Structure of the Nucleotide-Diphospho-Sugar Transferase, SpsA from *Bacillus subtilis*, in Native and Nucleotide-Complexed Forms[†,‡]

Simon J. Charnock and Gideon J. Davies*

Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York Y010 5DD, U.K.

Received February 4, 1999; Revised Manuscript Received March 22, 1999

ABSTRACT: The enzymatic formation of glycosidic bonds may be catalyzed by the transfer of the glycosyl moiety from an activated nucleotide-diphospho-sugar donor to a specific acceptor. SpsA is a glycosyltransferase implicated in the synthesis of the spore coat of *Bacillus subtilis*, whose homologues include cellulose synthase and many lipopolysaccharide and bacterial *O*-antigen synthases. The three-dimensional crystal structure of SpsA has been determined by conventional MIR techniques at a resolution of 1.5 Å. It is a two-domain protein with a nucleotide-binding domain together with an acceptor binding domain which features a disordered loop spanning the active site. The structures of SpsA in complex with both Mg−UDP and Mn−UDP have also been determined at 2.0 and 1.7 Å, respectively. These complexes, together with the sequence conservation, begin to shed light on the mechanism of this ubiquitous family of inverting glycosyltransferases.

The enzymatic formation of glycosidic bonds is, simply in terms of quantity, the most important reaction on earth. Glycosyltransferases are involved not only in biomass synthesis but also in many of the intricate details of cellular biochemistry such as protein and lipid glycosylation. They have attracted interest not only as drug targets but also because of their potential applications in the chemoenzymatic synthesis of oligosaccharides (*1*, *2*). But despite this synthetic utility and unquestionable medical importance, little is known about their structure or mechanism. The large body of information that is available for the DNA sequences encoding these enzymes is unfortunately accompanied by a severe lack of knowledge about the donor and acceptor specificities for the majority of the gene products. The derived amino acid sequences for those glycosyltransferases utilizing nucleotide-diphospho-sugars have, however, recently been classified into a number of sequence similarity-based families (*3*). Within each family, enzymes may exhibit a broad spectrum of substrate specificities, but it is likely that, by analogy with the glycoside hydrolases (*4*, *5*), related enzymes will share the same catalytic mechanism leading to either retention or inversion of anomeric configuration (*6*, *7*).

Members of glycosyltransferase family 2 are expected to use an inversion mechanism. It is one of the largest families (*3*) whose members range from the polysaccharide synthases responsible for the production of the world's most abundant polymers, such as cellulose and chitin, to enzymes involved in bacterial cell surface glycosylation. Family 2 enzymes are of medical importance; bacterial *O*-antigen synthases are potential drug targets and may be used for the chemoenzymatic production of vaccines (reviewed in ref *1*). Furthermore, mutations at their loci are implicated in vaccine resistance. Family 2 members are expected to use nucleotide-diphospho-α-D-sugars to generate β-linked products (Figure 1). The reaction mechanism is widely believed to be S_N2 with an oxocarbenium ion-like character to the transition state. Such a mechanism would require a general base, for activation of the nucleophilic acceptor species by deprotonation, together with some means of facilitating departure of the nucleotide-diphospho leaving group with divalent metal ions, particularly Mg^{2+} and Mn^{2+}, implicated (*7*). To date, the only structure for a nucleotide-diphospho-sugar transferase is the DNA-modifying β-glucosyltransferase from bacteriophage T4 (*8*). It exhibits a bilobed structure with two nucleotide-binding domains, one involved in the binding of the UDP-glucose donor and the other, presumably, the DNA acceptor species. The enzyme bares little sequence similarity with any other known sequence.

The *spsA* gene encodes a glycosyltransferase implicated in the production of the mature spore coat during the spore response of *Bacillus subtilis* (*9*, *10*). It encodes a 256-amino acid protein whose substrate specificity is undefined but whose sequence has been classified into family 2. Expression of *spsA* is under control of a σ^K-dependent promoter activated by the GerE protein (*9*). SpsA may be involved in peptidoglycan biosynthesis, but the most closely related sequences exhibit an extremely diverse array of activities with enzymes involved in teichoic and colanic acid synthesis, galactosyl transfer, and lipopolysaccharide production. Here we present the structure of SpsA at 1.5 Å resolution, in native and nucleotide-complexed forms, in the presence of both manganese and magnesium. The structures reveal the prototypical

[†] This work was funded, in part, by the Biotechnology and Biological Sciences Research Council and the University of York. Data collection was supported by the EU TMR/LSF programme. G.J.D. is a Royal Society University Research Fellow.

[‡] Coordinates for the structures described in this paper have been deposited with the Protein Data Bank (file names 1qg8, 2qg8, and 3qg8).

* Corresponding author. Telephone: 44-1904-432596. Fax: 44-1904 410519. E-mail: davies@yorvic.york.ac.uk.

10.1021/bi990270y CCC: $18.00
Published on Web 04/30/1999

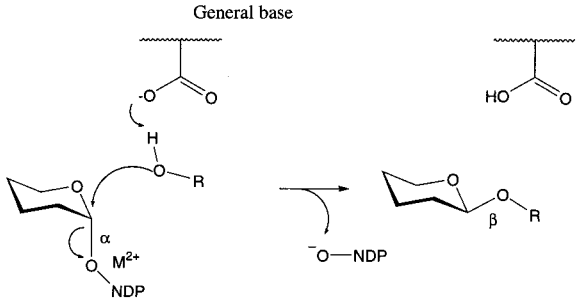

FIGURE 1: Likely catalytic mechanism for an inverting NDP-sugar transferase. An α-linked nucleotide-diphospho-sugar donor yields a product with β-configuration. The sugar acceptor is activated for nucleophilic attack by a general base. It is widely believed that a divalent metal ion (M^{2+}) assists leaving group departure.

framework for family 2 glycosyltransferases, while the UDP[1] complexes reveal the role of the invariant residues and provide insights into the likely catalytic mechanism.

EXPERIMENTAL PROCEDURES

Crystallization, Data Collection, and Refinement. Crystals of the native form of SpsA were grown as described previously (*11*). They exist in space group $C222_1$ and contain a single molecule in the asymmetric unit with the following cell dimensions: $a = 42.4$ Å, $b = 142.0$ Å, and $c = 81.4$ Å. A single crystal was transferred to a cryoprotectant stabilizing solution, mounted in a rayon-fiber loop, and immediately placed in a stream of N_2 gas at 100 K. Data (100°) were collected with an oscillation angle of 0.5°/image on beamline PX 9.6 ($\lambda = 0.870$ Å) at the Daresbury SRS. An ADSC CCD device was used as a detector. Data were processed with MOSFLM (A. Leslie, Medical Research Council Laboratory for Molecular Biology, Cambridge, U.K.) and scaled and reduced with SCALA from the CCP4 suite of programs (*12*). The X-ray diffraction data are summarized in Table 1.

Data for three derivatives, methyl mercury chloride (MeHgCl), ethyl mercury thiosalicylate (ETMS), and platinum terpyridine chloride (Pt), were collected on beamline PX 7.2 ($\lambda = 1.488$ Å) at the Daresbury SRS using a MAR research image plate as a detector. All data were processed with the HKL suite (*13*, *14*). Data for the two Hg derivatives extended to 2.25 Å and those for the Pt derivative to 3.3 Å. The derivative data were high-quality, with anomalous completeness in excess of 97% and an R_{merge} of <0.06 for all three derivatives (statistics not shown). Initial heavy metal positions were determined from manual inspection of the appropriate Pattersons which revealed the same single site for each of the two Hg derivatives. Phasing was performed using SHARP (*15*) and a single Pt site found by cross-phase Fourier transformation. Final phasing using all three derivatives gave an overall figure of merit of 0.59 for all data between 20 and 2.3 Å resolution.

Phase Extension, Model Building, and Refinement. MIR phases, to 2.3 Å resolution, were extended to 1.5 Å using solvent flattening and histogram matching in DM (*16*). At this point, the electron density map was of reasonably high quality and permitted identification of many elements of secondary structure and amino acid side chains. Main and side chain bones atoms were calculated using the X-FIT options of QUANTA (Molecular Simulations Inc.). Bones atoms were not, initially, used as an aid in model building but instead written out as a set of "seed" atoms for phase improvement. Initially, 801 seed atoms were treated as dummy waters and fed through 64 cycles of REFMAC (*17*). Five percent of the observations were set aside for cross validation analysis (*18*). ARP (*19*) was used to add between 30 and 50 water molecules per cycle in "dummy atom" mode where waters are permitted to lie within covalent bonding distance of adjacent atoms. During each refinement cycle, MIR phases were included as restraints. This iterative phase improvement procedure converged with 2021 seed atoms and an R_{cryst}/R_{free} of 0.25/0.34 for all data between 20 and 1.5 Å resolution. The electron density map calculated with the combined MIR/seed atom phases was of exceptionally high quality, and model building with X-FIT proved to be trivial. The structure was refined using maximum likelihood techniques with REFMAC with MIR phases included as additional restraints for the initial cycles. The behavior of R_{free} suggested that both inclusion of calculated hydrogen scattering from "riding" hydrogen atoms and the use of anisotropic modeling for the atomic displacement parameters were appropriate.

Nucleotide and Mn Complexes. Data were collected for crystals of SpsA soaked for 18 h in solutions of various nucleotide diphosphates (ADP, CDP, GTP, UDP, UDP-*N*-acetylglucosamine, and dTDP). Data were collected for crystals soaked in either 1 mM UDP, collected on beamline PX 7.2 at the Daresbury SRS to 2.0 Å resolution, or 10 mM UDP, collected on beamline BW7A at the EMBL Hamburg outstation to 2.25 Å resolution. The Mn−UDP complex was obtained by soaking crystals, for 1 h, in a solution containing 50 mM $MnCl_2$ and 10 mM UDP. Data were collected at the ESRF (Grenoble, France) on beamline ID14-4 using an ADSC Quantum 4 CCD instrument as a detector. Data for other nucleotide complexes were collected at the EMBL Hamburg outstation at the DESY synchrotron site. The complex structures were refined as described for the native enzyme, with the exception that riding hydrogen scattering and anisotropic treatment of the atomic displacement parameters were not justified by the R_{free} values and were not used.

RESULTS AND DISCUSSION

Description of the SpsA Structure. The structure of SpsA was determined by MIR with the inclusion of a novel phase extension procedure in which the initial skeletonization of the electron density map was used to generate starting seed atoms for phase extension (see Experimental Procedures). The structure determination proceeded smoothly, with the exception that two regions of the structure remain undefined in the electron density. These two disordered loops are a short turn between two β-strands, residues 134−136, and a long loop (residues 218−231), which must span the substrate-binding cleft. The final native SpsA structure has a crystallographic *R* value of 0.16 with an R_{free} value of 0.21 (Table 1).

SpsA is a two-domain molecule with an overall size of approximately 45 Å × 40 Å × 36 Å. The N-terminal domain

[1] Abbreviations: HPRT, hypoxanthine phophoribosyltransferase; MIR, multiple isomorphous replacement; SRS, synchrotron radiation source; UDP, uridine 5′-diphosphate.

Table 1: Data and Model Quality Statistics for SpsA from *B. subtilis*

	native (Mg)	UDP−Mn	UDP−Mg
diffraction data			
resolution (Å) (outer shell)	25−1.5 (1.6−1.5)	18−1.75 (1.81−1.75)	20−2.0 (2.02−2.0)
R_{merge}[a]	0.037 (0.27)	0.062 (0.37)	0.058 (0.32)
completeness (%)	97 (88)	99 (100)	96 (90)
$I/\sigma I$	22.5 (3.5)	23.0 (3.6)	18.3 (3.6)
multiplicity	3.7 (3.1)	5.5 (5.4)	3.7 (3.5)
radiation source	SRS, Daresbury, PX 9.6	ESRF, Grenoble, ID14-4	SRS, Daresbury, PX 7.2[b]
refinement			
R_{cryst}	0.16	0.17	0.18
R_{free}	0.21	0.23	0.24
rms deviation of 1−2 bonds (Å)	0.012	0.011	0.016
rms deviation 1−3 angle distances (Å)	0.032	0.028	0.037

[a] $R_{merge} = \sum_{hkl}\sum_i|I_{hkl} - \langle I_{hkl}\rangle|/\sum_{hkl}\sum_i\langle I_{hkl}\rangle$. [b] Data for this complex were also collected at the EMBL Hamburg outstation, to 2.2 Å resolution.

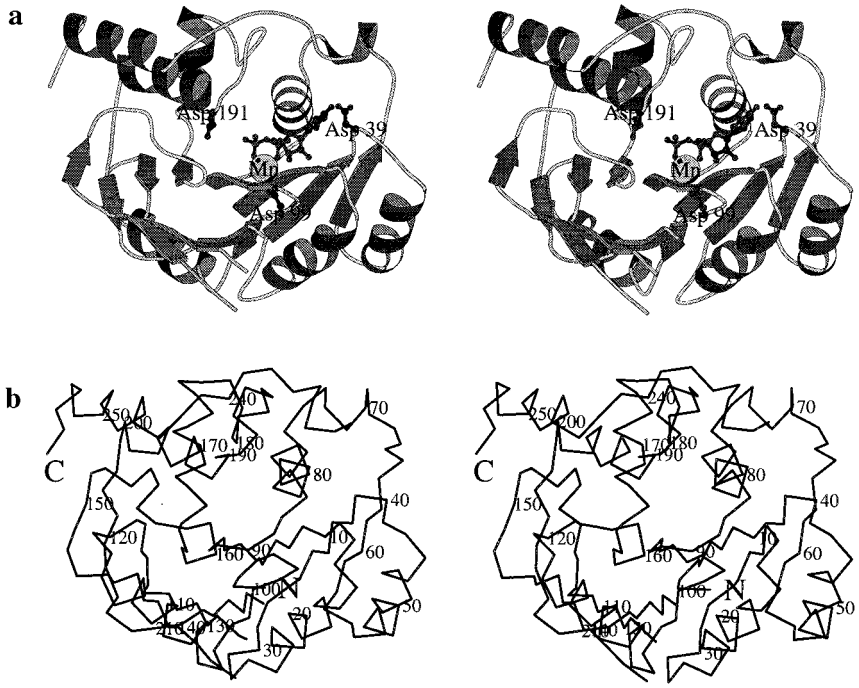

FIGURE 2: Structure of SpsA in complex with Mn−UDP. (a) In the divergent stereo cartoon representation, the UDP, Asp 39, Asp 99, and Asp 191 are depicted in ball-and-stick representations and the manganese ion is depicted as a shaded sphere. (b) Divergent stereo C_α trace of SpsA with every 10th residue labeled. This figure was drawn with MOLSCRIPT (*31*).

(residues 2−100) provides a nucleotide-binding fold consisting of four parallel β-strands flanked on either side by two α-helices (Figure 2). The remainder of the structure, residues 101−256, is composed of a mixed β-sheet flanked by three helices on one side and one on the other. The outer strand forms the boundary of a dimer formed by one of the crystallographic 2-fold axes. SpsA is a monomer in solution, although it exhibits non-ideal behavior after ultracentrigutation under extreme conditions of pH and buffer (data not shown). The C-terminal domain exhibits an open groove which is likely to be the binding site for the acceptor species. The disordered loop, residues 218−231, spans this groove, and one may speculate that loop closure in the ternary complex may provide the means for preventing abortive hydrolysis of the nucleotide-diphospho donor, as has been observed with mechanistically related enzymes such as the phosphoribosyltransferases (*20*). The structure of the nucleotide-binding domain of SpsA is very similar to others of that ilk, the most similar being those from pyruvate kinase and the 5′−3′ exonuclease domain of *Thermus aquaticus* DNA polymerase [calculated using DALI (*21*)]. To date, the only other available structure for a nucleotide-diphospho-sugar transferase is the DNA-modifying β-glucosyltransferase from bacteriophage T4 (*8*). The topology of the nucleotide-binding domain of SpsA is somewhat similar to that of the T4 enzyme, but this similarity does not extend to the mode of nucleotide binding, the conformation of the nucleotide, or the presentation of potential catalytic residues in the vicinity of the distal phosphate. The native enzyme structure, in the presence of 200 mM $MgCl_2$, also revealed a hexacoordinate magnesium ion, in the active site, coordinated via five solvent water molecules and a cryoprotectant glycerol molecule (which was essential for data collection from the fragile crystals). It is not clear whether the magnesium has a biological role, whether its position mimics that of the positivly charged oxocarbenium ion found during catalysis [as is often the case with divalent metals in the active sites of glycoside hydrolases (*22*, *23*)], or whether it merely binds in an adventitious fashion.

Nucleotide Binding and Sequence Conservation. As is the case with the majority of family 2 glycosyltransferases, the exact nucleotide-diphospho-sugar specificity of SpsA is not

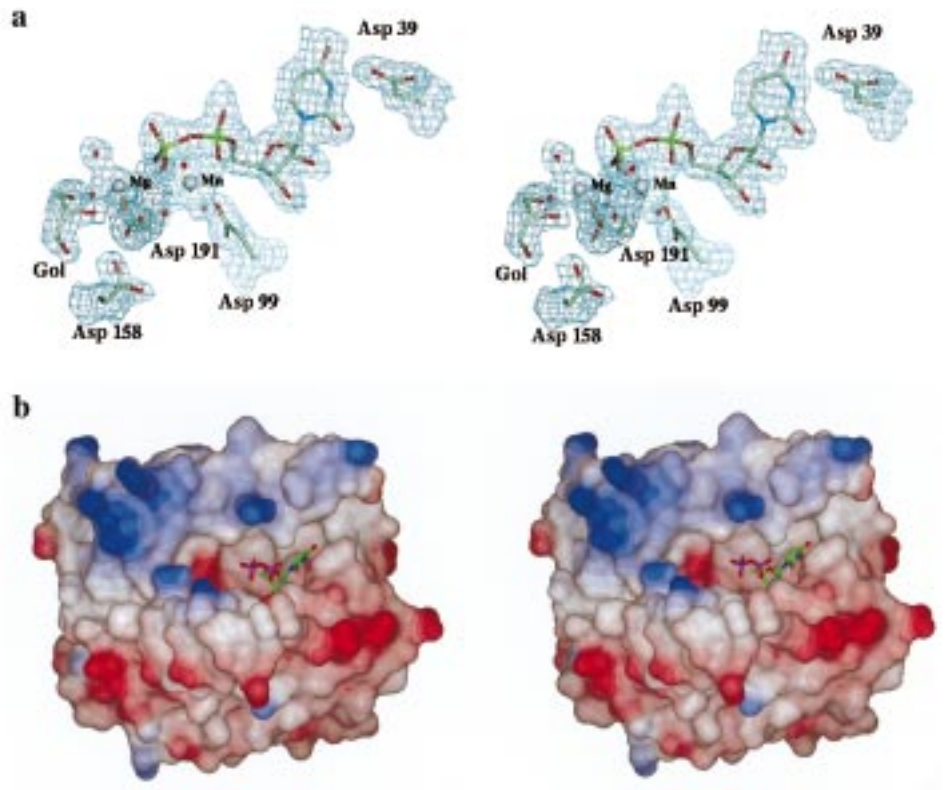

FIGURE 3: (a) Electron density for Mn−UDP bound to SpsA. The map shown is a maximum likelihood/σ_A weighted synthesis contoured at 0.4 electron/Å^3. Conserved residues Asp 39 and Asp 99, mentioned in the text, are denoted along with Asp 191, Asp 158, the active center glycerol molecule, and the additional Mg^{2+} ion and its ligands. (b) Electrostatic surface figure of SpsA, with UDP depicted in a "licorice" representation.

known. Sequence similarities suggest that the most likely nucleotide is UDP or dTDP. Several nucleotide diphosphates and NDP sugars were therefore screened by crystallography. Synchrotron data collection from crystals soaked in either Mg−UDP or Mn−UDP up to 1.7 Å resolution reveal details of the interactions between the enzyme and nucleotide. Preliminary investigations also show that dTDP binds similarly to UDP (S. J. Charnock and G. J. Davies, unpublished observations). In both the Mg−UDP and Mn−UDP complexes, the UDP binds to the nucleotide-binding domain in a deep cleft (Figure 3a,b). The ribose adopts a C2′-endo conformation with the base equatorial to the ring, markedly different from the orientation observed for the base in the bacteriophage T4 glucosyltransferase. The marked difference between the Mn−UDP and Mg−UDP complexes, of SpsA, is at the distal phosphate, which appears to prefer the greater ionic radius of Mn^{2+} as opposed to that of Mg^{2+} (Figure 3a).

The uracil base binds by aromatic stacking with Tyr 11 (Figure 4). Uracil carbonyl O4 forms a hydrogen bond to NH1 of Arg 71. Further interactions with the uracil moiety are provided by the invariant Asp 39 on the end of strand β-2, which makes a hydrogen bond to N3 of the uracil base. Asp 39 is one of two invariant aspartates whose function has been long speculated about, especially in the related cellulose synthase system (*24*); see below. The O2 carbonyl of the base appears to make no direct hydrogen bonds to the protein or solvent, which may perhaps aid the stability of the otherwise labile N1−C1′ bond. Instead, it sits in a hydrophobic pocket with Leu 80 at its base. Interaction with O3′ of ribose involves the main chain carbonyl and side chain hydroxyl of Thr 9. H-bonding to the O2′ hydroxyl is provided by the carboxylate of Asp 98. The diphosphate moiety makes just one direct H-bond to the protein, that from an oxygen of the proximal phosphate group to Lys 13. Family 2 sequences exhibit a remarkable sequence diversity in this region, implying that different ways of stabilizing the negative charge on the phosphates have evolved.

In the Mn−UDP complex, a single Mn^{2+} ion is found to be located between the two phosphates, which in turn interacts with the protein via Asp 99. This residue lies on β-strand 4, as was predicted using hydrophobic cluster analysis (*25*), and is strictly invariant throughout family 2. It seems likely that the role of this residue is to assist leaving group departure through coordination of the divalent manganese ion (Figure 1). Site-directed mutagenesis of the equivalent of Asp 99 in *Acetobacter xylinium* cellulose synthase led to a loss of cellulose synthase activity, implying a critical role in catalysis (*24*). In the Mg−UDP complex, in the absence of Mn^{2+}, the corresponding Mg^{2+} does not interact with both phosphates, instead showing incomplete coordination to the distal phosphate alone. As a result, the terminal phosphate is less well ordered (data not shown). This leads us to conclude that the optimal metal for SpsA is Mn^{2+}.

Further analysis of the active site of SpsA is complicated by the presence of a glycerol molecule coordinated to Asp 191, as discussed above. This, in turn, forms one of the axial ligands of a second metal, a hexacoordinate Mg^{2+} ion (equivalent to that seen in the native structure), which also interacts with the distal phosphate of UDP. Recent work on a mechanistically similar enzyme, the hypoxanthine phos-

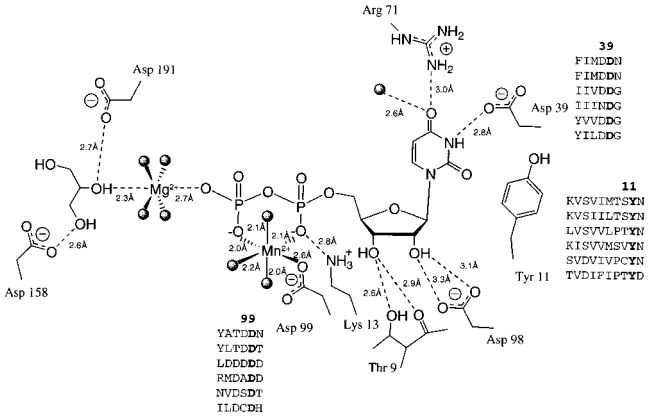

FIGURE 4: Interactions of SpsA with UDP−Mn. Glycosyltransferase family 2 contains many enzymes, so representitive sequence similarities for the conserved regions are given. The following sequences are depicted: (a) SpsA from *B. subtilis*, (b) CgeD from *B. subtilis*, (c) 334-amino acid hypothetical protein from *Pyrococcus horikoshii*, (d) glycosyltransferase involved in O-antigen synthesis from *Vibrio cholerea*, (e) root nodulation factor NodC from *Rhizobium* sp., and (f) cellulose synthase from *Acetobacter xylinus* (A subunit).

phribosyltransferase (HPRT) from *Trypanosema cruzi*, also revealed both Mn^{2+} and Mg^{2+} coordinated to the distal phosphate of the leaving group (*20*). In that case, the two are thought to be catalytic, but it is unclear whether SpsA, or other family 2 glycosyltransferases with known specificities, utilize a similar two-metal mechanism. It would seem unlikely, however, since this magnesium ion occupies the most likely position for donor sugar binding, in contrast to its location in the phosphoribosyltransferases.

Implications for the Catalytic Mechanism of Family 2 Inverting Transferases. The Mn−UDP complex of SpsA, interpreted in light of the known sequence conservation within family 2, allows us to speculate about the likely catalytic mechanism for this family. The vast majority of sequence conservation in family 2 is in the N-terminal nucleotide-binding domain, residues 2−100. Indeed, sequence conservation in this area is the defining feature of this family. Three clusters of invariance are particularly relevant (Figure 4), and all involve residues in direct contact with the nucleotide diphosphate donor. Tyr 11 is involved in stacking with the nucleotide base and Asp 39 in UDP binding, and Asp 99 sits adjacent to the distal phosphate where it coordinates the leaving group Mn^{2+} ion. The interaction between Asp 99 and the Mn^{2+}−phosphate strongly supports a role for Asp 99 in leaving group departure. A candidate for the general base is more difficult to identify. The structure exhibits four potential base functions: the "triad" of Asp 158, His 159, and Cys 160 and Asp 191. None of the residues of the Asp, His, Cys triad appear to be conserved, although sequence alignments in this region are difficult to interpret with certainty. Geometrically, Asp 191 is in an ideal location and may interact with the glycerol molecule in a way analogous to its coordination of the natural sugar, as is frequently the case with glycoside hydrolases (for example, see ref *26*). Additionally, iterative sequence searches with ψ-BLAST (*27*) suggest that this residue may be conserved throughout family 2, but again, sequence alignments may not be interpreted with great confidence. We are therefore tempted to speculate that Asp 191 functions as the catalytic base, and that the adjacent Mg^{2+} lies in the position occupied by the positive charge on the oxocarbenium ion-like transition state. Further structures of distantly related members of family 2 of the glycosyltransferases are required to permit further identification of residues that are invariant in the three-dimensional structure.

Glycosyl transfer is one of the key reactions of living cells, yet the machinery and mechanisms involved are poorly understood. NDP-sugar transferases are active in aqueous solution, yet they prevent the abortive transfer of water to the NDP sugar. The disordered loop of SpsA may provide one such mechanism by ensuring that a catalytically active complex is only formed in the presence of the appropriate substrates. Similar mechanisms exist for phosphoryltransferases which face the same problem of transfer to water and have recently been shown to exist in the phosphoribosyl transferases such as HPRT (*20*).

Implications for the Synthesis of Cellulose and Chitin. One of the most important and well-studied sequence homologues of SpsA is cellulose synthase. Two hypotheses exist for the role of this enzyme in the formation of cellulose. On the basis of pulse-chase experiments, it has been proposed that cellulose is elongated at the reducing end, with cellulose synthase itself playing a role in the shuttling of glucose between lipid pyrophosphate intermediates (*28*). The alternative hypothesis is that cellulose synthase acts as a UDP-glucose transferase, elongating the polymer at the nonreducing end (*24*, *25*, *29*, *30*). This latter hypothesis also proposes that cellulose synthase achieves the 2-fold rotation between adjacent glucose units in the cellulose crystal by the addition of two UDP-Glc molecules by two distinct catalytic centers within the same polypeptide chain. Such a dual addition is also required to explain the biosynthesis of the alternating polysaccharide hyaluronic acid (β-1,4-GlcNAc-β-1,4-glucuronic acid)$_n$ and for the formation of the Nod-factor chito-oligosaccharides and chitin itself. The structure of SpsA provides little support for either proposal, although speculation is difficult when the substrate specificity of SpsA is unknown.

The sequence of cellulose synthase is similar to those of other well-characterized NDP-sugar transferases. The majority of the sequence similarity, including residues that are essential for catalytic function, lies in the N-terminal domain of SpsA. We have shown this to be the binding site for nucleotide with the invariant residues intimately involved with UDP binding. It seems extremely unlikely that cellulose synthase displays sequence conservation through this domain if its role is, as proposed by Han and Robyt (*28*), in the transfer of lipid pyrophosphate-bound glucose species and not as an NDP-sugar transferase. The alternative proposals, in which a double addition of glucose is performed at the nonreducing end, are derived from sequence analyses which revealed conserved aspartate-containing motifs within the A and B "domains" of cellulose synthase and related "processive" glycosyltransferases. These were interpreted as equivalent catalytic functions within the two domains, one for the addition of each sugar (*24*, *25*, *29*, *30*). While the structure of SpsA strongly favors a model for direct addition at the nonreducing end, it does not support the "double-addition" mechanism for cellulose synthesis.

The A domain conservation of the processive enzymes, such as cellulose synthase, consists of residues 1−170 of SpsA, with the first two invariant aspartates clearly corre-

sponding to Asp 39 and 99. It seems likely, as discussed above, that while residues 1−100 of SpsA make up the nucleotide-binding domain further residues are required to provide the full catalytic functionality. The most likely candidate for the catalytic base is Asp 191, but even irrespective of the identity of the base, the minimum requirement for catalytic competence would appear to be approximately 200 residues. The signature motifs of the processive transferases are not found in SpsA, but they lie within the first, equivalent, 200 amino acids which are part of the single catalytic module of SpsA. It would therefore seem unlikely that they are able to constitute a discrete second catalytic center. It remains a possibility, however, that these additional residues in the processive transferases may assist in the dimerization of independent molecules, an alternative structural framework for polymerization by double addition. Interestingly, a number of other sequences do indeed possess tandem repeats of an SpsA catalytic "module", suggesting that a "two-catalytic centres within one peptide" model will be valid in some cases.

In addition to its similarity to polysaccharide synthases and enzymes involved in bacterial surface antigen synthesis, distant sequence similarity can be also be detected between SpsA and the ceramide glucosyltransferases which catalyze the first step in the production of gangliosides. This suggests that SpsA may provide the prototypical framework not only for family 2 glycosyltransferases but also for a wide range of distantly related enzymes which together will form a "clan" of related glycosyltransferases. Furthermore, the N-terminal domain of SpsA exhibits low, but significant, sequence similarity with family 27 of the glycosyltransferases (polypeptide-*N*-acetylgalactosamine transferases) which act with net retention of anomeric configuration. This opens up the intriguing possibility of domain swapping between transferases during evolution, in which a common nucleotide-leaving group module may have combined with a retaining catalytic module to provide a new catalytic function.

NOTE ADDED IN PROOF

Since the submission of this paper, two papers relevant to family 2 glycosyltransferases have appeared. The potential similarity with family 27 retaining glycosyltransferases is strengthened by site-directed mutagenesis of the UDP-*N*-acetyl-D-galactosamine polypeptide *N*-acetylgalactosamine-transferase (*32*). In an unrelated study, support for a mechanistic model for direct transfer from an NDP sugar to the nonreducing end of the growing polymer has been elegantly demonstrated for the family 2 enzyme NodC using *p*NPGlcNAc as the artificial acceptor species (*33*).

ACKNOWLEDGMENT

We thank the staff of the Daresbury SRS, the EMBL Hamburg outstation, and the ESRF (Grenoble) for provision of data collection facilities and the referees for useful comments.

REFERENCES

1. Boons, G.-J. (1996) *Tetrahedron 52*, 1095−1121.
2. Reichel, F., and Boons, G.-J. (1998) *Chem. Br. 34*, 43−51.
3. Campbell, J. A., Davies, G. J., Bulone, V., and Henrissat, B. (1997) *Biochem. J. 326*, 929−942.
4. Henrissat, B., and Davies, G. J. (1997) *Curr. Opin. Struct. Biol. 7*, 637−644.
5. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. 316*, 695−696.
6. Sinnott, M. L. (1990) *Chem. Rev. 90*, 1171−1202.
7. Davies, G., Sinnott, M. L., and Withers, S. G. (1997) in *Comprehensive Biological Catalysis* (Sinnott, M. L., Ed.) pp 119−209, Academic Press, London.
8. Vrielink, A., Rüger, W., Driessen, H. P. C., and Freemont, P. S. (1994) *EMBO J. 13*, 3413−3422.
9. Stragier, P., and Losick, R. (1996) *Annu. Rev. Genet. 30*, 297−341.
10. Roels, S., and Losick, R. (1995) *J. Bacteriol. 177*, 6263−6275.
11. Charnock, S. J., and Davies, G. J. (1998) *Acta Crystallogr. D55*, 677−678.
12. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
13. Otwinowski, Z. (1993) in *Data Collection and Processing: proceedings of the CCP4 study weekend* (Sawyer, L., Isaacs, N., and Bailey, S., Eds.) Science and Engineering Research Council, Daresbury, U.K.
14. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology: Macromolecular Crystallography, part A* (Carter, J., and Sweet, R. M., Eds.) pp 307−326, Academic Press, London and New York.
15. De La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol. 276*, 472−494.
16. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr. D49*, 148−157.
17. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D53*, 240−255.
18. Brünger, A. T. (1992) *Nature 355*, 472−475.
19. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D49*, 129−147.
20. Focia, P. J., Craig, S. P., III, and Eakin, A. E. (1998) *Biochemistry 37*, 17120−17127.
21. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
22. Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L., and Woldike, H. F. (1990) *Biochemistry 29*, 6244−6249.
23. Brzozowski, A. M., and Davies, G. J. (1997) *Biochemistry 36*, 10837−10845.
24. Saxena, I. M., and Brown, R. M. (1997) *Cellulose 4*, 33−49.
25. Saxena, I. M., Brown, R. M., Fevre, M., Geremia, R. A., and Henrissat, B. (1995) *J. Bacteriol. 177*, 1419−1424.
26. Burmeister, W. P., Cottaz, S., Driguez, H., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663−675.
27. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
28. Han, N. S., and Roybt, J. F. (1998) *Carbohydr. Res. 313*, 125−133.
29. Koyama, M., Helbert, W., Imaj, T., Suginama, J., and Henrissat, B. (1988) *Proc. Natl. Acad. Sci. U.S.A. 94*, 9091−9095.
30. Carpita, N., and Vergara, C. (1998) *Science 279*, 672−673.
31. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
32. Hagen, F. K., Hazes, B., Raffo, R., deSa, D., and Tabak, L. A. (1999) *J. Biol. Chem. 274*, 6797−6803.
33. Kamst, E., Bakkers, J., Quaedvleig, N. E. M., Pilling, J., Kijne, J. W., Lugtenberg, B. J. J., and Spaink, H. P. (1999) *Biochemistry 38*, 4045−4052.

BI990270Y