Chemistry & Biology, Vol. 9, 1337–1346, December, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S1074-5521(02)00292-2

Insights into Trehalose Synthesis Provided by the Structure of the Retaining Glucosyltransferase OtsA

Robert P. Gibson, Johan P. Turkenburg, Simon J. Charnock,[2] Ruth Lloyd,[3] and Gideon J. Davies[1]
Structural Biology Laboratory
Department of Chemistry
The University of York
Heslington, York YO10 5YW
United Kingdom

Summary

Trehalose is a nonreducing disaccharide that plays a major role in many organisms, most notably in survival and stress responses. In *Mycobacterium tuberculosis*, it plays a central role as the carbohydrate core of numerous immunogenic glycolipids including "cord factor" (trehalose 6,6′-dimycolate). The classical pathway for trehalose synthesis involves the condensation of UDP-glucose and glucose-6-phosphate to afford trehalose-6-phosphate, catalyzed by the retaining glycosyltransferase OtsA. The configurations of two anomeric positions are set simultaneously, resulting in the formation of a double glycoside. The three-dimensional structure of the *Escherichia coli* OtsA, in complex with both UDP and glucose-6-phosphate, reveals the active site at the interface of two β/α/β domains. The overall structure and the intimate details of the catalytic machinery reveal a striking similarity to glycogen phosphorylase, indicating a strong evolutionary link and suggesting a common catalytic mechanism.

Introduction

Trehalose, α-D-glucopyranosyl α-D-glucopyranoside, is an unusual, nonreducing disaccharide widely distributed in a large range of organisms and characterized most notably in insect, plant, and microbial cells. Its production is often associated with physiological stresses such as osmotic and heat shock, sporulation, and dehydration. It has unique water-retention properties, allowing resurrection plants to survive almost complete desiccation [1], which has led to its widespread use in the food and pharmaceutical industries [2, 3]. In insects, it is not only a major energy storage form for glucose, where it substitutes for glycogen, but emerging work also shows a role in anoxic tolerance and development [4]. Trehalose and its metabolism are increasingly relevant to bacterial persistence processes [5], and perhaps the most well-documented role for trehalose is in mycobacteria, where it is produced constitutively and involved in the formation of numerous glycolipids [6, 7] including the potently-immunogenic trehalose 6,6′-dimycolate (cord factor). While *M. tuberculosis* presents three alternative pathways for trehalose synthesis [7], most organisms synthesize trehalose in a two-step process catalyzed by the OtsA and OtsB proteins.

OtsA is a UDP-glucose (UDP-Glc) dependent glycosyltransferase. It catalyzes the synthesis of trehalose-6-phosphate using UDP-Glc as donor and the α anomer of glucose-6-phosphate (Glc-6-P) as the acceptor. The reaction occurs with net retention of the anomeric configuration of the donor sugar (Figure 1A) and results in the formation of a double glycosidic linkage (Figure 1B), generating trehalose-6-phosphate as product (Figure 1C). Trehalose-6-phosphate is subsequently dephosphorylated in a reaction catalyzed by OtsB, a trehalose phosphatase to yield trehalose. Seventy-three (as 08/28/02) independent OtsA-like sequences are known, derived from across a wide taxonomic landscape, and are classified into the carbohydrate active enzymes (CAZy) family GT-20; one of the 60 sequence-derived families for open-reading frames (ORFs) encoding glycosyltransferases [8–10].

Mechanistic considerations are central to glycosyltransferase research. They underpin both fundamental studies and applications while also providing the framework for therapeutic inhibition of these enzymes. The glycosyl transfer reaction may proceed with one of two possible stereochemical outcomes: the configuration of the anomeric center may either be inverted or retained during the transfer event (Figure 1). Different enzymes perform the reaction with strict stereochemical control and are thus classified as inverting or retaining. Glycosyl transfer with inversion of configuration is a conceptually straightforward reaction involving a concerted S_N2-like single displacement in which the acceptor sugar attacks the anomeric center from the opposite face concomitant with leaving-group departure. In addition to stabilization of the oxocarbenium-ion like transition state, the enzyme provides a general base function, typically a carboxylate, to activate the acceptor nucleophile through deprotonation [reviewed in 11].

The mechanism of glycosyl transfer with net retention of anomeric configuration is less clear [for discussion see 12, 13]. Although the hydrolysis of glycosides with net retention of anomeric configuration clearly involves the formation and subsequent breakdown of a covalent reaction intermediate [14, and for review see 15, 16], evidence for a similar double-displacement reaction for glycosyl transfer from activated sugar donors, such as nucleotide sugars, has not been forthcoming. This has led to consideration of a variety of mechanisms including classical double displacement with enzymatic nucleophile [17, 18], double-displacement with intramolecular attack of phosphate [19], or the acceptor itself [20]; S_Ni "internal return-like" mechanisms [13, 20]; and long-lived glycosyl cation intermediates (as proposed for glycogen phosphorylases [21]). So far, none of the mechanistic candidates is ultimately persuasive.

[1]Correspondence: davies@ysbl.york.ac.uk
[2]Present address: School of Applied and Molecular Sciences, Northumbria University, The Ellison Building, Newcastle upon Tyne NE1 8ST, United Kingdom.
[3]Present address: Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, United Kingdom.

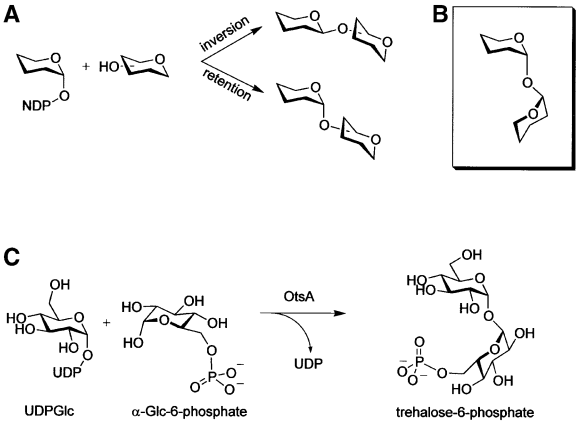


Figure 1. Glycosyltransfer with Inversion and Retention

(A) The synthesis of glycosidic bonds from activated nucleotide-sugar donors may proceed with either inversion or retention of anomeric configuration. (B) The formation of nonreducing double glycosides is unusual in that it involves the formation of two glycosidic linkages. (C) The reaction catalyzed by OtsA: the transfer of glucose, from UDP-Glucose to glucose-6-phosphate to generate trehalose-6-phosphate. T-6-P is subsequently dephosphorylated by OtsB to yield α,α-1,1 trehalose.

The mechanistic conundrums surrounding glycosyl transfer, together with the vast importance of the reactions themselves, have, in recent years, led to increased interest in the three-dimensional structures of glycosyltransferases. Yet, in contrast to the abundance of sequence data, three-dimensional structures for these enzymes remains scarce. To date, structures for representatives of only nine of the CAZy glycosyltransferases families (compared to approximately fifty of the eighty glycoside hydrolase families) have been determined, and seven of these are for the conceptually simpler inverting class of glycosyltransferase [reviewed in 11]. Just two families of retaining nucleotide-sugar-dependent enzymes have representative three-dimensional structures, GT-6 and GT-8. Two family GT-8 structures have been reported, the α-1,4-galactosyl transferase LgtC from *Neisseria meningitides* [20] and rabbit muscle glycogenin [18], and three family GT-6 structures have been reported, the bovine α-1,3-galactosyl-transferase [13, 17] and the human blood group glycosyltransferases GTA and GTB [22]. It is notable that, in stark contrast to the large diversity of folds displayed by the glycoside hydrolase families [23, 24], only two protein topologies have been observed for nucleotide-sugar-dependent glycosyltransferases [25]. These canonical folds have been described as "fold family GT-A and GT-B" corresponding to the prototypical SpsA [26] and T4 Glc transferase [27] structures, respectively. Elegant sequence work suggests that many of the remaining fifty-plus sequence-derived families will also utilize one of these two scaffolds [25, 28]. Thus far there have been no structures determined for retaining nucleotide-sugar glycosyltransferases that display the twin-Rossmann GT-B fold.

OtsA, whose structure we report here, displays the twin-Rossman GT-B fold, previously reported for some inverting glycosyltransferases. Both the overall fold and the intimate details of the catalytic center show a striking similarity to glycogen and maltodextrin phosphorylases, indicative of a conserved catalytic mechanism for glycogen breakdown and the synthesis of glycosidic bonds from UDP-sugar precursors.

Results and Discussion

The Three-Dimensional Structure of Trehalose-6-Phosphate Synthase OtsA

E. coli OtsA in a $(His)_6$-tagged form may be overexpressed in a standard pET-based system [29]. Modification of the growth conditions permitted structure solution using the multiple-wavelength anomalous dispersion (MAD) method from a selenomethionyl-derived protein The *E. coli* OtsA trehalose-6-phosphate synthase structure, in abortive complex with both UDP and Glc-6-P, was determined by MAD with three wavelengths harnessing data to 3.1 Å resolution and including single wavelength data from a second Se-Met OtsA crystal to 2.45 Å (see Experimantal Procedures).

Light scattering data (not shown) show that overexpressed His-tagged OtsA exists as both dimers and tetramers in equilibrium. It is not clear if this has any biological significance, perhaps reflecting the ability of a dimer to interact with the phosphatase OtsB (although in *E. coli*, OtsA most likely exists as a discrete entity; in some other organisms, such as *Saccharomyces cerevisae*, OtsA and OtsB are associated into a multicomponent complex [30]) or are merely an overexpression artifact. The $P2_12_12_1$ crystal form contains four independent copies of OtsA arranged as a 222 tetramer (Figure 2A). The noncrystallographic symmetry was harnessed throughout structure solution and was essential both for phase improvement through noncrystallographic symmetry averaging and for refinement through the incorporation of tight NCS-restraints (see Experimental Procedures).

The final model structure, from residues 1–456 inclusive (The C-terminal residues 457–478 are not visible in density), has an R_{cryst} of 0.19 with R_{free} 0.24. Two residues are in forbidden regions of the Ramachandran plot, Asp130 and Asn364; both are involved in substrate binding, described below. OtsA displays two domains, each of the β/α/β Rossmann-fold type, with a deep fissure at the interface forming the catalytic center. The two domains of each of the four molecules of OtsA, described below, display varying degrees of rigid-body motion in-crystal, highlighting the potential for domain opening and closing even within the constraints of the crystal lattice. Such motion of the N- and C-terminal domains of OtsA necessitated use of TLS (translation, libration, screw-rotation) refinement [31]. Mean translational and librational components to the TLS parameters range from 0.1 Å and 3^{o2} for the N-terminal domain of molecule A through to 1 Å and over 6.6^{o2} for the N-terminal domain of molecule C, presumably arising from the comparative paucity of molecule C lattice-contacts and reflected in extremely poor electron density for this highly mobile domain (Table 1).

The N-terminal domain, residues 1–225, has a central core of seven parallel β strands flanked at either end of

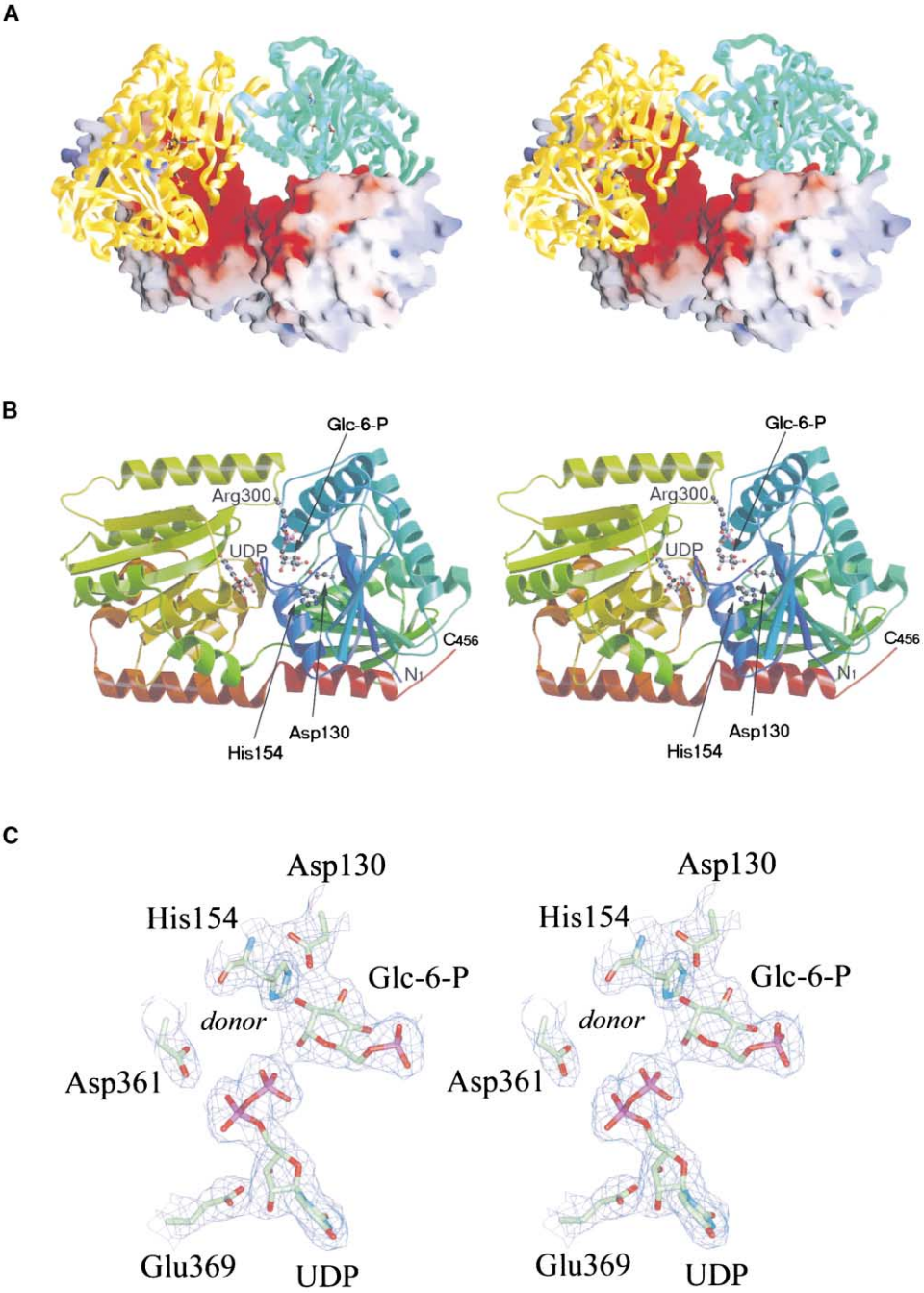


Figure 2. Three-Dimensional Structure of Trehalose-6-Phosphate Synthase

(A) The tetramer of OtsA as a mixed electrostatic surface/protein cartoon figure. Ligands, which bind at the interface of the N- and C-terminal domains of each monomer, are shown in liquorice. This figure, in divergent (wall-eyed) stereo, was prepared using GRASP [48] (electrostatic contour range from −15 [red] to +13 [blue]). (B) Three-dimensional structure of *E. coli* OtsA in divergent (wall-eyed) stereo, color-ramped from residue 1 at the N terminus to residue 456 at the C terminus. The ligands, Glc-6-P and UDP, are shown in ball-and-stick representation, as are residues Asp130, His154, and Arg300. This figure was drawn with BOBSCRIPT [49]. (C) Observed electron density at the catalytic center of OtsA in divergent (wall-eyed) stereo. The UDP and Glc-6-P are shown together with Asp130 and His154, which are conserved in glycogen and maltodextrin phosphorylases[20,21]. The location of the donor site, as observed in the maltodextrin phosphorylase[20], is indicated. Asp361 and Glu369, proposed GPGTF superfamily motif elements [28], are included for reference. The map shown is a maximum-likelihood [45]/σ_A [50] weighted $2F_{obs}$-F_{calc} map contoured at 0.23 electrons / Å^3 (1.5σ).

the sheet with a single, anti-parallel strand, together with six associated helices (Figure 2B). The C-terminal domain (226–456) is also of Rossmann-fold topology with a central parallel β sheet of six strands flanked by seven flanking α helices. The C-terminal helix undergoes a kink at residue 438 at the interface of the two domains

Table 1. Multiple-Wavelength Anomalous Dispersion Data Quality and Statistics for the UDP and Glc-6-Phosphate Complex of OtsA

	Peak	Inflection	Remote	High Resolution		
Data quality						
Wavelength (Å)	0.97926	0.97950	0.93920	0.93920		
Resolution of data (Å)	20–3.1	20–3.1	20–3.1	20.0–2.45		
(Resolution of outer shell)	3.2–3.1	3.2–3.1	3.2–3.1	2.54–2.45		
[a]R_{merge} (outer shell)	0.08 (0.42)	0.09 (0.49)	0.063 (0.30)	0.056 (0.43)		
Mean I/σI	15 (3.3)	14 (2.7)	19 (4.9)	17 (2.4)		
Completeness %	99.6 (99.0)	99.6 (99.0)	99.6 (98.6)	99.9 (99.5)		
Multiplicity[1]	3.7 (3.4)	3.7(3.4)	3.6 (3.3)	2.8 (2.3)		
Refinement statistics						
R_{cryst}				0.19		
R_{free}				0.24		
Rms bond deviation (Å)				0.019		
Rms angle deviation (°)				1.95		
Rms chiral volumes (Å^3)				0.145		
NCS deviations, (N domain/C domain, Å)				A–B	A–C	A–D
				0.03/0.06	0.02/0.07	0.03/0.05
Ramachandran Outliers (%)[2]				2.8		
Mean B protein atoms molecules A–D (Å^2)				A B C D		
				33 33 33 35		

[1]These values reflect treating hkl and −h, −k−l independently during scaling. Overall multiplicity is therefore approximately twice the value shown.

[2]Calculated using Uppsala Ramachadran Server. See Kleywegt, G.J., and Jones, T.A. (1996). Phi/Psi-chology: Ramachandran revisited. Structure *4*, 1395–1400.

(a general feature of fold family GT-B enzymes), with the final 18 residues crossing over to the N-terminal domain and continuing as α helix. The interface between the N- and C-terminal domains provides the location for the bound ligands UDP and Glc-6-P and the catalytic apparatus.

Sequence Considerations: From GT-20 to Beyond

As of 08/28/02, CAZy family GT-20 contains seventy-three unique sequences. An alignment of four representative sequences (*E. coli, Candida albicans* [34% identity with *E. coli*], *M. tuberculosis* [32%], and *Arabidopsis thaliana* [33%]) is shown in Figure 3. Most of the highly conserved regions are involved in substrate binding and catalysis, with clusters of invariance centered around Arg9, Trp40, Tyr 76, Trp85, and Arg300 involved in the binding of the Glc-6-phosphate acceptor, while conserved clusters around Gly22, Asp130, His154, Arg262, Asp361, and Glu369 are all involved in the binding of UDP, described in light of the complex of OtsA with UDP and Glc-6-P below.

Many of the residues that form the A-B dimer interface of the tetramer are conserved, notably the cluster around Pro156 and Glu161. The A-D interface is formed by the helix 305-323 and the subsequent loop to 331 (and the same residues from molecule D). Although residues 318-322 are invariant, their side chains (where present) point inwards toward the A molecule and do not interact with molecule D in a significant way. This suggests that the tetramer organization is not conserved across species, consistent with the known intimate involvement of some OtsAs with their phosphatase partner. Furthermore, there are a number of patches of conservation whose function is unclear. Clusters around residues Leu181, Pro214, and Arg319, for example, all lie in exposed surface regions. Such sites may represent a potential interaction surface for the *E. coli* dimer with the phosphatase subunit(s), although this is, as yet, undemonstrated for the *E. coli* system. It is conceivable that these sites could also be involved in the binding of a regulatory factor (some OtsA homologs are reported to change their activity through the binding of large polyanions, for example [32]), but again we have no experimental evidence for such a role.

Classical sequence similarity detection methods using OtsA fail to detect similarities to other families of glycosyltransferase. In a powerful predictive study, however, iterative PSI-BLAST [33] searches, based-upon family GT-41 *O*-GlcNAc transferases, allowed Wrabl and Grishin [28] to propose that up to one-third of all GT families, including the OtsA family GT-20, would adopt the same dual-Rossmann domain 3-D fold. Furthermore, they identified a signature, termed the glycogen phosphorylase/glycosyltransferase (GPGTF) motif, diagnostic of this fold family. Two positions in the C-terminal domain, a conserved Asp/Glu/Lys in the middle of helix 4 termed position 1 and a conserved Asp/Glu between strand 4 and helix 4 termed position 2, were identified as particularly significant. Both positions are indeed present in OtsA and play key roles in binding of the UDP-Glc donor species, described below. This fold family (also termed fold family GT-B [25]) is not, however, diagnostic of catalytic mechanism for it contains both inverting and retaining glycosyltransferases. OtsA, as a retaining glycosyltransferase, provides the missing link of this superfamily.

UDP Binding: The Nucleotide-Sugar Donor Site

Electron density at the active-center clearly reveals bound UDP and Glc-6-P (Figure 2C) and allows description of the enzyme-ligand interactions (Figure 4). UDP binding interactions are predominantly with the C-terminal domain (Figure 2B). The uracil moiety is deeply buried within the core, and specificity is conferred

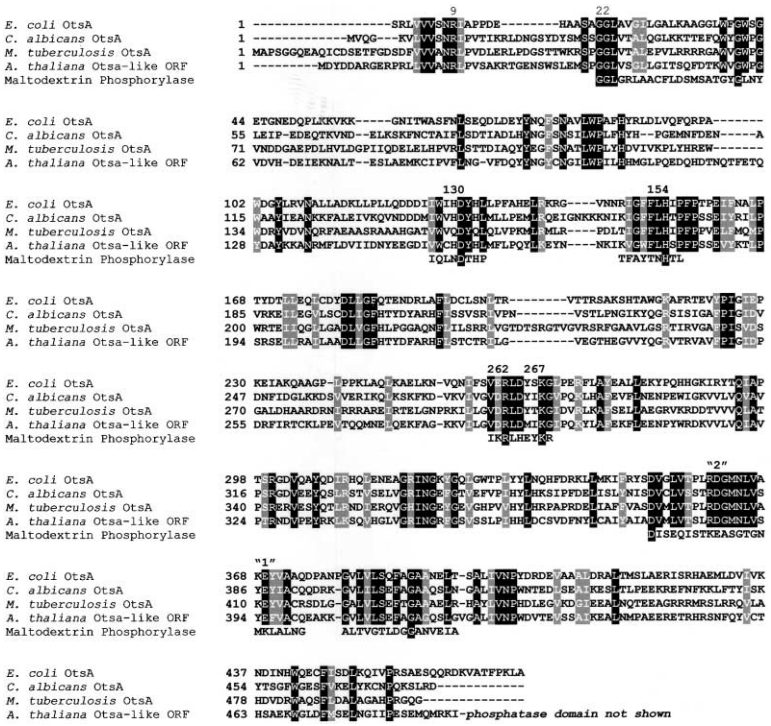


Figure 3. Primary Sequences of Some Family GT-20 Trehalose-6-Phosphate Synthases

Partial sequence alignment of family GT-20 sequences. Four representative sequences are shown *Escherichia coli* (P31677), *Candida albicans* (Q92410), *Mycobacterium tuberculosis* R37Rv (O06353), and the *Arabidopsis thaliana* ORF At1g16980/F6I1.1. Invariant and highly conserved regions are shaded. A partial sequence alignment of the *E. coli* maltodextrin phosphorylase is shown below containing just those residues mentioned in the text. Amino acids corresponding to Wrabl and Grishin's "glycogen phosphorylase glycosyltransferase" (GPGTF) motif positions 1 and 2 [28] are indicated.

through interactions of O4 and N3 with the main-chain amide hydrogen and carbonyl of residue Phe339. It is significant that there is no interaction with O2, and consequently one may easily model in GDP in place of UDP with N1 and O6 of GDP taking the place of N3 and O4 of UDP. This may account for the rather surprising observation that some OtsAs display relaxed nucleotide specificity [32]. Both O2 and O3 hydroxyls of the ribose interact with the carboxylate of Glu369. This is the residue identified as GPGTF superfamily motif position 1 in the sequence work of Wrabl and Grishin [28]. This residue is a carboxylate in all nucleotide-sugar-dependent enzymes of the superfamily. In glycogen phosphorylases, a lysine is observed at this position and plays the crucial role in forming the Schiff's base to the pyridoxal phosphate cofactor. The endocyclic oxygen of the ribose makes a 3.1 Å hydrogen bond to the main-chain amide of Gly21 found in a conserved (Figure 3) Gly_{21}-Gly_{22}-Leu_{23} motif.

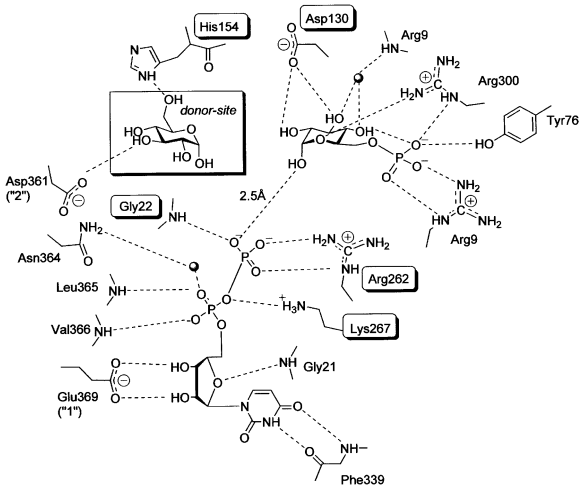


Figure 4. Schematic Diagram of the Catalytic Center of Trehalose-6-Phosphate Synthase

Schematic diagram of the catalytic center of OtsA. Residues invariant in the active center of glycogen and maltodextrin phosphorylases are labeled in shadowed boxes. The putative donor-site glucosyl moiety (from the overlap with *E. coli* maltodextrin phophorylase) is included for reference. GPGTF superfamily motif elements 1 and 2 are indicated [28].

O1 and O2 of the proximal phosphate group make interactions with the main-chain amides of Leu265 and Val366, respectively, and through a solvent-mediated hydrogen bond to Asn362, one of the two Ramachandran outliers described above. The ribose and distal phosphate environment thus contains no DxD motif, as observed in many other glycosyltransferases [11] and often claimed to be a signature motif for glycosyltransferases. The distal phosphate group (which would be bonded to the donor glucoside if present) interacts with the side-chains of both Arg262 and Lys267 and, consistent with the absences of a DxD motif, we observe no density corresponding to a divalent metal-ion coordinated to the two phosphate-groups. The distal phosphate also interacts with the main-chain amide of Gly22, the central residue in the conserved Gly-Gly-Leu motif described above. The twin glycines in this motif provide obvious cross-talk between UDP and the N-terminal domain and may thus be involved in domain closure upon ligand binding. Although a single imidazole group presumably from the crystallization buffer has been modeled adjacent to the distal phosphate, it lies in the position almost certainly occupied by the donor sugar in a productive complex (see below), and we consider its observation adventitious as opposed to catalytic.

Donor sugar interactions may not be directly observed in this complex, but the striking similarity with the active-center of glycogen and maltodextrin phosphorylases, described more fully below, allows them to be modeled with some confidence. O6 of the donor sugar most likely interacts with the side-chain of His154 and O3 with the carboxylate of Asp361. This latter residue is the second signature motif (position 2) for members of the GPGTF superfamily [28]. Wrabl and Grishin further observe that a carboxylate at position 2 is correlated with catalysis with net retention of anomeric configuration. The location of this residue appears indicative of an equatorial 3-hydroxyl in the donor sugar and must also be closely involved in transition-state stabilization as the sugar ring geometry changes during catalysis. It would not appear to play any role in the direction of attack of the acceptor (which governs stereochemical outcome), so previous correlations may be casual rather than causal. O2 and O4 interactions are harder to predict. The main-chain amide of Asn364, one of the Ramachandran outliers, blocks off the glucosyl donor site and may interact with O4 while His89, invariant in GT-20 enzymes, (Figure 3) may interact with O2.

Glucose-6-Phosphate Binding: The Acceptor Site

Glucose-6-phosphate, the acceptor, binds predominantly on the N-terminal domain. The phosphate lies in a positively charged pocket, interacting directly with Arg9 and Arg300. The latter interaction, donated from the C-terminal domain, may further facilitate domain closure in the ternary complex in conjunction with the UDP-sensing Gly-Gly-Leu motif described above. The hydroxyl of Tyr76 and water-mediated interactions, notably with His132, complete the phosphate coordination. All these residues are conserved in family GT-20 enzymes (Figure 3).

Both O2 and O3 hydroxyls of the glucosyl moiety hydrogen-bond to OD2 of Asp130. O3 also makes a solvent-mediated interaction, the water molecule hydrogen-bonding to both the main-chain amide of Arg9 and the 4-position hydroxyl group. O4 also interacts closely ($\sim$2.4 Å) with one of the 6-phosphate oxygens (Figure 4). Perhaps the most interesting interactions are at the O1 position. OtsA is notable for its ability to select just the α anomer of Glc-6-phosphate in order to generate α-α, 1-1 disaccharides as products. Ile155 and especially Trp85, the latter invariant, form a hydrophobic environment legislating against binding of the β anomer of Glc-6-P. The O1 hydroxyl is thus seen only in its α-configuration in appropriate position for nucleophilic attack on a donor sugar (were one bound to the distal phosphate). The α anomer is further favored by a close, 2.5 Å charged hydrogen bond to distal phosphate O1B of UDP. It is likely that this would be the glycosidic oxygen bridging to the phosphorous atom in a true donor complex, the mechanistic implications of which are discussed below. The interactions of the glucosyl moiety (notably that with Asp130) and of the distal phosphate of UDP are strikingly similar to the equivalent interactions seen in the catalytic center of glycogen and maltodextrin phosphorylases described below.

The entrance/exit of the acceptor site is partially enclosed by invariant Trp40, which may act as a ligand-driven gate to enclose the active-center and reduce solvent access. Furthermore, the deep burial of UDP(-Glc), with an exit adjacent only to Glc-6-P in this structure, suggests a compulsory order to the binding and reaction mechanism in which UDP-Glc binding is followed by Glc-6-P and, subsequent to turnover, release of trehalose-6-P is followed by UDP. Such a sequential reaction may be facilitated through ligand-gated domain movement influenced by the cross-talk provided by interactions of the N terminally bound acceptor with Arg300 from the C domain and the distal phosphate of UDP interaction with Gly22 from the N-terminal domain, above.

Similarity with Glycogen Phosphorylase and Other Enzymes

OtsA is similar to neither the family GT-8 transferases such as the *N. meningitidis* LgtC [20] nor the family GT-6 enzymes exemplified by the bovine β-1,3 Gal transferase [13, 17]. Distance matrix searches using DALI [34] instead reveal strong similarity to a number of twin Rossmann-fold enzymes, as had been proposed from sequence alone [25, 28]. The UDPGlcNAc 2′ epimerase [35] is structurally the most closely related (DALI Z score 20.3 with 335/376 overlapping C alphas displaying an rms of 4.4 Å), but OtsA also displays strong structural similarity with the other twin Rossman domain, inverting glycosyltransferases, T4 Glc transferase [27] (Z14.6, 266/333, 4.0 Å), MurG [36] (Z13.3, 281/351, 5.8 Å), and GtfB [37] (Z14.9, 292/382, 4.3 Å) (Figure 5A). With such flexible two-domain proteins, these rms values, based upon overlap of the whole structure, obviously represent the an upper limit for the deviations. The OtsA structure provides the missing link in the glycogen phosphorylase/glycosyltransferase superfamily whose members now include both inverting and retaining nucleotide-sugar glycosyltransferases, (retaining) sugar phosphorylases, and the UDPGlcNAc 2′ epimerase; all four enzyme classes linked by similar transition-state structures [35].

Many of the interactions of the UDP are invariant between members of this fold family. The uridine base is buried deeply within a hydrophobic pocket with main-chain hydrogen bonds to N3 and O4. The glutamic acid which binds the ribose O2 and O3 hydroxyls (GPGTF superfamily position 1) is found in all but the GtfB structure. Glycine repeats, in a flexible loops that have been termed G-loops [37], bind the phosphates in a similar way in the superfamily members. In the case of OtsA, the amide hydrogen of Gly22 hydrogen bonds to the O1B atom of the distal phosphate. The most striking similarity is, however, that observed with glycogen and maltodextrin phosphorylases. In this case, the similarity extends beyond the fold and sequence conservation predicted previously. Almost the complete catalytic center is invariant, suggesting a common evolutionary origin and catalytic mechanism for these enzymes [38].

Glycogen phosphorylase catalyzes the reversible phosphorylation of glycogen using inorganic phosphate in the presence of a pyridoxal phosphate cofactor. The UDP and PLP groups bind in equivalent locations in the OtsA and glycogen phosphorylase structures, respec-

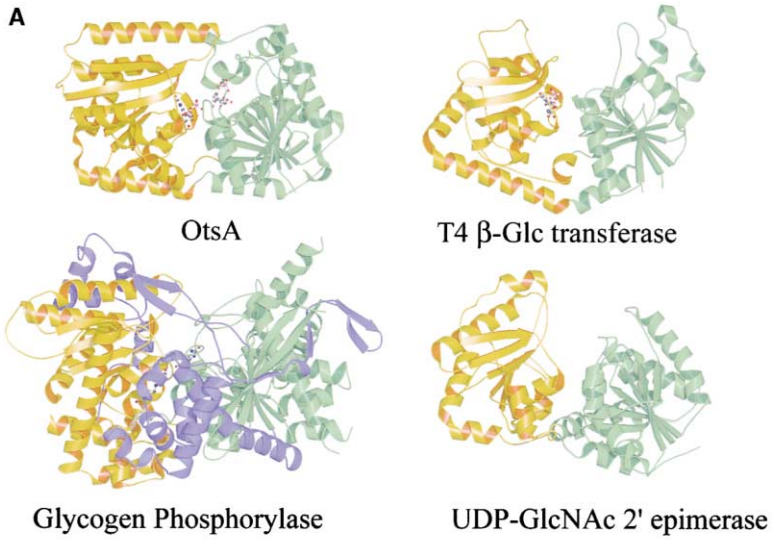
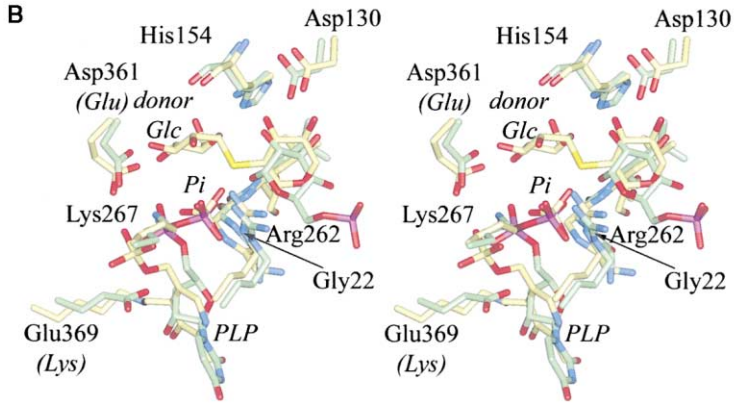


Figure 5. A Fold Family Linked by a Common Transition State

(A) Twin Rossmann-fold enzymes closely related to OtsA (1gz5.pdb), as determined using DALI [34], include the Phage T4 β-glucosyl transferase (1c3j.pdb), glycogen phosphorylase (1a8i.pdb), and UDP-GlcNAc 2′ epimerase (1f6d.pdb). The equivalent N- and C-terminal domains are colored in gold and green, respectively, with the insertions present in glycogen phosphorylases in blue. Ligands where present are shown in ball-and-stick representation. This figure was drawn with BOBSCRIPT [49]. (B) Active center overlap with maltodextrin phosphorylase [21]. The product complex (a nonhydrolyzable thiooligosaccharide plus phosphate plus PLP) of *E. coli* maltodextrin phosphorylase is shown with pale yellow bonds and the abortive complex (UDP and Glc-6-P) of OtsA in pale green. Residue labels for OtsA are shown. Differences observed the maltodextrin phosphorylase (Asp for Glu at position 2, Glu for Lys at position 1, PLP, and Pi) are indicated in italics.



tively, while the distal phosphate of UDP in OtsA overlaps the inorganic phosphate observed in glycogen phosphorylase (Figure 5B). The ribose and the pyridoxal group likewise lie in equivalent locations. A number of interactions appear mechanistically invarient, as revealed by both sequence and structural identity (Figures 3, 4, and 5). The interactions of the distal phosphate with Arg262 and the main-chain amide hydrogen of Gly22 are conserved in phosphorylase, as is the interaction of Lys267 with the bridging oxygen. Gly22 is the central residue in a Gly-Gly-Leu motif, also present in phosphorylase.

Interactions of the acceptor sugar are also conserved. Asp 130 hydrogen bonds to both O2 and O3 hydroxyls of the Glc-6-P, the equivalent residue Asp339 (rabbit muscle phosphorylase numbering), likewise interacts with the O3 and O2 hydroxyls of the acceptor-site glucose of phosphorylase. Both Asp130 (OtsA) and Asp339 (phosphorylase) show equivalent and unusual main-chain phi/psi angles of 90°, 174° and 61°, 184°, respectively. The significant difference is the orientation of the acceptor sugar. Whereas glycogen phosphorylase presents the O4 hydroxyl for nucleophilic attack, OtsA places the O1 hydroxyl in attacking position. This reversed orientation of the Glc-6-P acceptor in the OtsA is favored by the burial of the 6-phospho group. At the putative donor site, conservation is also revealed. Glu672 of phosphorylase, which interacts with the donor 3-OH, likewise finds direct equivalent in the carboxylate of Asp361 of OtsA (GPGTF superfamily position 2).

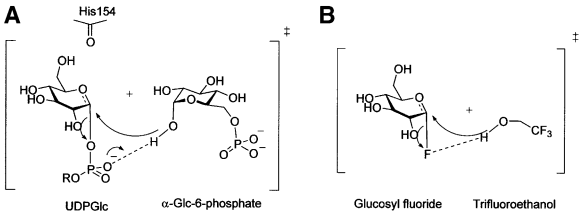


Figure 6. Putative Transition State for Glycosyltransfer with Retention of Anomeric Configuration

(A) A putative transition-state structure for glycosyl transfer catalyzed by OtsA, featuring a late oxonium-ion transition state with asynchronous front-face attack (B). Chemical precedent for concerted but necessarily asynchronous front-face attack is provided by the solvolysis of glucosylfluorides in trifluoroethanol[23]. Arrows are included solely to show the direction of electron movement prior to the transition state.

His154 is isosteric with its counterpart His377 in phosphorylase, suggesting that its side chain most likely interacts with the O6 hydroxyl of the donor glucoside in OtsA (Figures 4 and 5B). It is of particular importance that the main-chain carbonyl of His154, whose counterpart is strongly implicated in transition-state stabilization and catalysis in phosphorylase [38, 39], lies in identical position in OtsA. The catalytic centers of OtsA and glycogen/maltodextrin phosphorylases are thus almost identical in the disposition of catalytic and substrate binding residues, which points to a conserved catalytic mechanism.

Mechanistic Considerations

Glycosyltransferases harness activated sugars to drive the synthesis of di, oligo, and polysaccharides, and glycoconjugates. Although the activating group is typically an α linked (in the case of D sugars), nucleotide-, phosphate-, pyrophosphate-, and lipid-phosphate-linked donors are also known. In this context we can see that in the direction of glycosidic bond synthesis, the reactions catalyzed by glycogen phosphorylases and OtsA are chemically equivalent. Indeed, for this reason glycogen phosphorylase sequences are included as retaining glycosyltransferases in CAZy family GT-35 [10]. The invariant active center and catalytic machinery shared by OtsA and glycogen phosphorylases point strongly toward a shared reaction mechanism for these two enzymes. Furthermore, the structurally unrelated retaining glycosyltransferases of fold family GT-A possess an isostructural but side-chain-derived carbonyl function above the sugar ring, namely Gln189 in LgtC [20] and Glu317 in the bovine α-1,3 Gal transferase [13] (although this residue has also been proposed as a candidate nucleophile in a double-displacement mechanism [17]), located in an identical position to the main-chain carbonyl of His154 in OtsA, hinting that perhaps these enzymes too share a common catalytic strategy.

The catalytic mechanism of both glycogen phosphorylase and retaining glycosyltransferases in general remains a challenging area of enzymology. Although retention of anomeric configuration during the hydrolysis of glycosides is known to be achieved in a two-step double-displacement reaction via a covalent intermediate [reviewed in 16], there is no convincing experimental evidence that supports a similar mechanism for the synthesis of glycosidic bonds from phosphate-activated sugar donors. Perhaps the most widely accepted mechanism for glycogen phosphorylase, distilled from crystallographic and inhibitor studies, involves a late oxonium-ion-like transition state in which leaving group departure and nucleophilic attack, occurring in a concerted but necessarily asynchronous (orbital considerations presumably legislate against simultaneous front face attack and departure) manner on the same face of the glycoside (Figure 6A). Although unusual, a similar mechanism has been demonstrated for the solvolosis of glycosyl fluorides in trifluoroethanol with retention of configuration [40] (Figure 6B), a reaction which likewise involves departure of an anionic leaving group. Thus, in the absence of any convincing support for a covalent intermediate, such an internal return S_Ni-like mechanism is perhaps the strongest mechanistic candidate. An S_Ni-like mechanism would be further facilitated by the formation of a close hydrogen bond between the anionic leaving group and the incoming nucleophile. Yet, with little genuine information available about the reaction mechanism, the nature of this transition state remains uncertain. Information derived from inhibitor studies begins to shed some light on transition-state structure. In an intriguing study, the binding of nojirimycin tetrazole (a potent inhibitor of glycogen phosphorylase) displayed kinetic properties consistent with it binding as a transition state analog [39]. In contrast, glycogen phosphorylase is very poorly inhibited by acarbose, whose $^2H_3$ half-chair conformation instead renders it a potent glycoside hydrolase inhibitor.

Significance

Trehalose plays an important, often critical role in many bacteria, insects, and plants. In *M. tuberculosis*, trehalose-containing glycolipids, including the potent immunogen cord-factor (trehalose 6,6′-dimycolate), are not only implicated in cellular damage in the human host, but may also play a critical role in cell wall biosynthesis. In this organism, trehalose has been proposed to act as a mycolic acid carrier, transporting this long-chain fatty acid from its point of synthesis in the cytoplasm to the cell surface, where it is appended to the growing arabinogalactan chain by the action of mycolyltransferases that belong to the antigen 85 complex [41]. Trehalose biosynthesis represents a novel target for new therapeutic agents against mycobacteria. Interestingly, the trehalose analogs 6-azido-6-deoxy-α,α-trehalose [41] and trehalosamine are inhibitors of mycobacterial growth and are believed to act by inhibition of the mycolyl transferases of the antigen 85 complex. Given the supposed importance of trehalose in the primary metabolism of mycobacteria, it is unfortunate that *M. tuberculosis* possesses two alternative strategies for trehalose biosynthesis in addition to the OtsA-OtsB pathway, involving the isomerization of maltose and maltose-derived oligosaccharides to trehalose by transglycosylation [7]. Thus, it is unlikely that compounds could be discovered that are capable of inhibiting all three pathways simultaneously. *M. leprae*, however, possesses only two pathways for the biosynthesis of trehalose and, so, therapeutic strategies based on inhibition of trehalose biosynthesis may have more success. Trehalose synthesis also plays a crucial role in the survival of other pathogenic bacteria such as *Salmonella enterica* [5]. In the context of therapeutic intervention, the OtsA structure unlocks the vast body of data on glycogen phosphorylase inhibition to inform the design of new therapeutic agents directed toward trehalose biosynthesis. Despite this, our meager understanding of the mechanistic fundamentals of this class of enzymes remains a serious limitation.

Experimental Procedures

Crystallization and Data Collection

Selenomethionyl-derived OtsA was prepared essentially as described for the native enzyme [29] but using the methionine auxo-

troph *E. coli* B834 (DE3; Novagen). Crystals of the orthorhombic form of OtsA (cell dimensions a = 103.6 Å, b = 125.4 Å, c = 176.6 Å, $\alpha = \beta = \gamma = 90°$) were grown from 14% (v/v) PEG 4000, 0.15 M NaOAc, 0.1 M imidazole buffer (pH 7.5), and 200 mM NaCl, with the ligands Mg-UDP and glucose-6-phosphate at 10 mM. A cryoprotectant solution was produced by the addition of 25% (v/v) PEG 400 to the growth buffer. Crystals were mounted in rayon-fiber loops and placed into a N_2 stream at 120 K for data collection. A three-wavelength MAD experiment at 100 K was conducted on beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF) at Grenoble, using a ADSC Quantum-4 charge-coupled device as detector. The wavelengths for the MAD experiment were chosen by scanning through the absorption edge of the Se-OtsA crystal. MAD data, to 3.1 Å resolution, were collected at the wavelength corresponding to the minimum *f′*, the maximum *f″* and a reference wavelength at an energy above the absorption edge to maximize dispersive differences. A final higher resolution dataset, to 2.45 Å, was collected from a second crystal on the same beamline at the high-energy remote wavelength (Table 1). Data were processed and reduced using the HKL suite of programs [42]. All subsequent computing used the CCP4 suite [43], unless otherwise stated.

Structure Solution and Refinement

The three unmerged datasets for the selenium-MAD experiment were input to SOLVE [44]. Automated Patterson searching, using data to 2.8 Å, located 16 Se positions (SOLVE score 56.7) corresponding to four molecules of OtsA in the asymmetric unit. Initial phases at 2.45 Å resolution had a figure of merit of 0.4. SOLVE and RESOLVE were used to perform noncrystallographic symmetry averaging, using operators based upon the four sets of Se positions. The quality of the averaged map was greatly superior, reflected in the RESOLVE figure of merit of 0.56. The C-terminal domain, which contains all four Se sites, was traced in this averaged map using the X-AUTOFIT module in QUANTA (Accelrys, Inc. San Diego, CA). Partial structure refinement, including experimental phases as restraints, was performed using REFMAC [45], with five percent of the data first having been set aside for cross-validation [46]. Updated operators, based upon the C-terminal domain coordinates, were used to reaverage the electron density map with DM [47] and the resultant averaged, combined-phase map used to complete tracing of the structure. Further refinement maintained both phase and tight noncrystallographic symmetry restraints. The cross-validation reflections, as reflected in R_{free}, were used to monitor refinement strategies such as the strength of NCS restraint values and suggested tight restraints throughout.

Domain motion was incorporated through refinement of TLS parameters as implemented in REFMAC [31] for each domain, and refined TLS values are included in the PDB header. Ligands, both UDP and Glc-6-phosphate, were added and solvent-water molecules built into NCS averaged mF_{obs}-F_{calc} difference maps. A consequence of the TLS strategy is that ligands and solvent are excluded from the TLS treatment and consequently obtain temperature factors, which are artificially elevated compared to the surrounding protein atoms. Data and Refinement statistics are given in Table 1.

Acknowledgments

The Wellcome Trust and the Biotechnology and Biological Sciences Research Council are thanked for funding. The staff of the European Synchrotron Radiation Facility are acknowledged for provision of data collection facilities. Professors Michael Sinnott and Duilio Aragoni, and Dr. Spencer Williams are thanked for insightful discussions. G.J.D. is a Royal Society University Research Fellow.

Received: September 17, 2002
Revised: October 25, 2002
Accepted: October 25, 2002

References

1. Wingler, A. (2002). The function of trehalose biosynthesis in plants. Phytochemistry *60*, 437–440.
2. Singer, M.A., and Lindquist, S. (1998). Thermotolerance in Saccharomyces cerevisiae: the yin and yang of trehalose. Trends Biotechnol. *16*, 460–468.
3. Benaroudj, N., Lee, D.H., and Goldberg, A.L. (2001). Trehalose accumulation during cellular stress protects cells and cellular proteins from damage by oxygen radicals. J. Biol. Chem. *276*, 24261–24267.
4. Chen, Q., Ma, E., Behar, K.L., Xu, T., and Haddad, G.G. (2002). Role of trehalose phosphate synthase in anoxia tolerance and development in *Drosophila melanogaster*. J. Biol. Chem. *277*, 3274–3279.
5. Howells, A.M., Bullifent, H.L., Dhaliwal, K., Griffen, K., García de Castro, A., Frith, G., Tunnacliffe, A., and Titball, R.W. (2002). Role of Trehalose biosynthesis in environmental survival and virulence of *Salmonella enterica* serovar *typhimurium*. Res. Microbiol. *153*, 281–287.
6. Chatterjee, D. (1997). The mycobacterial cell wall: structure, biosynthesis and sites of drug action. Curr. Opin. Chem. Biol. *1*, 579–588.
7. De Smet, K.A.L., Weston, A., Brown, I.N., Young, D.B., and Robertson, B.D. (2000). Three pathways for trehalose biosynthesis in mycobacteria. Microbiology *146*, 199–208.
8. Coutinho, P.M., and Henrissat, B. (1999). Carbohydrate-Active Enzymes server at URL: http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html.
9. Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino-acid similarities. Biochem. J. *326*, 929–942.
10. Coutinho, P.M., and Henrissat, B. (1999). Carbohydrate-active enzymes: an integrated approach. In Recent Advances in Carbohydrate Engineering, H.J. Gilbert, G.J. Davies, B. Svensson, and B. Henrissat, eds. (Cambridge, UK: Royal Society of Chemistry), pp. 3–12.
11. Tarbouriech, N., Charnock, S.J., and Davies, G.J. (2002). Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. J. Mol. Biol. *314*, 665–661.
12. Davies, G.J. (2001). Sweet secrets of synthesis. Nat. Struct. Biol. *8*, 98–100.
13. Boix, E., Swaminathan, G.J., Zhang, Y., Natesh, R., Brew, K., and Acharya, K.R. (2001). Structure of UDP complex of UDP-galactose: β-galactoside-α-1,3-galactosyltransferase at 1.53-Å resolution reveals a conformational change in the catalytically important C-terminus. J. Biol. Chem. *276*, 48608–48614.
14. Vocadlo, D.J., Davies, G.J., Laine, R., and Withers, S.G. (2001). Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. Nature *412*, 835–838.
15. Zechel, D.L., and Withers, S.G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc. Chem. Res. *33*, 11–18.
16. Davies, G., Sinnott, M.L., and Withers, S.G. (1997). Glycosyl transfer. In Comprehensive Biological Catalysis, Volume 1, M.L. Sinnott, ed. (London: Academic Press) pp. 119–209.
17. Gastinel, L.N., Bignon, C., Misra, A.K., Hindsgaul, O., Shaper, J.H., and Joziasse, D.H. (2001). Bovine α-1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. EMBO J. *20*, 638–649.
18. Gibbons, B.J., Roach, P.J., and Hurley, T.D. (2002). Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. J. Mol. Biol. *319*, 463–477.
19. Waldscheck, B., Streiff, M., Notz, W., Kinzy, W., and Schmidt, R.R. (2001). α(1-3)-galactosyltransferase inhibition based on a new type of disubstrate analogue. Angew. Chem. Int. Ed. *40*, 4007–4011.
20. Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C.J. (2001). Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. Nat. Struct. Biol. *8*, 166–175.
21. Watson, K.A., McCleverty, C., Geramia, S., Cottaz, S., Driguez, H., and Johnson, L.N. (1999). Phosphorylase recognition and

phosphorolysis of its oligosaccharide substrate: answers to a long outstanding question. EMBO J. *18*, 4619–4632.
22. Patenaude, S.I., Seto, N.O.L., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M., and Evans, S.V. (2002). The structural basis for specificity in human ABO(H) blood group biosynthesis. Nat. Struct. Biol. *9*, 685–690.
23. Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. Structure *3*, 853–859.
24. Davies, G.J., and Henrissat, B. (2002). Structural enzymology of carbohydrate-active enzymes: implications for plant glycogenomics. Biochem. Soc. Trans. *30*, 291–297.
25. Bourne, Y., and Henrissat, B. (2001). Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr. Opin. Struct. Biol. *11*, 593–600.
26. Charnock, S., and Davies, G.J. (1999). The structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis,* in native and nucleotide-complexed forms. Biochemistry *38*, 6380–6385.
27. Vrielink, A., Rüger, W., Driessen, H.P.C., and Freemont, P.S. (1994). Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J. *13*, 3413–3422.
28. Wrabl, J.O., and Grishin, N.V. (2001). Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. J. Mol. Biol. *314*, 365–374.
29. Gibson, R.P., Lloyd, R.M., Charnock, S.J., and Davies, G.J. (2002). Characterisation of *E. coli* OtsA, a trehalose-6-phosphate synthase from glycosyltransferase family 20. Acta Crystallogr. D *58*, 349–351.
30. Bell, W., Sun, W.N., Hohmann, S., Wera, S., Reinders, A., De Virgilio, C., Wiemken, A., and Thevelein, J.M. (1998). Composition and functional analysis of the *Saccharomyces cerevisiae* trehalose synthase complex. J. Biol. Chem. *273*, 33311–33319.
31. Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta. Crystallogr. D *57*, 122–133.
32. Pan, Y.T., Drake, R.R., and Elbein, A.D. (1996). Trehalose-P synthase of mycobacteria: its substrate specificity is affected by polyanions. Glycobiology *6*, 453–461.
33. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
34. Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.
35. Campbell, R.E., Mosimann, S.C., Tanner, M.E., and Strynadka, N.C.J. (2000). The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferases. Biochemistry *39*, 14993–15001.
36. Ha, S., Walker, D., Shi, Y., and Walker, S. (2000). The 1.9A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan synthesis. Protein Sci. *9*, 1045–1052.
37. Mulichak, A.M., Losey, H.C., Walsh, C.T., and Garavito, R.M. (2001). Structure of the UDP-Glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. Structure *9*, 547–557.
38. Withers, S.G., Madsen, N.B., Sykes, B.D., Takagi, M., Shimomura, S., and Fukui, T., (1981). Evidence for direct phosphate-phosphate interaction between pyridoxal phosphate and substrate in the glycogen phosphorylase catalytic mechanism. J. Biol. Chem. *256*, 10759–10762.
39. Mitchell, E.P., Withers, S.G., Ermert, P., Vasella, A.T., Garman, E.F., Oikonomakos, N.G., and Johnson, L.N. (1996). Ternary complex crystal structures of glycogen phosphorylase with the transition state analogue nojirimycin tetrazole and phosphate in the T and R states. Biochemistry *35*, 7341–7355.
40. Sinnott, M.L., and Jencks, W.P. (1980). Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol. J. Am. Chem. Soc. *102*, 2026–2032.
41. Belisle, J.T., Vissa, V.D., Sievert, T., Takayama, K., Brennan, P.J., and Besra, G.S. (1997). Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. Science *276*, 1420–1422.
42. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
43. CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.
44. Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D *55*, 849–861.
45. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum likelihood method. Acta Crystallogr. D *53*, 240–255.
46. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.
47. Cowtan, K.D., and Main, P. (1996). Improvement of macromolecular electron density maps by the simultaneous application of real and reciprocal space constraints. Acta Crystallogr. D *49*, 148–157.
48. Sharp, K., Fine, R., and Honig, B. (1987). Computer-simulations of the diffusion of a substrate into the active-site of an enzyme. Science *236*, 1460–1463.
49. Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. J. Mol. Graph. *15*, 133–138.
50. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A *42*, 140–149.

Accession Numbers

Atomic coordinates and structure factors for the UDP/Glc-6-phosphate complex of OtsA have been deposited at the Macromolecular Structures Database with accession code (1GZ5.PDB).