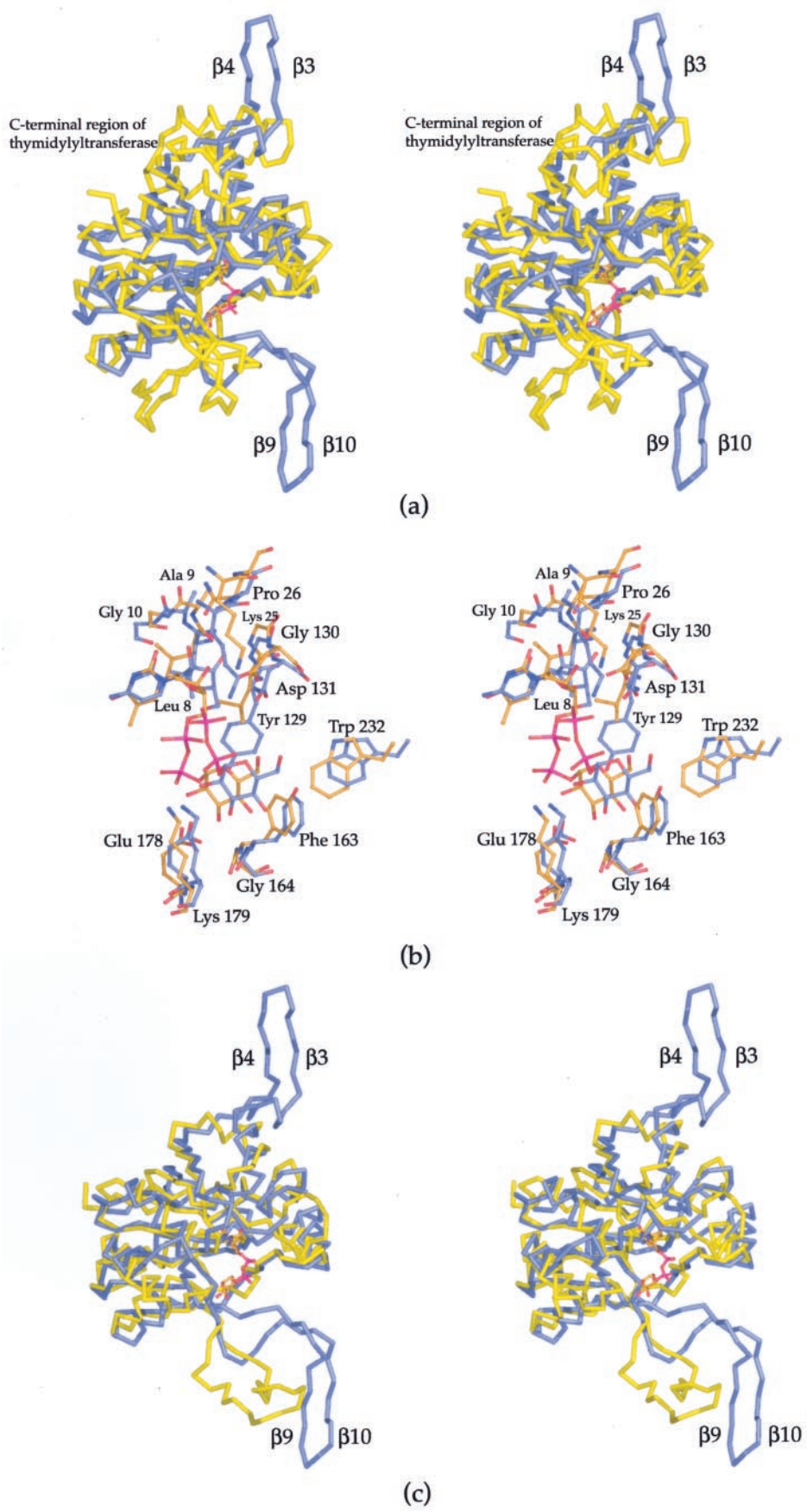


FIG. 4. **Comparison with other nucleotidylyltransferases.** A superposition of one subunit of the cytidylyltransferase (*blue*) onto the *E. coli* glucose-1-phosphate thymidylyltransferase (*yellow*) is given in *a*. X-ray coordinates for the *E. coli* thymidylyltransferase were obtained from the Protein Data Bank (accession number 1H5T). The bound ligand, shown in a *ball-and-stick* representation, corresponds to the CDP-glucose observed in the cytidylyltransferase. The active site regions for these two enzymes were superimposed as shown in *b*. The cytidylyltransferase and thymidylyltransferase are displayed in *blue* and *gold*, respectively. The numbering corresponds to that observed in the cytidylyltransferase. A superposition of the α-carbon traces for cytidylyltransferase (*blue*) and CDP-ME synthetase (*yellow*) is given in *c*. X-ray coordinates for CDP-ME synthetase were obtained from the Protein Data Bank (accession number 1INI).

thymidylyltransferase. Interestingly, whereas the thymidylyltransferase also contains a β-hairpin structure, instead of crossing the dimer interface as in the cytidylyltransferase it merely folds back upon itself, laying in a hydrophobic pocket formed by side chains from an α-helix and three β-strands within the same monomer. Because of the change in the direc-

tion and placement of the β-hairpin motif in the thymidylyltransferase, the monomer has a more spherical appearance and contacts it neighbors differently thereby contributing to the difference in quaternary structures exhibited by these two classes of nucleotidyltransferases. Additionally, these differences result in the active site of the thymidylyltransferase being completely contained within one monomer, whereas in the cytidylyltransferase it is wedged between two subunits.

One of the more structurally conserved regions between these two enzymes is the immediate vicinity surrounding the CDP-glucose ligand. An overlay of this area for the two proteins is displayed in Fig. 4*b*. Note that Glu^{178} and Lys^{179} in the cytidylyltransferase are contributed by the second subunit, whereas the structurally equivalent residues in the thymidylyltransferase (Glu^{162} and Lys^{163}) are contained within one monomer. Another notable difference is the replacement of Tyr^{129} in the cytidylyltransferase for Leu^{109} in the thymidylyltransferase.

The cytidylyltransferases typically contain ~260 amino acids, whereas the thymidylyltransferases are larger with ~300 residues. The last 30 residues of the *E. coli* thymidylyltransferase fold into three small α-helices, which pack together and occupy an area close to the placement of the first β-hairpin wing in the cytidylyltransferase (Fig. 4*a*). The packing of these helices against the core of the thymidylyltransferase creates a second shallow diphosphonucleotide binding site, which is apparently absent in the cytidylyltransferase. In addition, this auxillary binding site is positioned at one of the main subunit-subunit interfaces of the thymidylyltransferase and therefore has been speculated to be a regulatory site (17–19). The nucleotidyltransferases, regardless of their nucleotide preference, are subject to feedback inhibition and are thought to serve as control points for regulation of the deoxysugar biosynthetic pathways. Previous reports on the cytidylyltransferases from a number of sources have demonstrated that these enzymes are inhibited by a variety of CDP-deoxyhexoses (8, 9, 12). The observed inhibition is only partially reversible by increasing amounts of CTP, suggesting that the CDP-deoxyhexoses may bind to a second regulatory site on the enzyme (8, 9). The location of this putative site on the *S. typhi* cytidylyltransferase is not apparent from the present study.

A search with the DALI server (29) revealed that the *S. typhi* cytidylyltransferase also demonstrates molecular similarity to 4-diphosphocytidyl-2-*C*-methylerythritol (CDP-ME) synthetase from *E. coli* (30). This enzyme, which catalyzes the formation of CDP-ME from CTP and 2-*C*-D-methyl-erythritol 4-phosphate, is involved in the mevalonate-independent pathway for isoprenoid biosynthesis in a number of different organisms. Shown in Fig. 4*c* is a superposition of a monomer of CDP-ME synthetase (Protein Data Bank accession code 1INI) onto the cytidylyltransferase subunit. These two models correspond with a root mean square deviation of 2.2 Å for 173 structurally equivalent α-carbon atoms. Like the cytidylyltransferase and the thymidylyltransferase, CDP-ME synthetase contains the same α/β platform. CDP-ME synthetase also contains a β-hairpin motif like the second wing in the cytidylyltransferase that is involved in subunit-subunit interactions, although the manner in which it mediates dimer formation is quite different.

Catalytic Mechanism—Catalysis by glucose-1-phosphate cytidylyltransferase was originally proposed to occur via a ping-pong Bi Bi mechanism, whereby in the first half CTP binds and leads to the formation of a CMP-enzyme complex and subsequent release of pyrophosphate (13). The second half was proposed to occur via a nucleophilic attack of the glucose 1-phosphate phosphoryl oxygen on the CMP-enzyme intermediate. Several independent kinetic and x-ray crystallographic studies on the thymidylyltransferases, however, have supported a sequential ordered Bi Bi catalytic mechanism (17–19). Accordingly, for the thymidylyltransferases, the mechanism is believed to proceed by direct nucleophilic attack of the glucose 1-phosphate phosphoryl oxygen on the α-phosphorus of the nucleotide resulting in the products pyrophosphate and dTDP-glucose. Given the similarities between the active sites for both of these enzymes, it is intriguing that the cytidylyltransferase reaction mechanism is reported to be ping-pong. In the enzyme/complex described here, the only potential base near the α-phosphorus is Lys^{25} for which the electron density is quite weak following C-γ of its side chain. This lysine is conserved in the *E. coli* thymidylyltransferase as Lys^{23} with its ϵ-nitrogen positioned at 3.9 Å from the α-phosphorus of dTTP (according to the x-ray coordinates deposited under Protein Data Bank accession code 1MC3). A detailed kinetic analysis of the cytidylyltransferase from *S. typhi* is presently underway to further probe its catalytic mechanism.

Acknowledgments—We thank W. W. Cleland for carefully reading this manuscript and Dr. James B. Thoden for helpful advice throughout the course of this investigation. Use of the Argonne National Laboratory Structural Biology Center beam lines at the Advanced Photon Source was supported by the United States Department of Energy, Office of Biological and Environmental Research, under Contract Number W-31-109-ENG-48.

REFERENCES

1. Hanessian, S. (1966) *Adv. Carbohydr. Chem. Biochem.* **21,** 143–207
2. Williams, N. R., and Wander, J. D. (1980) in *The Carbohydrates*: *Chemistry and Biochemistry* (Pigman, W., and Horton, D., eds) Vol. 1b, Academic Press, New York
3. Varki, A. (1993) *Glycobiology* **3,** 97–130
4. Luderitz, O., Staub, A. M., and Westphal, O. (1966) *Bacteriol. Rev.* **30,** 192–255
5. Ashwell, G., and Hackman, J. (1971) in *Microbial Toxins* (Weinbaum, G., Kodis, S., and Aji, S. A., eds) Vol. 4, pp. 235–266, Academic Press, New York
6. Bishop, C. T., and Jennings, H. J. (1982) in *The Polysaccharides* (Aspinall, G. O., ed) Vol. 1, pp. 291–330, Academic Press, New York
7. Lindberg, B. (1990) *Adv. Carbohydr. Chem. Biochem.* **48,** 279–318
8. Mayer, R. M., and Ginsburg, V. (1965) *J. Biol. Chem.* **240,** 1900–1904
9. Kimata, K., and Suzuki, S. (1966) *J. Biol. Chem.* **241,** 1099–1113
10. Nikaido, H., and Nikaido, K. (1966) *J. Biol. Chem.* **241,** 1376–1385
11. Matsuhashi, S., Matsuhashi, M., and Strominger, J. L. (1966) *J. Biol. Chem.* **241,** 4267–4274
12. Rubenstein, P. A., and Strominger, J. L. (1974) *J. Biol. Chem.* **249,** 3789–3796
13. Lindqvist, L., Kaiser, R., Reeves, P. R., and Lindberg, A. A. (1994) *J. Biol. Chem.* **269,** 122–126
14. Thorson, J. S., Kelly, T. M., and Liu, H. W. (1994) *J. Bacteriol.* **176,** 1840–1849
15. Liu, H. W., and Thorson, J. S. (1994) *Annu. Rev. Microbiol.* **48,** 223–256
16. He, X. M., and Liu, H. W. (2002) *Annu. Rev. Biochem.* **71,** 701–754
17. Zuccotti, S., Zanardi, D., Rosano, C., Sturla, L., Tonetti, M., and Bolognesi, M. (2001) *J. Mol. Biol.* **313,** 831–843
18. Barton, W. A., Lesniak, J., Biggins, J. B., Jeffrey, P. D., Jiang, J., Rajashankar, K. R., Thorson, J. S., and Nikolov, D. B. (2001) *Nat. Struct. Biol.* **8,** 545–551
19. Blankenfeldt, W., Asuncion, M., Lam, J. S., and Naismith, J. H. (2000) *EMBO J.* **19,** 6652–6663
20. Sivaraman, J., Sauve, V., Matte, A., and Cygler, M. (2002) *J. Biol. Chem.* **277,** 44214–44219
21. Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D., Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T. T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T. S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrell, B. G. (2001) *Nature* **413,** 848–852
22. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124
23. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
24. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
25. Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 965–972
26. Roussel, A., and Cambillau, C. (1998) in *Silicon Graphics Geometry Partners Directory* (Graphics, S., ed) pp. 77–79, Silicon Graphics, Mountain View, CA
27. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. Sect. A* **43,** 489–501
28. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400
29. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
30. Richard, S. B., Bowman, M. E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A. M., Cane, D. E., and Noel, J. P. (2001) *Nat. Struct. Biol.* **8,** 641–648