Structure of a *Streptomyces* Maltosyltransferase GlgE: a Homologue of a Genetically Validated Anti-Tuberculosis Target*S*

Karl Syson‡, Clare E. M. Stevenson‡, Martin Rejzek‡, Shirley A. Fairhurst‡, Alap Nair‡, Celia J. Bruton‡, Robert A. Field‡, Keith F. Chater§, David M. Lawson‡, and Stephen Bornemann‡

From the ‡Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, Norfolk, NR4 7UH, United Kingdom, and the §Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, Norfolk, NR4 7UH, United Kingdom

Running title: Structure of maltosyltransferase GlgE

To whom correspondence may be addressed: Stephen Bornemann, Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, Norfolk, NR4 7UH, United Kingdom. Tel: +44 (0)1603 450741. Fax: +44 (0)1603 450018. E-mail: stephen.bornemann@jic.ac.uk.

**Background:** GlgE is a maltosyltransferase involved in bacterial α-glucan biosynthesis and is a genetically validated anti-tuberculosis target.

**Results:** We have determined the catalytic properties of *Streptomyces coelicolor* GlgE and solved its structure.

**Conclusions:** The enzyme has the same catalytic properties as *Mycobacterium tuberculosis* GlgE and the structure reveals how GlgE functions.

**Significance:** The structure will help guide the development of inhibitors with therapeutic potential.

GlgE is a recently identified (1→4)-α-D-glucan:phosphate α-D-maltosyltransferase involved in α-glucan biosynthesis in bacteria and is a genetically validated anti-tuberculosis drug target. It is a member of the GH13_3 CAZy sub-family for which no structures were previously known. We have solved the structure of GlgE isoform I from *Streptomyces coelicolor* and shown that this enzyme has the same catalytic and very similar kinetic properties to GlgE from *Mycobacterium tuberculosis*. The *S. coelicolor* enzyme forms a homodimer with each subunit comprising five domains including a core catalytic α-amylase-type domain A with a (β/α)₈ fold. This domain is elaborated with domain B and two inserts that are specifically configured to define a well conserved donor pocket capable of binding maltose. Domain A, together with domain N from the neighbouring subunit, forms a hydrophobic patch that is close to the maltose binding site and capable of binding cyclodextrins. Cyclodextrins competitively inhibit the binding of maltoligosaccharides to the *S. coelicolor* enzyme, showing that the hydrophobic patch overlaps with the acceptor binding site. This patch is incompletely conserved in the *M. tuberculosis* enzyme such that cyclodextrins do not inhibit this enzyme, despite acceptor length specificity being conserved. The crystal structure reveals two further domains, C and S, the latter being a helix bundle not previously reported in GH13 members. The structure provides a framework for understanding how GlgE functions and will help guide the development of inhibitors with therapeutic potential.

The crucial need to develop new drugs against tuberculosis (1), one of the world’s most pervasive and lethal infectious diseases (2), drives much research into the causative agent, *Mycobacterium tuberculosis*. In this context, we recently identified a new α-glucan pathway in this bacterium (Fig. 1 (3)). Its defining enzyme, GlgE, is a (1→4)-α-D-glucan:phosphate α-D-maltosyltransferase and member of the glycoside hydrolase family sub-family GH13_3 (4). It is capable of transferring maltosyl units not only from maltose 1-phosphate to maltoligosaccharides but also between maltoligosaccharides. We have genetically validated GlgE to be a potential new drug target (3) that has some attractive features as discussed at length elsewhere (5). The bactericidal mechanism of the blockage of GlgE is novel because rather than preventing the formation of an essential metabolic product, it is the auto-amplified build up of GlgE’s donor substrate, maltose 1-phosphate, which leads to pleiotropic effects, toxicity and cell death.

The GlgE pathway generates a branched α-glucan from trehalose (Fig. 1) (3). *M. tuberculosis* is known to produce three α-glucans; cytosolic
glycogen, capsular α-glucan and methylglucose lipopolysaccharide (6). These are either involved or implicated in the storage of carbon (7), evasion of the immune system (8-11) and chaperoning/regulating fatty acid biosynthesis (12), respectively. It is not yet known how much the GlgE pathway contributes to the biosynthesis of each of the three α-glucans. Nevertheless, synthetic lethality has been observed between the GlgE and methylglucose lipopolysaccharide pathways, implying the essentiality of at least one type of α-glucan and the role of GlgE in its biosynthesis (3).

The GlgE pathway is present in many other actinomycetes. For example, it is involved in carbon management in Streptomyces coelicolor (13-15). The genes of this pathway are duplicated and separately developmentally regulated in this organism, such that each is respectively associated with transient glycogen deposition at the initiation of aerial growth (phase I) and during the first stages of sporulation (phase II). The pathway is not restricted to actinomycetes and is remarkably widespread (6). Fourteen percent of sequenced microbial genomes contain all of the GlgE pathway genes, which are usually clustered, making the pathway half as common as the more well-known glycogen pathway involving GlgA and GlgC.

Structures have not previously been reported of GlgE or any other GH13_3 sub-family member. In parallel studies of the mycobacterial and Streptomyces GlgE enzymes, we have found that Streptomyces coelicolor GlgE isoform I is particularly amenable to structural analysis. This enzyme comprises domains in common with other members of the GH13 α-amylase family of enzymes together with a helix-bundle domain that is novel in this structural context. The location of the donor binding site has been defined together with a site capable of binding cyclodextrins that overlaps with the acceptor binding site. The structure is consistent with evidence that maltoligosaccharide acceptors are extended at their non-reducing ends. The S. coelicolor and M. tuberculosis GlgE enzymes have the same catalytic and very similar kinetic properties, with well conserved donor-binding sites. This allows the structure of the former to be used to guide inhibitor development for the latter in the search for new therapies against tuberculosis.

**EXPERIMENTAL PROCEDURES**

**Chemical Synthesis**—α and β-maltose 1-phosphate, 1a and 1b, were synthesised from 2,3,6,2',3',4',6'-hepta-O-acetyl-D-maltose (2) that was readily prepared from D-maltose using known procedures (see supplemental experimental procedures for details of preparation of 2) (16). α-Maltosyl fluoride (17) was prepared via 2,3,6,2',3',4',6'-hepta-O-acetyl-α-D-maltosyl fluoride using published procedures (18). TLC was performed on pre-coated silica plates (Merck 60 F254, 0.25 mm) containing a fluorescence indicator. Compounds were visualised under UV (254 nm) and/or by heating after dipping in a solution of 5% H2SO4 in ethanol. Flash column chromatography was performed on silica gel columns (Biotage KP-Sil™ silica, 60 Å, 32-63 μm) fitted to a Biotage SP1® Automated Purification System (Uppsala, Sweden). High resolution MS was carried out using a Thermo Fisher Scientific (Waltham MA, USA) LTQ Orbitrap XL. Low resolution mass spectra were recorded with a Thermo Fisher Scientific Finnigan LCQ Deca XP Plus ion trap mass spectrometer. 1H and 13C NMR spectra were recorded at 300 K on a Bruker Avance II 600 MHz spectrometer with Bruker TCI cryoprobe (Bruker Biospin Ltd.). Water peaks were suppressed with pre-saturation and data were analyzed with Topspin 2.1 software (Bruker Biospin Ltd.). Chemical shifts are reported in ppm relative to TMS (δH 0) or, for samples in D2O, residual water (δH 4.70). Full assignment of 1H and 13C spectra was achieved with the aid of COSY, DEPT, HSQC and HSQC-TOCSY experiments. 31P spectra were obtained at 161 MHz on a Bruker Avance III 400 MHz spectrometer with and without decoupling and were referenced with external D3PO4 (δP 0.0). J values are in Hz.

Dibenzyl 2,3,6,2',3',4',6'-Hepta-O-acetyl-α-D-maltosyl Phosphate (3a) and Dibenzyl 2,3,6,2',3',4',6'-Hepta-O-acetyl-β-D-maltosyl Phosphate (3b)—Lithium diisopropylamide [4.15 ml of a 2 M solution in THF/ethylebenzene (Sigma-Aldrich Company Ltd., Dorset, England), 8.3 mmol] was slowly added to a solution of anhydrous 2,3,6,2',3',4',6'-hepta-O-acetyl-D-maltose (2) (3.1 g, 4.9 mmol), in the minimum volume of absolute THF required to solubilise it.
(220 ml) at -80 °C under a N2 atmosphere. After 10 min of stirring, a solution of tetrabenzyl pyrophosphate (3.7 g, 6.9 mmol) in THF (20 ml) was slowly added with cooling. The mixture was stirred with continued cooling for 40 min before being allowed to slowly warm to +4 °C followed by stirring for 16 h. An off-white precipitate of Li2PO4(O)OBn was removed by filtration. The resulting solution was evaporated to dryness under reduced pressure and the product was re-dissolved in ethyl acetate (20 ml). The organic layer was washed with saturated aqueous NaHCO3 (20 ml) and then by stirring for 16 h. An off-white precipitate of pyrophosphate (3.7 g, 6.9 mmol) in THF (20 ml) was slowly added with cooling. The mixture was stirred with continued cooling for 40 min before being warmed to +10 °C. An off-white precipitate of pyrophosphate (3.7 g, 6.9 mmol) in THF (20 ml) was slowly added with cooling. The mixture was stirred with continued cooling for 40 min before being warmed to +10 °C.

The overall isolated yield of the anomeric mixture 3 was 51%, β:α ratio = 2:1 according to 31P NMR spectroscopy, m/z (HR ESI) 199.2397 ([M+Na]+); C40H49NaO21P requires 919.2396. The α and β phosphate anomers, 3a and 3b, were partially separated with the β anomer 3b eluting first. Fractions containing a given anomer were enriched by repeating the chromatographic step twice. Anomer-enriched samples were evaporated to dryness under reduced pressure. Each anomer was further purified by HPLC using a Phenomenex semi-preparative silica column (Luna 250 × 10 mm, 10 µm) fitted to a Dionex Ultimate 3000. Compounds were eluted with a 40-67% ethyl acetate gradient and monitored by UV absorbance at 265 nm giving the α anomer 3a, δα (600 MHz; CDCl3) 7.40-7.36 (10 H, m, 2 C6H5), 5.82 (1 H, dd, Jα,β 3.4, Jα,β 7.0, 1-H), 5.54 (1 H, dd, Jα,β 10.0, Jα,α 10.0, 3-H), 5.42 (1 H, d, Jα,β 4.0, 1'-H), 5.37 (1 H, dd, Jα,β 10.0, Jα,β 10.0, 3'-H), 5.13 (1 H, m, 4'-H), 5.09 (4 H, m, 2 CH2C6H5), 4.83 (1 H, dd, Jα,β 10.0, 2-H), 4.84 (1 H, dd, Jα,β 4.0, Jα,β 10.0, 2'-H), 4.30 (1 H, dd, Jα,β 12.4, 6'-H), 4.24 (1 H, dd, Jα,β 3.5, Jα,β 12.4, 6'-H), 4.11 (1 H, dd, Jα,β 3.5, Jα,β 12.4, 6'-H), 4.02 (1 H, dd, Jα,β 3.5, Jα,β 12.4, 6'-H), 4.00 (1 H, dd, Jα,β 10.0, Jα,β 10.0, 4'-H), 3.95 (1 H, m, 5'-H), 3.90 (1 H, m, 5'-H), 2.10 (3 H, s, CH3), 2.08 (3 H, s, CH3), 2.07 (3 H, s, CH3), 2.04 (3 H, s, CH3), 2.02 (3 H, s, CH3), 2.016 (3 H, s, CH3), 1.88 (3 H, s, CH3); δp (150 MHz; CDCl3) 170.3 (CO2CH3), 169.9 (2 C, CO2CH3), 169.8 (CO2CH3), 169.4 (CO2CH3), 128.8-128.0 (C6H5), 95.7 (1'-C), 93.8 (1'-C), 71.8 (4'-C), 71.7 (3'-C), 70.0 (2'-C), 69.74 (CH2C6H5), 69.69 (5'-C), 69.3 (3'-C), 68.6 (5'-C), 67.9 (4'-C), 62.1 (6'-C), 61.3 (6'-C), 29.7 (CH3), 20.9 (CH3), 20.72 (CH3), 20.69 (CH3), 20.64 (CH3), 20.61 (CH3), 20.3 (CH3); δp (161 MHz; CDCl3) -2.8 (2H, dd, J1H,β 7.0, Jα,β ~7.8); m/z (ESI) 919.4 ([M+Na]+, 100%), 641.5 (9), MS2 641.1 (- (BnO)2PO2H), MS3 581.0 (- CH3CO2H); [α]20 + 70.7 (c 0.92, CHCl3); Rf (TLC; ethyl acetate:petroleum ether, 2:1, v/v) 0.70, and the β anomer 3b, δh (600 MHz; CDCl3) 7.36-7.29 (10 H, m, 2 C6H5), 5.41 (1 H, d, Jα,β 4.0, 1'-H), 5.38 (1 H, dd, Jα,β 7.4, Jα,β 7.4, 1-H), 5.35 (1 H, dd, Jα,β 7.0, Jα,β 9.0, 3-H), 5.06 (1 H, m, 4'-H), 5.02 (4 H, m, 2 CH2C6H5), 4.96 (1 H, dd, Jα,β 7.4, Jα,β 9.0, 2-H), 4.86 (1 H, dd, Jα,β 7.4, Jα,β 10.0, 2'-H), 4.50 (1 H, dd, Jα,β 6.5, Jα,β 12.2, 2-H), 4.05 (1 H, dd, Jα,β 4.0, Jα,β 12.2, 2'-H), 4.04 (1 H, m, 4-H), 3.94 (1 H, m, 5'-H), 3.78 (1 H, m, 5-H), 2.10 (3 H, s, CH3), 2.049 (3 H, s, CH3), 2.047 (3 H, s, CH3), 2.03 (3 H, s, CH3), 2.004 (3 H, s, CH3), 2.002 (3 H, s, CH3), 1.87 (3 H, s, CH3); δc (150 MHz; CDCl3) 170.6 (CO2CH3), 170.5 (CO2CH3), 170.3 (CO2CH3), 169.9 (CO2CH3), 169.8 (CO2CH3), 169.4 (CO2CH3), 128.8-128.0 (C6H5), 95.8 (1'-C), 95.86 (1'-C), 74.8 (3'-C), 73.0 (5-C), 72.3 (4-C), 73.0 (5-C), 72.1 (2-C), 70.1 (2'-C), 70.0 (CH2C6H5), 69.3 (3'-C), 68.6 (5'-C), 68.0 (4'-C), 62.2 (6-C), 61.4 (6'-C), 20.8 (CH3), 20.7 (CH3), 20.62 (CH3), 20.61 (CH3), 20.59 (CH3), 20.58 (CH3), 20.4 (CH3); δp (161 MHz; CDCl3) -3.2 (dd, J1H,β 7.4, Jα,β ~7.8); m/z (ESI) 919.4 ([M+Na]+, 100%), 641.5 (21), MS2 641.1 (- (BnO)2PO2H), MS3 581.1 (- CH3CO2H); [α]20 + 70.7 (c 1.1, CHCl3); Rf (TLC; ethyl acetate:petroleum ether, 2:1, v/v) 0.71.

α-D-Maltose 1-Phosphate, Disodium Salt (1a)—Pd/C catalyst (10 wt. %, ~15 mg) was added to a solution of dibenzyl 2,3,6,2'-3,4',6'-hepta-O-acetyl-α-D-maltosyl phosphate (3a) (57 mg, 64 µmol) in methanol (20 ml). The atmosphere within the flask was replaced with H2 gas and the reaction mixture was stirred vigorously for 24 h at ambient temperature. The reaction went to completion according to TLC. The mixture was filtered and three drops of triethylamine were added. The solution was evaporated to dryness under reduced pressure and the product was re-dissolved in ethyl acetate (20 ml).
(TLC; dichloromethane: methanol: water, 6:3:1, v/v) 0.94, which was used directly in the next steps. Compound 4a was dissolved in methanol/water/triethylamine (7:3:1; 10 ml) and stirred at ambient temperature for 24 h. The reaction went to completion according to TLC and the mixture was evaporated to dryness to yield α-D-maltosyl phosphate, bis(triethylammonium) salt (5a) as a white solid. The sample was dissolved in H2O (1 ml) applied to a Dowex Marathon C column (Na+ form), eluted with water (10 ml) and freeze-dried to yield 1a as a white solid, δH (600 MHz; D2O) 5.36 (1 H, d, J1,P 3.0, J1,P 7.0, 1-H), 5.33 (1 H, d, J1,P 3.4, 1'-H), 3.94 (1 H, m, 3-H), 3.94 (1 H, m, 5-H), 3.79 (1 H, d, J6a,6b 12.8, 6a-H), 3.76 (1 H, dd, J6a,6b 12.6, 6a-H), 3.70 (1 H, d, J6a,6b 12.8, 6b-H), 3.67 (2 H, m, 5', 6'b-H), 3.61 (1 H, m, 3'-H), 3.55 (1 H, m, 4-H), 3.47 (1 H, d, J6a,6b 12.8, 6'-H), 3.31 (1 H, m, 5'-H), 3.62 (1 H, m, 6b-H), 3.60 (1 H, m, 3'-H), 3.54 (1 H, m, 5'-H), 3.49 (1 H, d, J9,9 4.0, 4-H), 3.48 (1 H, d, J9,9 4.6, 6'-H), 3.40 (1 H, m, 3'-H), 3.31 (1 H, d, J9,9 4.6, 4'-H), 3.25 (1 H, dd, J1,P 7.7, J2,P 10.0, 2-H); δC (150 MHz; D2O) 99.5 (1'P, 1-C), 77.1 (4-C), 75.9 (3-C), 74.2 (5-C), 72.8 (3'-C), 72.6 (5'-C), 72.1 (P, 7.7, 1-H), 67.8 (4-C), 73.6 (3-C), 72.8 (3'-C), 72.6 (5'-C), 72.1 (P, 6.6, 2-C), 71.8 (2'-C), 70.4 (5-C), 69.3 (4'-C), 60.7 (6-C), 60.4 (6'-C); δP (161 MHz; D2O) 2.1 (Jp,H1 7.0); m/z (ESI) 421.2 ([(R-OPO3Na2+)+H]+; C12H22Na2O14P2 requires 421.0537; m/z (ESI) 467.0537; m/z (ESI) 467.0537); m/z (ESI) 421.2 ([(R-OPO3H2-]) with 499 expected).

**Expression and Purification of GlgE**—The genes for both isoforms of GlgE from *S. coelicolor* strain M145 were each sub-cloned into a pET15b vector using BamHI and NdeI restriction sites to allow the expression of the enzyme with an N-terminal His-tag and thrombin-cleavage site. Both *glgE* genes in the final expression plasmids were confirmed by DNA sequencing. Protein expression was carried out as described previously (3) except that *Escherichia coli* BL21(DE3) pLysS was used. Selenomethionine-labelled GlgE isoform I was obtained by the metabolic inhibition method (20). The method used to express GlgE from *M. tuberculosis* (3) was also used to express *M. smegmatis* GlgE. The enzymes were purified using nickel-affinity and size exclusion chromatographies (3).

**Assay of GlgE Activity**—GlgE activity was monitored using a quantitative stopped assay to determine Pi release with malachite green (3). Reaction mixtures comprised enzyme, substrates and 100 mM bis-Tris phosphate, pH 7.0, containing 50 mM NaCl at 30 °C. Reactions were monitored over an 8 min period and progressed linearly with time for at least 4 min when donor consumption was typically < 5%. Acceptor preferences were determined in triplicate using 7.5 mM maltohexaose and between 22 and 80 nM enzyme. Activity was also monitored qualitatively using MALDI TOF MS to detect extension of maltohexaose (3). Product oligosaccharide linkage analysis was carried out using reaction mixtures that were quenched by heating to 99 °C for 15 min before being subjected to 1H NMR spectroscopy at 600 MHz (3). Pi was removed from buffers using an ‘Pi mop’ consisting of bacterial purine nucleotide phosphorylase and 1 mM 7-methylguanosine (21). α-Maltose 1-phosphate was generated from α-maltosyl fluoride (10 mM) in the presence of Pi (50 mM) and enzyme, and identified using MALDI TOF MS; m/z 499 ([(R-OPO3K2+)+H]+ with 499 expected).

**Protein Size Determination**—Size exclusion chromatography was carried out using a Superdex 200 10/300 column (GE Healthcare, Amersham, United Kingdom) using 100 mM bis-Tris propane buffer.
buffer, pH 7.0 containing 50 mM NaCl. Dynamic light scattering was carried out using a Dynapro Titan molecular-sizing instrument at 298 K (Wyatt Technology) with an enzyme concentration of 2 mg ml⁻¹ in the buffer described above. Data were analysed using the DYNAMICS software package (Wyatt Technology). Analytical ultracentrifugation (AUC) experiments were performed using a Beckman Optima XL-I analytical ultracentrifuge (High Wycombe, United Kingdom) equipped with absorbance optics and an An-50 Ti rotor. Experiments were performed at 20 °C and 10,000 r.p.m. with a protein concentration of 1 mg ml⁻¹ in 20 mM bis-Tris propane, pH 7.0, containing 100 mM NaCl. The partial specific volume of GlgE was calculated from the amino acid sequence using SEDNTERP (www.jphilo.mailway.com). Ultrascan II (www.ultrascan.uthscsa.edu) was used to fit the experimental sedimentation equilibrium profiles to a single species model.

Protein Crystallization and Cryoprotection—Crystallization screens and optimizations were performed using a protein concentration of ~5 mg ml⁻¹ and a temperature of 20 °C. Crystals of GlgE (both apo and SeMet-labelled) were obtained from 15% (w/v) polyethylene glycol (PEG) 3350, 0.2 M sodium citrate and 15% ethylene glycol. Ligand-bound structures of GlgE were obtained by co-crystallization under the same conditions with ligand concentrations of 5 mM. Ligand-bound structures of GlgE were obtained by co-crystallization under the same conditions with ligand concentrations of 5 mM.

Structure Determination and Refinement—All crystals were flash-cooled in Litholoops (Molecular Dimensions) by plunging into liquid nitrogen and transported in Unipuck cassettes before being robotically mounted onto the goniostat on either station I02, I03 or I04 at the Diamond Light Source (Oxford, UK), whereupon they were maintained at -173 °C with a Cryojet cryocooler (Oxford Instruments). Diffraction data were recorded using an ADSC Quantum 315 CCD detector. The resultant data were integrated using MOSFLM (22) and scaled with SCALA (23). Analysis in POINTLESS (23) suggested that the space group was P4₁2₁2/P4₃2₁2, although statistical tests in TRUNCATE (24) indicated that the crystals were usually hemihedrally twinned (operator: k, h, -l), and must therefore belong to a lower symmetry space group. Nevertheless, it proved to be more tractable to determine experimental phases and build a preliminary model in P4₁2₁2/P4₃2₁2.

A three-wavelength anomalous dispersion data set was collected to 2.8 Å resolution from a single crystal of SeMet-substituted protein (Table S1). The data were processed in space group P422 with approximate cell parameters of a = b = 113.8 Å, c = 316.7 Å. Experimental phases were determined using the SHELX suite (25). The two possible enantiomorphs gave comparable statistics, however, P4₁2₁2 was ultimately chosen based on superior electron density map quality. SHELXD located 15 selenium sites, being consistent with two copies of the GlgE protomer (based on 8 methionines per subunit) in the asymmetric unit (ASU), with a corresponding solvent content of 63% (based on a subunit molecular mass of 75,290 Da). After phasing with SHELXE and density modification, with two-fold non-crystallographic symmetry (NCS) averaging in DM (26), the figure-of-merit was 0.794 to 2.8 Å resolution. After automated building with BUCCANEER (27) and several iterations of: (i) rebuilding in COOT (28), (ii) restrained refinement against the SeMet peak data set with REFMAC5 (29), (iii) combination of experimental and model phases using SIGMAA (30), and (iv) two-fold averaging in DM, a model comprising 1062 residues with corresponding R_work and R_free values of 36.8% and 40.0%, respectively, at 2.8 Å resolution, was produced.

Several data sets were collected from crystals obtained by co-crystallization with potential ligands. Of these, only the complex with α-cyclodextrin alone yielded a data set that was essentially untwinned and, therefore, could be justifiably treated as belonging to space group P4₁2₁2. This data set was collected to 2.3 Å resolution and was used to complete the building and refinement of the first GlgE model. This was performed with REFMAC5 using two-fold NCS restraints, and TLS parameters (four domains per monomer). The final model consisted of 1298 residues in two subunits and two α-cyclodextrin molecules, having final R_work and R_free values of 17.3% and 20.1%, respectively (Table 2). From inspection, the biological unit of GlgE is a homodimer. However, the two subunits in the ASU of this model represent halves of two separate dimers, with individual dimers being completed through the application of two-fold crystallographic symmetry.

The highest resolution data set was collected
from a crystal obtained by co-crystallization with 6'-α-D-glucosyl-maltotriose (Megazyme, Bray, Ireland) and processed to 1.8 Å resolution. The structure was solved using PHASER (31) with a single GlgE subunit from the α-cyclodextrin complex as the search model. Although the expectation was that the space group would be primitive tetragonal, acceptable solutions were found in space groups P212121 and C2221, in addition to P41, in each case giving four subunits per ASU (arranged as two dimers) and very similar crystal packing. However, the log-likelihood-gain value (calculated to 3 Å resolution) was higher for the P212121 solution and it gave a lower clashscore in MOLPROBITY (32). Therefore, P212121 was chosen for refinement of the model in REFMAC5, employing intensity-based twin refinement, four-fold NCS restraints, and TLS parameters (four domains per monomer). Unfortunately, electron density maps were heavily biased due to the high twin fraction (0.48) (33). Thus, model building was performed cautiously and water molecules were added sparingly. No evidence was seen for the added ligand and, therefore, this model was subsequently treated as a reference apo structure. Data sets from three further co-crystallizations yielded new complexes. These were all twinned, and handled as for the apo structure, which was also used as the starting point for refinement in each case. The X-ray data collection and refinement statistics for all structures are summarized in Table 2. Structural figures were generated using PyMOL (34).

RESULTS

Synthesis of Maltose 1-Phosphate—In order to assay GlgE activity, it was necessary to obtain the donor substrate, α-maltose 1-phosphate 1a. A protection-deprotection strategy was used to allow the phosphorylation of the 1-position of maltose using tetrabenzyl pyrophosphate (Fig. 2). This yielded a mixed anomer product 3, from which pure anomers were obtained using silica column chromatography. Following deprotection, this route allowed the production of both α and β-maltose 1-phosphate, 1a and 1b. The properties of the synthetic α anomer 1a (NMR and MS spectra, and TLC Rf) were indistinguishable from those of the material obtained from M. smegmatis assigned as being α-maltose 1-phosphate (3).

Crystallization of GlgE—Recombinant GlgE from M. tuberculosis and M. smegmatis were subjected to crystallisation trials but failed to yield protein crystals. GlgE isoforms I and II from another actinomycete, S. coelicolor, were subsequently entered into trials. Although isoform II proved to be too insoluble to obtain crystals, isoform I readily yielded crystals.

A Comparison of the Catalytic and Kinetic Properties of S. coelicolor GlgE Isoform I and M. tuberculosis GlgE—Before embarking on solving the structure of S. coelicolor GlgE isoform I, it was important to determine its properties and compare them with those of GlgE from M. tuberculosis. While it seems likely that homologous enzymes from actinomycetes would share similar properties, GlgE isoforms I and II from S. coelicolor (with 86% amino acid sequence identity between them) each share only 51% identity with GlgE from M. tuberculosis.

GlgE isoform I from S. coelicolor was heterologously expressed with an N-terminal His-tag in E. coli and purified to homogeneity. According to assays based on Pi release, it possessed GlgE activity. While the pH optimum (7.0; Fig. S1) and slight activation by NaCl (~20% at 50 mM; Fig. S2) were common to isoform I and GlgE from M. tuberculosis, their temperature optima reflected the lifestyles of the source organisms [~30 °C for S. coelicolor isoform I (Fig. S3) and ~37 °C for M. tuberculosis GlgE (3)]. The acceptor preferences of these two enzymes were similar such that a degree of polymerisation (DP) of ≥ 4 gave the most significant rates of reaction (Fig. 3 and Fig. S4). The acceptor length specificities were very similar, with only a marginal shift of the optimum from DP 5 to 6 in isoform I. Isoform II from S. coelicolor behaved very similarly to isoform I as would be expected given their high sequence identities (Fig. 3).

The $K_m^{app}$ values for α-maltose 1-phosphate with isoform I and the M. tuberculosis enzyme were very similar (0.25 ± 0.05 and 0.30 ± 0.06 mM in the presence of 1 mM maltotetraose; Table 1 and Fig. S5). (It is noteworthy that an enzyme activity consistent with GlgE detected in Mycobacterium smegmatis extracts exhibited a comparable $K_m^{app}$ for the donor substrate of 0.25 mM using glycogen as the acceptor (35).) Isoform I had $k_{cat}^{app}$ values up to an order of magnitude greater rate than those of the M. tuberculosis
enzyme. Both $K_m^{\text{app}}$ and $k_{cat}^{\text{app}}$ for α-maltose 1-phosphate increased with increasing maltohexaose concentration consistent with isoenzyme I (data not shown), consistent with a ping-pong (substituted) enzyme mechanism. The $K_m^{\text{app}}$ for maltohexaose in the presence of 5 mM α-maltose 1-phosphate was 23-fold lower with isoenzyme I. In general, the Michaelis-Menten parameters for isoenzyme II were broadly similar to those of isoenzyme I, except that the $k_{cat}^{\text{app}}/K_m^{\text{app}}$ values were between ~3 and ~5-fold lower.

Isoform I formed exclusively α-1,4 linkages according to NMR spectroscopy (Fig. S6) and neither β-maltose 1-phosphate nor α-D-glucose 1-phosphate served as donor substrates. It catalyzed disproportionation reactions through maltosyl transfer between maltooligosaccharides (Fig. S7) with chain length specificities indistinguishable from those of GlgE from *M. tuberculosis* (donor DP ≥4 but preferentially ≥6 with acceptor DP ≥4 (3)). Disproportionation occurred just as efficiently in the presence of a ‘Pi mop’ consisting of purine nucleoside phosphorylase and 7-methylguanosine (21), providing evidence that maltosyl transfer occurred directly from donor to acceptor rather than via a maltose 1-phosphate intermediate.

The above analyses showed that GlgE isoenzyme I from *S. coelicolor* is a (1→4)-α-D-glucan:phosphate α-D-maltosyltransferase that has very similar kinetic properties to GlgE from *M. tuberculosis*. The only differences were in $k_{cat}^{\text{app}}$, $K_m^{\text{app}}$ for maltohexaose and temperature optimum together with a small shift in the acceptor chain length specificity.

*S. coelicolor* GlgE Isoform I Extends a Primer at its Non-Reducing End—In the absence of a primer accepting, GlgE forms only very small amounts of oligomeric product after many hours of incubation with maltose 1-phosphate (data not shown). It is likely that this occurs via hydrolysis of maltose 1-phosphate and extension of the resulting maltose; both very slow processes. Therefore, self priming, though possible, is not efficient with GlgE.

In order to test how GlgE extends acceptors, maltotetraol (Fig. 4a), which has no reducing end, was exposed to isoenzyme I and α-maltose 1-phosphate. Maltotetraol could be detected using MALDI TOF MS before the addition of enzyme (Fig. 4b). After the addition of enzyme, a series of products were observed with masses consistent with maltotetraol-extension by one maltosyl unit at a time (Fig. 4c). 1H NMR spectroscopy before and after the addition of maltose 1-phosphate and enzyme (Fig. 4d and Fig. 4e, respectively) showed a net ~3-fold increase in normal α-1,4 linkages consistent with extension at the non-reducing end of maltotetraol. There was very little reducing end generated in the reaction mixture (6% reducing end α resonance at ~5.25 ppm compared with that of the α-glucosidic link to the glucitol moiety at ~5.08 ppm) and these were likely formed by slow hydrolytic side reactions. The Pi-release assay indicated that maltose was transferred to maltotetraol at a rate 35% of that with of maltotetraose implying a +4 sub-site has a preference for a glycopyranose ring over the ring-opened glucitol. Overall, these observations strongly support the preference for a maltooligosaccharide acceptor that is extended at its non-reducing end by GlgE.

**α-Maltosyl Fluoride is an Efficient Donor**—It is noteworthy that α-maltosyl fluoride, which bears a better leaving group than the normal substrate, was a donor (Fig. 5a) and was able to extend maltotetraose to give longer products than α-maltose 1-phosphate under the same conditions (Fig. 5b). Products from α-maltosyl fluoride of >DP34 (well beyond the limit of aqueous solubility of ~DP18) were conspicuous because the solution became visibly white and turbid. This donor was also capable of generating maltose 1-phosphate, according to MS, in the presence of enzyme and Pi but in the absence of an acceptor (data not shown). It could therefore be of utility in the enzymatic synthesis of α-maltose 1-phosphate and maltooligosaccharides as well as in monitoring GlgE activity using a fluoride electrode.

**Solving the Structure of S. coelicolor GlgE Isoform I**—The structure of ligand-free apo-GlgE was determined by the multi-wavelength anomalous dispersion method using selenomethionine-substituted protein (Table S1). A number of ligand-bound structures were subsequently obtained by co-crystallisation. Although the majority of data sets were hemihedrally twinned, structure solution and refinement was achievable (Table 2).
was also a dimer in solution (172 kDa by analytical ultracentrifugation, 120 kDa by size exclusion chromatography and 103 kDa by dynamic light scattering with 151 kDa predicted for the His-tagged dimer; data not shown). The *M. tuberculosis* and *M. smegmatis* enzymes also formed dimers in solution according to analytical ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size formed dimers in solution according to analytical

ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size

ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size

was also a dimer in solution (172 kDa by analytical ultracentrifugation, 120 kDa by size exclusion chromatography and 103 kDa by dynamic light scattering with 151 kDa predicted for the His-tagged dimer; data not shown). The *M. tuberculosis* and *M. smegmatis* enzymes also formed dimers in solution according to analytical ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size formed dimers in solution according to analytical ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size formed dimers in solution according to analytical ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size formed dimers in solution according to analytical ultracentrifugation (K. Syson and S. Bornemann, unpublished). Some variance in the oligomer size
end sugar ring (Fig. 7a). The conformation was similar to that of the major species found in solution (42), indicating a low energy conformation of maltose bound to GlgE. Despite maltose being present as a mixture of α and β anomers in solution (with an α:β ratio of ~1:2 at equilibrium according to NMR spectroscopy; F. Miah and S. Bornemann, unpublished), the enzyme bound the α anomer, consistent with this pocket being tailored to bind, break and make α-1-linked bonds. The orientation of the maltose, compared with other ligand-bound structures of the GH13 family, is consistent with it being the donor pocket comprising -1 and -2 sugar-binding sub-sites (37). The reducing end of maltose sits between Asp394 and Glu423 (Fig. 7a). Using sequence and structural comparisons with other family members, these residues of GlgE are predicted to be the nucleophile/base and proton donor, respectively (Fig. 8), associated with the typical double displacement mechanism of such retaining enzymes (43). The mean distance between the carboxyl side chain oxygen atoms of Asp 394 and Glu 423 was 4.9 Å. This is within the range observed in other retaining glycosidases of 4.8 to 5.3 Å and contrasts with that of inverting glycosidases of 9.0 to 9.5 Å (43).

The -1 sub-site is lined with amino acids side chains that include Asp480, which forms hydrogen bonding interactions with the C-2 and C-3 hydroxyls of the reducing end sugar of maltose (Fig. 7 and Fig. S10). A carboxylate side chain in this position is highly conserved within this enzyme family and is thought to assist in catalysis by stabilizing the oxocarbenium ion-like transition state and also for maintaining the Glu base in the correct protonation state (37). The maltose molecule is sandwiched between the hydrophobic side chains of Trp281 from insert 1 and Tyr357 from the domain B lid. Thus domain A, domain B and, to a lesser extent, insert 1 have a role in defining sub-site -1.

Sub-site -2 is defined by domain B and inserts 1 and 2 (Fig. S10). There is no sub-site -3 because of the presence of domain B and insert 1 within this region of the protein providing a reason why GlgE is specific for maltose as the donor. Overall donor specificity is therefore defined by domains A and B and inserts 1 and 2, a typical arrangement that determines specificity in GH13 enzymes (37-38).

The location of the +1 sub-site can be predicted to be adjacent to the -1 sub-site, projecting from the reducing end anomeric α-hydroxyl of maltose and by analogy with other family members. This site must be able to bind the phosphate of α-maltose 1-phosphate, promote its cleavage and yet also be able to bind and deprotonate the non-reducing end of an acceptor maltooligosaccharide without activating water. Polar residues likely to define the phosphate-binding site include Asn352 and Tyr357 of the domain B lid as well as other candidates from domain A (Fig. S10).

The Acceptor Site—In order to define the site where an acceptor binds, the protein was crystallized in the presence of maltooligosaccharides and analogues thereof. However, no extra density was observed in structures solved from co-crystallizations with either maltotriose, 6'-α-D-glucosyl-maltotriose (which yielded the apo-GlgE structure) or acarbose, for example. This is perhaps not surprising given that they are neither acceptors nor inhibitors. Maltotetraose, maltopentaose and maltooctaose each gave ligand-bound structures (data not shown), but they were all indistinguishable from the mal-GlgE structure. It would appear that over the time-scale of the crystallization, GlgE hydrolyzed these oligomers to generate sufficient maltose to occupy the donor pocket.

The interaction of cyclodextrins (cyclic maltooligosaccharides) with GlgE was then explored. According to MALDI TOF MS, cyclodextrins were not converted to any products by GlgE. However, α-cyclodextrin was shown to inhibit the extension of 1 mM maltohexaose with an IC50 of ~19 mM, according to the Pi-release assay (Fig. S11a). Both β and γ-cyclodextrins were also inhibitory, each with an IC50 value of ~6 mM (Fig. S11b and Fig. S11c). Their lower IC50 values suggest slightly more favourable protein contacts with the larger diameter cyclodextrins. The dependence of inhibition by α-cyclodextrin on donor and acceptor concentrations was then tested. The percentage inhibition almost halved when the acceptor concentration was increased four-fold (Fig. S12). This is consistent with α-cyclodextrin competing with the acceptor for a common binding site on GlgE. Inhibition was more pronounced when the donor concentration was increased four-fold. This is consistent with an
increase in $K_m$ for the acceptor when the donor concentration increases in a ping-pong reaction, allowing the inhibitor to compete with the donor more effectively. These observations strongly suggest that the acceptor binding site overlaps with the α-cyclodextrin binding site of the S. coelicolor enzyme. Interestingly, the M. tuberculosis GlgE enzyme was not significantly inhibited by the cyclodextrins in the concentration range tested (data not shown).

Co-crystallization of GlgE with α-cyclodextrin yielded a ligand-bound structure, αCD-GlgE, that was solved to 2.3 Å resolution. There were no significant changes within the protein compared with the apo-GlgE and mal-GlgE structures. The α-cyclodextrin was bound to a largely hydrophobic ridge near the donor pocket (Fig. 7b). This ridge comprises largely non-polar side chains of domain A and Gly84 of domain N of the neighbouring subunit (Fig. 7b, Fig. 9b and Fig. S10b). Thus domain N not only participates in enzyme dimerization but also appears to be involved in specificity. Similar roles for domain N have been identified in a maltogenic amylase from Thermus sp. (44), despite its role in other enzymes being unclear (37). The orientation of the cyclodextrin-GlgE interaction was close to and parallel to the linear binding cleft, near to the predicted +1 subsite and roughly orthogonal to the orientation of the maltose (Fig. 6b and Fig. 7c).

There are two additional features either side of the cyclodextrin binding patch worthy of note. There is a linear cleft that extends from the exit of the donor pocket and through what is predicted to be the +1 subsite (Fig. 6b and Fig. 7c). It is defined by domains A and B at its origin and extends between domains N and S of the neighbouring subunit. There is also a diagonal cleft that runs across both subunits of the dimer and intersects both of the linear clefts at the points where they exit the protein (Fig. 6b). These clefts could therefore be involved in binding a growing α-glucan chain.

Co-crystallization of GlgE with α-cyclodextrin and maltose yielded a structure showing density for both ligands consistent with both individual ligand-bound structures (data not shown). However, the highest resolution structure with both of these ligands bound, αCD-mal-GlgE, happened to be obtained from a co-crystallization with α-cyclodextrin and maltodextrin that was solved to 2.2 Å resolution. Co-crystallization of GlgE with β-cyclodextrin yielded a ligand-bound structure, βCD-mal-GlgE, that was solved to 2.5 Å resolution. The β-cyclodextrin interacted with GlgE in a manner very similar to that of α-cyclodextrin (Fig. S9c). Electron density within the donor pocket was consistent with the presence of maltose, which presumably was a contaminant from the β-cyclodextrin.

**DISCUSSION**

**Relationship Between GlgE Activity, GlgE Structure and GH13_3 Membership**—We have determined the structure of GlgE isoform I from S. coelicolor, which is the first example from the GH13_3 sub-family (4). There are a large number of structures of other GH13 sub-family members in the PDB database (145 non-redundant structures similar to GlgE with a Z score of $\geq 10$ according to DALI (39)). The S domain is a novel feature of GlgE, and the particular configuration of domain B and inserts 1 and 2 is specific to GlgE. For example, the protein with the most similar structure to GlgE according to both DALI (39) and SSM (40) is annotated as an α-amylase from Lactobacillus plantarum (PDB code 3dhu). Despite it possessing a B domain and inserts 1 and 2 at the same junctions of domain A they are different in length, sequence and conformation such that there is almost no conservation of the residues defining the maltose binding site. When only domains A and B together with inserts 1 and 2 of GlgE were used as the query, DALI gave the same top hit. SSM gave Thermatoga maritima 4-α-glucanotransferase (PDB code 1lwj (45)), which again has significantly different elaborations of its active site. Inspection of other relevant and high scoring hits revealed even greater structural diversity in and around their active sites; e.g. maltogenic amylase, which binds maltose in its -1 and -2 sub-sites (GH13_20; PDB code 1gvi (46)) and amylomaltase, which generates an α-1,4 glucan polymer (GH13_4; PDB code 1g5a (47)).

We have failed to detect GlgE activity in other GH13 enzymes that are capable of disproportionating maltoligosaccharides (K. Syson and S. Bornemann, unpublished), such as Thermatoga maritima maltosyltransferase (48). Therefore the ability to use maltose 1-phosphate as a donor may be restricted to members of the
GH13_3 sub-family. The majority of glgE genes are clustered with either one or all of the other genes of the GlgE pathway (6). In addition, there is substantial overlap between the set of proteins encoded by these genes and those defined as GH13_3 members by the CAZy database. This lends weight to the likelihood that most, if not all, GH13_3 sub-family members have GlgE activity.

Catalytic Centre—There was some doubt about the presence of the entire catalytic machinery in the GH13_3 sub-family of proteins (4) before GlgE was discovered (3). However the key side chains can now be clearly identified in the structure of GlgE, whereby Asp394 and Glu423 are well placed to carry out the roles of nucleophile/base and proton donor, respectively (Fig 7a and Fig. 8). This arrangement is consistent with the evidence for extension of acceptors at their non-reducing ends and the ability of GlgE to use α–maltosyl fluoride as a donor.

GlgE catalyzes glycosyl transfer reactions to acceptors other than water despite it being a GH class member. Although some phosphorylases are members of the GT class (e.g. GT35 glycogen phosphorylase with a distinct GT_B fold) there are examples of others in the GH class (49), such as GH13_18 sucrose phosphorylase (4,50). The way in which sucrose phosphorylase’s +1 sub-site is tailored to bind phosphate as a leaving group and yet also accept sugar acceptors involves local conformational changes (51). Whether this is also the case with GlgE remains to be seen because the +1 sub-sites of these two enzymes are very different. While it is not clear from our structures how GlgE kinetically suppresses hydrolytic reactions, the way other phosphorylase-type enzymes achieve this appears equally elusive at this time.

It is noteworthy that phosphorylases are generally associated with phosphorylisis rather than saccharide polymerisation primarily due to a relatively high cytosolic concentration of Pi. However, flux through the GlgE pathway has been demonstrated (3) and this is presumably driven principally by the ATP-requiring maltose kinase step preceding GlgE (see the supplemental discussion for a more extensive consideration of equilibria within the GlgE pathway).

Binding of Substrates—While the donor pocket of GlgE is well conserved and highly tailored to bind maltose, it is less clear what defines acceptor specificity. The general location of the +1 sub-site can be identified by inspection of the trajectory of the reducing end of maltose as it emerges from the donor pocket (Fig. 7c and Fig. S10). The observation that cyclodextrins compete with linear maltooligosaccharide acceptors provides strong evidence that their binding sites overlap. Thus some of the +n sub-sites are likely to be located in or very near the cyclodextrin binding patch, which is ~12 Å (the equivalent of ~3 sub-sites) away from the +1 sub-site. It is possible that linear acceptors bind in the same orientation as cyclodextrins, with identical sugar-protein interactions. This orientation is certainly consistent with acceptors being extended at their non-reducing ends. However, in order to connect the donor and acceptor sub-sites, there would have to be a significant bend in the acceptor, for which there is some precedence in GH13 glucan-binding sites (45). Alternatively, it is possible that the cyclodextrins, which are conformationally restricted, bind in a different orientation to that of acceptors. For example, linear acceptors could bind in an orientation orthogonal to that of cyclodextrins, removing the need for a bend. Other GH13 enzymes have binding sites in such an orientation (e.g. porcine pancreatic α-amylase isozyme II complexed with trestatin A-derived pseudo-octasaccharide V-1532 (52)). Some support for this possibility comes from the observation that the M. tuberculosis enzyme is not inhibited by cyclodextrins and yet its acceptor specificity is quite similar to that of S. coelicolor GlgE. While most of the cyclodextrin binding patch is well conserved (Fig. 7c and Fig. S8), its end distal to the donor site is likely to be different in the mycobacterial enzyme due to the presence of a variable loop (Fig. 7c). This loop bears the Gly84 backbone that interacts with cyclodextrins in the structures (Fig. 7b) but includes an insertion of nine amino acid residues in the M. tuberculosis enzyme (Fig. S8). The 23-fold lower $K_{\text{m}}^{\text{app}}$ for maltotetraose and an order of magnitude higher $k_{\text{cat}}^{\text{app}}$ with the S. coelicolor enzyme likely reflect the effect of the variable loop. Nevertheless, despite such a significant amino acid insertion, it remains likely that the conserved elements of the patch form part of the acceptor binding site and help define acceptor length specificity.

The GlgE pathway ultimately generates a branched α-glucan, so it is conceivable that GlgE...
also extends Y-shaped branched glucans. The arms of such acceptors could occupy conspicuous diagonal and/or linear clefts (Fig. 6b and Fig. 7c), the latter being more highly conserved and partly defined by the novel domain S. Further work is clearly required to fully identify the acceptor site.

Relevance to GlgEs from Other Organisms—

The glgE gene is widespread among bacteria (6). Most of these genes are similar in length to that of S. coelicolor, making it possible to generate homology models of GlgE proteins based on our structure. Interestingly, there are some examples that are about 60% longer, such as BPSL2074 of the human pathogen Burkholderia pseudomallei K96243. Inspection of the protein sequence encoded by this gene showed that it has an N-terminal extension that results from a partial duplication. This extension is unlikely to exhibit GlgE activity however, because it lacks most of the catalytic machinery, most of the residues defining the maltose binding pocket, and domains N and S. Whether it serves some other function remains to be seen.

The GlgEs from S. coelicolor and M. tuberculosis are very similar in length and share very similar properties, allowing one to be used as a structural model for the other. Indeed, the very high degree of conservation between the maltose binding site residues of these enzymes (Fig. S8) and their similar $K_m$ values for maltose 1-phosphate (Table 1) illustrate this. This allows the structure of the S. coelicolor enzyme to be used to guide inhibitor design for the M. tuberculosis enzyme, which has been genetically validated as a potential novel drug target (3). Although the maltose site is largely hydrophilic, it includes two aromatic residues that sandwich maltose and provide potential hydrophobic surfaces to enhance the binding of inhibitors to the GlgE of M. tuberculosis and of other animal and plant pathogens (6). Importantly, the distinct configuration of the donor site of this GH13_3 enzyme provides the opportunity to develop inhibitors that do not target the many other GH13 sub-family enzymes present in mammals and plants.

REFERENCES


FOOTNOTES

*This work was supported by the United Kingdom Biotechnology and Biological Sciences Research Council through a grant (BB/I012850/1) and the MET Institute Strategic Programme Grant to the John Innes Centre.

The on-line version of this article (available at http://www.jbc.org) contains supplemental discussion of equilibria within the GlgE pathway, synthetic experimental procedures, Figs. S1-S12 and crystallographic Tables S1-S2.

The atomic coordinates and structure factors (codes 3zss, apo-GlgE; 3zst, αCD-GlgE; 3zt5, mal-GlgE; 3zt6, αCD-mal-GlgE; and 3zt7, βCD-GlgE) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rscb.org/).

The abbreviations used are: PDB, Protein Data Bank; DP, degrees of polymerization; mal, maltose; αCD, α-cyclodextrin; βCD, β-cyclodextrin.

ACKNOWLEDGEMENTS—We thank beamline scientists at the Diamond Light Source for assistance with X-ray data collection, Lionel Hill for recording mass spectra, Tom Clarke for running the analytical ultracentrifuge, Ainhoa Fernandez for preparing α-maltosyl fluoride, and Rainer Kalscheuer, Stephen G. Withers and Bernd Nidetzky for helpful discussions.

FIGURE LEGENDS

FIGURE 1. α-Glucan pathways of actinomycetes. The classical GlgA cytosolic glycogen pathway and the newly identified GlgE pathway (highlighted in red (3)) are common to both S. coelicolor and M. tuberculosis. The Rv3032 pathway for methylglucose lipopolysaccharide biosynthesis is present in M. tuberculosis. Which pathway is responsible for capsular glucan biosynthesis in M. tuberculosis is not yet clear and there may be redundancy between the pathways.

FIGURE 2. Synthesis of β-maltose 1-phosphate (1). Compound numbers in the text with suffixes a and b refer to α and β anomers, respectively.

FIGURE 3. Acceptor specificity of GlgE. Panels A and B show acceptor specificity of S. coelicolor isoforms I and II, respectively. Enzyme activity with maltooligosaccharide acceptor substrates with different degrees of polymerization (DP) was determined by monitoring Pi release in triplicate. The same trends were observed with the MALDI TOF MS assay (e.g. Fig. S4). The bars indicate means ± SEM. For comparison, data from M. tuberculosis GlgE (3) are shown in C.
FIGURE 4. Maltotetraitol is an acceptor for GlgE isoform I. The structure of maltotetraitol is shown in A. Panel B shows MALDI TOF MS of maltotetraitol where the mass of the starting material (m/z 691; [M+Na]+; highlighted with a star) is amongst the peaks from the matrix and other reaction mixture components. Panel C shows the spectrum after incubation with enzyme and α-maltose 1-phosphate revealing a series of peaks (highlighted with arrows) with the successive addition of m/z 324 corresponding to maltosyl units. The DP of each peak (including the glucitol chain) is indicated. Panels D and E respectively show 1H NMR spectra of maltotetraitol before and after incubation with enzyme and α-maltose 1-phosphate. Peak assignments are indicated.

FIGURE 5. Ability of GlgE isoform I to use α-maltosyl fluoride as a donor. Panel A shows MALDI TOF MS of a reaction mixture containing maltotetraose (5 mM) after 10 min exposure to enzyme and α-maltosyl fluoride (5 mM). Panel B shows a control with α-maltose 1-phosphate (5 mM). The successive addition to the acceptor of m/z 324 was observed, which corresponds to maltosyl units. The degree of polymerization (DP) associated with each peak is highlighted. α-Maltosyl fluoride was an efficient donor yielding longer polymers than α-maltose 1-phosphate under these conditions.

FIGURE 6. Structure of S. coelicolor GlgE isoform I. Panel A shows the GlgE homodimer in ribbon representation and in wall-eyed stereo highlighting domain N (residues 1-108 and 192-205), domain S (109-191), domain A (200-253, 300-322, 536-512 and 553-573), insert 1 (254-299), domain B (323-367), insert 2 (513-552), and domain C (574-675). Panel B shows a space-filling representation. Various features are highlighted including the gap which the domain B lid could potentially occupy in order to allow access to the donor site.

FIGURE 7. Maltose and α-cyclodextrin bound to GlgE. Panel A shows α-maltose in mal-GlgE and panel B α-cyclodextrin in αCD-GlgE. Difference electron density 'omit' maps were generated for bound ligands using phases calculated from the final models minus the ligand coordinates after simulated annealing refinement. This was performed from a starting temperature of 5000 K after applying random shifts to the model ('shake' term set to 0.3) using PHENIX (53). The resultant maps were non-crystallographic symmetry averaged to improve connectivity. The corresponding stereo images are shown in Fig. S9. Most amino acids interacting with the ligands are highlighted but some are omitted here for clarity (all are shown in Fig. S10). Panel C shows the relative orientations of maltose and α-cyclodextrin in the αCD-mal-GlgE structure (comparable with the lower part of Fig. 6b). GlgE is shown in space-filling mode and coloured by sequence conservation between the S. coelicolor and M. tuberculosis enzymes (conserved in blue through to similar in green and dissimilar in red). The donor pocket is highly conserved, the linear cleft is well conserved and some of the cyclodextrin binding patch is well conserved except for a variable loop as indicated.

FIGURE 8. Proposed mechanism of GlgE. The extension of a maltooligosaccharide acceptor by α-maltose 1-phosphate is shown. The reversibility of the second step allows disproportionation reactions to occur.
TABLE 1
Michaelis-Menten kinetic analysis of *S. coelicolor* GlgE isoforms I and II
Enzyme activity was monitored by detecting Pi release in triplicate and expressed as the mean and standard error (see Fig. S5).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>$K_{m\text{app}}$</th>
<th>$k_{cat\text{app}}$</th>
<th>$k_{cat\text{app}}/K_{m\text{app}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mM</td>
<td>s$^{-1}$</td>
<td>M$^{-1}$ s$^{-1}$</td>
</tr>
<tr>
<td><em>S. coelicolor</em> isoform I</td>
<td>maltose 1-phosphate$^a$</td>
<td>0.30 ± 0.06</td>
<td>12.3 ± 0.5</td>
<td>41,000 ± 8,000</td>
</tr>
<tr>
<td></td>
<td>maltotriose$^b$</td>
<td>1.5 ± 0.3</td>
<td>53 ± 2</td>
<td>36,000 ± 7,000</td>
</tr>
<tr>
<td><em>S. coelicolor</em> isoform II</td>
<td>maltose 1-phosphate$^a$</td>
<td>1.2 ± 0.2</td>
<td>10.0 ± 0.6</td>
<td>8,000 ± 1,700</td>
</tr>
<tr>
<td></td>
<td>maltotriose$^b$</td>
<td>2.3 ± 0.4</td>
<td>23.5 ± 1.1</td>
<td>10,000 ± 2,000</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> $^c$</td>
<td>maltose 1-phosphate$^a$</td>
<td>0.25 ± 0.05</td>
<td>1.26 ± 0.07</td>
<td>5,000 ± 1,000</td>
</tr>
<tr>
<td></td>
<td>maltotriose$^b$</td>
<td>35 ± 8</td>
<td>15.4 ± 1.1</td>
<td>440 ± 100</td>
</tr>
</tbody>
</table>

$^a$In the presence of 1 mM maltotriose

$^b$In the presence of 5 mM M1P

$^c$From (3)
<table>
<thead>
<tr>
<th>Data set</th>
<th>apo-GlgE</th>
<th>αCD-GlgE</th>
<th>mal-GlgE</th>
<th>αCD-mal-GlgE</th>
<th>βCD-mal-GlgE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P2₁2₁2₁</td>
<td>P4₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell parameters (Å)</td>
<td>a = 113.1, b = 113.0, c = 314.2</td>
<td>a = b = 113.2, c = 314.5</td>
<td>a = 113.8, b = 113.6, c = 315.7</td>
<td>a = 113.9, b = 114.2, c = 315.6</td>
<td>a = 113.3, b = 113.4, c = 315.0</td>
</tr>
<tr>
<td>Beamline</td>
<td>103</td>
<td>102</td>
<td>104</td>
<td>104</td>
<td>102</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9709</td>
<td>0.9795</td>
<td>0.9763</td>
<td>0.9763</td>
<td>0.9795</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>(53.15 - 1.80)</td>
<td>(71.35 - 2.30)</td>
<td>(56.89 - 2.10)</td>
<td>(63.99 - 2.20)</td>
<td>(71.42 - 2.50)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>351,739</td>
<td>85,772</td>
<td>236,917</td>
<td>206,349</td>
<td>140,547</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.0 (72.9)</td>
<td>93.7 (66.0)</td>
<td>99.5 (96.7)</td>
<td>98.9 (93.2)</td>
<td>99.8 (99.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.6 (5.5)</td>
<td>14.3 (13.2)</td>
<td>7.3 (4.7)</td>
<td>5.4 (4.1)</td>
<td>4.9 (5.0)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.110 (0.762)</td>
<td>0.088 (0.327)</td>
<td>0.121 (0.426)</td>
<td>0.151 (0.602)</td>
<td>0.143 (0.688)</td>
</tr>
<tr>
<td>Rmeas</td>
<td>0.118 (0.840)</td>
<td>0.092 (0.340)</td>
<td>0.130 (0.482)</td>
<td>0.167 (0.692)</td>
<td>0.160 (0.767)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>13.4 (2.1)</td>
<td>23.6 (7.9)</td>
<td>12.0 (3.2)</td>
<td>7.7 (2.1)</td>
<td>10.0 (2.3)</td>
</tr>
<tr>
<td>Wilson B value (Å²)</td>
<td>21.6</td>
<td>32.2</td>
<td>26.1</td>
<td>33.8</td>
<td>42.4</td>
</tr>
<tr>
<td>Twin fraction</td>
<td>0.39</td>
<td>0.04</td>
<td>0.36</td>
<td>0.27</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**Refinement**

<table>
<thead>
<tr>
<th>Reflections: working/free</th>
<th>333,972/17,659</th>
<th>81,357/4,301</th>
<th>224,955/11,847</th>
<th>195,790/10,447</th>
<th>133,513/6,958</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rwork</td>
<td>0.231</td>
<td>0.173</td>
<td>0.204</td>
<td>0.208</td>
<td>0.191</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.249</td>
<td>0.201</td>
<td>0.228</td>
<td>0.236</td>
<td>0.221</td>
</tr>
<tr>
<td>Ramachandran favoured/allowed (%)</td>
<td>98.6/100.0</td>
<td>98.7/100.0</td>
<td>98.8/100.0</td>
<td>98.7/100.0</td>
<td>98.5/100.0</td>
</tr>
<tr>
<td>Ramachandran outliers</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>rmsd bond distances (Å)</td>
<td>0.013</td>
<td>0.016</td>
<td>0.015</td>
<td>0.015</td>
<td>0.014</td>
</tr>
<tr>
<td>rmsd bond angles (°)</td>
<td>1.34</td>
<td>1.53</td>
<td>1.44</td>
<td>1.46</td>
<td>1.45</td>
</tr>
<tr>
<td>Twin fraction</td>
<td>0.48</td>
<td>n/a</td>
<td>0.49</td>
<td>0.45</td>
<td>0.49</td>
</tr>
</tbody>
</table>

**Contents of model**

| Protein residues | 4 × 649 | 2 × 649 | 4 × 649 | 4 × 649 | 4 × 649 |
| Glucans | 0 | 2 × αCD | 4 × mal | 4 × αCD; 4 × mal | 4 × βCD; 4 × mal |
| Ethylene glycol | 0 | 1 | 0 | 0 | 0 |
| Water molecules | 540 | 731 | 486 | 459 | 369 |

**Average atomic displacement parameters (Å²)**

| Main chain atoms | 25.0 | 47.1 | 28.6 | 32.7 | 31.8 |
| Side chain atoms | 25.6 | 49.2 | 29.1 | 33.9 | 32.5 |
| Glucans | - | 78.5 | 25.1 | αCD: 69.3; mal: 28.3 | βCD: 41.9; mal: 37.5 |
| Ethylene glycol | - | 49.8 | - | - | - |
| Water molecules | 20.2 | 44.9 | 21.7 | 24.4 | 21.4 |
| Overall | 25.2 | 48.3 | 28.7 | 33.5 | 32.1 |

**PDB accession code**

3zss, 3zst, 3zt5, 3zt6, 3zt7

---

**Notes:**

- Space group used for refinement.
- I02, I03, I04 = beamlines at the Diamond Light Source (Oxfordshire, UK).
- The figures in brackets indicate the values for outer resolution shell.
- \[ R_{\text{merge}} = \sum_{i} \left( I_{i}(hkl) - \langle I(hkl) \rangle \right)/ \sum_{i} I_{i}(hkl) \text{, where } I_{i}(hkl) \text{ is the } i \text{th observation of reflection } hkl \text{ and } \langle I(hkl) \rangle \text{ is the weighted average intensity for all observations } i \text{ of reflection } hkl. \]
- \[ R_{\text{meas}} = \sum_{i} \left[ N(N-1) \right]^{-1} \sum_{i} \left( I_{i}(hkl) - \langle I(hkl) \rangle \right)/ \sum_{i} I_{i}(hkl) \text{, where } N \text{ is the number of observations of reflection } hkl. \]
- As calculated by TRUNCATE (24).
- As calculated using MOLPROBITY (32).
- Refined values from REFMAC5 (29).

---

1. Space group used for refinement.
2. I02, I03, I04 = beamlines at the Diamond Light Source (Oxfordshire, UK).
3. The figures in brackets indicate the values for outer resolution shell.
4. \[ R_{\text{merge}} = \sum_{i} \left( F_{i} - F_{\text{calc}} \right)/ \sum \left| F_{i} \right| x 100, \text{ where } F_{i} \text{ and } F_{\text{calc}} \text{ are the observed and calculated structure factor amplitudes, respectively.} \]
5. As calculated using MOLPROBITY (32).
6. Refined values from REFMAC5 (29).
Figure 2

1) LDA, THF
2) [(BnO)₂PO]₂O
3) separation of anomers

1) Pd/C, H₂, MeOH

2) Et₃N

3) separation of anomers

Dowex Na⁺

5: R = Et₃NH⁺
1: R = Na⁺
Figure 3
**Figure 4**

A = Maltotetraotol

- $\alpha$-1,4 from residual M1P from product
- $\alpha$-1,4 from residual M1P from product
- (α-1,4)-Glc from maltotetraotol
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product
- (α-1,4)-Glc from product

- ε = Maltotetraotol

**Relative intensity (%)**

- 0
- 50
- 100

**m/z**

- 0
- 100
- 200
- 1,000
- 2,000

- ppm

- 5.4
- 5.0
- 4.6

**From:**

- Product
- Residual M1P
- Maltotetraotol
- α-reducing end
Figure 5
Figure 7
Structure of a *Streptomyces* maltosyltransferase GlgE: a homologue of a genetically validated anti-tuberculosis target

Karl Syson, Clare E. M. Stevenson, Martin Rejzek, Shirley A. Fairhurst, Alap Nair, Celia J. Bruton, Robert A. Field, Keith F. Chater, David M. Lawson and Stephen Bornemann

*J. Biol. Chem. published online September 13, 2011*

Access the most updated version of this article at doi: 10.1074/jbc.M111.279315

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2011/09/13/M111.279315.DC1