Crystal structure of the bifunctional *N*-acetylglucosamine 1-phosphate uridylyltransferase from *Escherichia coli*: a paradigm for the related pyrophosphorylase superfamily

Kieron Brown, Frédérique Pompeo¹,
Suzanne Dixon, Dominique
Mengin-Lecreulx¹, Christian Cambillau and
Yves Bourne²

AFMB-CNRS, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20
and ¹CNRS, Biochimie Structurale et Cellulaire, Université Paris-Sud,
91405 Orsay Cedex, France

²Corresponding author
e-mail: yves@afmb.cnrs-mrs.fr

N-acetylglucosamine 1-phosphate uridylyltransferase (GlmU) is a cytoplasmic bifunctional enzyme involved in the biosynthesis of the nucleotide-activated UDP-GlcNAc, which is an essential precursor for the biosynthetic pathways of peptidoglycan and other components in bacteria. The crystal structure of a truncated form of GlmU has been solved at 2.25 Å resolution using the multiwavelength anomalous dispersion technique and its function tested with mutagenesis studies. The molecule is composed of two distinct domains connected by a long α-helical arm: (i) an N-terminal domain which resembles the dinucleotide-binding Rossmann fold; and (ii) a C-terminal domain which adopts a left-handed parallel β-helix structure (LβH) as found in homologous bacterial acetyltransferases. Three GlmU molecules assemble into a trimeric arrangement with tightly packed parallel LβH domains, the long α-helical linkers being seated on top of the arrangement and the N-terminal domains projected away from the 3-fold axis. In addition, the 2.3 Å resolution structure of the GlmU–UDP-GlcNAc complex reveals the structural bases required for the uridylyltransferase activity. These structures exemplify a three-dimensional template for the development of new antibacterial agents and for studying other members of the large family of XDP-sugar bacterial pyrophosphorylases.

Keywords: acetyltransferase/bifunctional/crystallography/
drug design/pyrophosphorylase

Introduction

UDP-GlcNAc, the cytoplasmic nucleotide-activated form of *N*-acetylglucosamine (GlcNAc), is an essential precursor for the peptidoglycan and lipopolysaccharide biosynthesis pathways in Gram-positive and Gram-negative bacteria, respectively (Raetz, 1996). In Gram-negative bacteria, this precursor is found in the enterobacterial common antigen and some O antigen (Rick and Silver, 1996; van Heijenoort, 1996), whereas in Gram-positive bacteria it is found in teichoic acids (Fisher, 1990). The assimilation of GlcNAc, *N*-acetylmannosamine

and *N*-acetylneuraminic acid in bacteria all converge at the step of *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P) for the biosynthesis of UDP-GlcNAc (Plumbridge and Vimr, 1999), suggesting that enzymes involved in its biosynthesis represent attractive targets for the development of new antibacterial agents. Conditional lethal mutants of *Escherichia coli* altered in the biosynthesis of UDP-GlcNAc were characterized by a cell lysis phenotype under restrictive growth conditions (White, 1968; Sarvas, 1971; Wu and Wu, 1971). The four-step formation from fructose-6-P to UDP-GlcNAc is fully elucidated in *E. coli* and involves three enzymes: (i) a GlcN-6-P synthase (GlmS) (Dutka-Malen *et al.*, 1988); (ii) a phosphoglucosamine mutase (GlmM) (Mengin-Lecreulx and van Heijenoort, 1996); and (iii) a GlcN-1-P acetyltransferase and GlcNAc-1-P uridylyltransferase (GlmU) (Mengin-Lecreulx and van Heijenoort, 1994) which catalyses the two-step mechanism in the presence of MgCl₂, according to the following reactions:

Escherichia coli GlmU, a cytoplasmic 49 kDa protein, represents the first reported bifunctional enzyme in the cytoplasmic steps of the peptidoglycan and lipopolysaccharide synthesis pathways. The recombinant protein has been overexpressed and purified to homogeneity and shown to be dimeric/trimeric in solution (Mengin-Lecreulx and van Heijenoort, 1993). Homologue sequences were reported for *Haemophilus influenzae* (Fleischmann *et al.*, 1995), *Neisseria gonorrhoeae* (Ullrich and van Putten, 1995) and *Bacillus subtilis* (Hove-Jensen, 1992) with sequence identities of 69, 54 and 42%, respectively (Figure 1). The enzyme activity is maximal in the presence of magnesium and phosphate ions; in addition, the acetyltransferase activity is inactivated by thiol-specific reagents (Mengin-Lecreulx and van Heijenoort, 1994; Pompeo *et al.*, 1998), indicating that cysteine residues might be involved in the acetyl transfer reaction. Mutagenesis experiments, aimed at investigating the role of the four cysteine residues of the *E. coli* GlmU enzyme, have demonstrated the important role of Cys307 and Cys324 for acetyltransferase activity (Pompeo *et al.*, 1998), but have not revealed the exact location of the active site or the catalytic residues involved in the uridylyltransferase activity. GlmU has also been used to synthesize azido-substituted nucleotide-sugar analogues for the identification and characterization of glycosyltransferases (Sunthakar *et al.*, 1998) and to produce large amounts of *N*-acetyl-labelled UDP-GlcNAc for industrial purposes (Leiting *et al.*, 1998).

Biochemical experiments (Mengin-Lecreulx and van Heijenoort, 1994; Gehring *et al.*, 1996) have shown that GlmU is organized into two domains: (i) a pyro-

Fig. 1. GlmU sequence conservation, secondary structure, fold and function. The sequence of the enzyme of *E.coli* is aligned with that of three bacterial species, *H.influenzae* (H.i.), *N.gonorrhoeae* (N.g.) and *B.subtilis* (B.s.). Invariant and conserved residues in all sequences are highlighted in white with a red background and white with a yellow background, respectively. Secondary structure elements of *E.coli* GlmU are indicated beneath the sequences with the consensus sequence motif (red bar). Residue functions implied by the GlmU structure are identified by triangles that are black for mutated residues, green for residues in the acetyltransferase domain, and red for those in the pyrophosphorylase domain. Residues buried at the trimer interface are displayed as filled black circles, and residues involved in UDP-GlcNAc binding as filled orange and blue circles for those in contact with the nucleotide and the sugar moiety, respectively. The first and last residue well defined in the electron density map, Asn3 and Ile326, are indicated as filled magenta triangles. Deletions are indicated by dashes.

phosphorylase N-terminal domain that shares homology with other pyrophosphorylase (PPase) enzymes over residues Met1–Ala120 with the strict conservation of the motif G-X-G-T-(RS)-(X)₄-P-K; and (ii) an acetyltransferase domain that possesses 23 times the hexapeptide repeat (LIV)-(GAED)-X₂-(STAV)-X typically found in other bacterial acetyltransferases (Vaara, 1992).

In eukaryotes, the equivalent UDP-GlcNAc biosynthesis pathway is accomplished by four distinct enzymes, two of which are involved in the reaction catalysed by GlmU: (i) a UDP-*N*-acetylglucosamine PPase (Uap1p) that preserves the consensus sequence motif found in all PPase amino acid sequences (Mio *et al.*, 1998); and (ii) a glucosamine-6-phosphate acetyltransferase (GNA1) (Mio *et al.*, 1999), which possesses little structural homology with other known bacterial acetyltransferases. The UDP-GlcNAc nucleotide-sugar is a substrate of chitin synthase whose product is essential for fungal cell wall, and the

GlcNAc moiety is found in N-linked glycosylation and the glycerophosphatidylinositol (GPI) anchor of cellular proteins (Cabib *et al.*, 1982; Herscovics and Orlean, 1993). Human UDP-GlcNAc PPase, AGX1, which causes human male infertility (Diekman and Goldberg, 1994), is able to synthesize both UDP-GlcNAc and UDP-GalNAc by alternative splicing of the *AX1* gene resulting in a 17 amino acid insertion in the C-terminal region of the protein (Wang-Gillam *et al.*, 1998). Other nucleotide sugars, such as UDP-Glc, play a crucial role in the quality control of newly synthesized glycoproteins (Hammond and Helenius, 1995). Deficiencies in the nucleotide-sugar UDP-Glc occur in insulin-dependent tissues of diabetic organisms. A single mutation, Gly115 to Asp, dramatically impairs the enzymatic activity of Chinese hamster UDP-Glc PPase (Flores-Diaz *et al.*, 1997); yet the consequences of this deficiency remain unclear.

The three-dimensional structures of four bacterial

Fig. 2. (A) Experimental electron density map for a GlmU L β H domain region at 2.8 Å resolution based on the solvent-flattened, NCS-averaged MAD phases (contoured at 1.2 σ); the current model is superimposed and shows residues Gly264–Glu281 located in coil C2 (see B). (B) Sequence alignment of the L β H domain of GlmU identifying equivalent residues in each coil. The nomenclature PB1, PB2, PB3 and T1, T2, T3 denotes parallel β -strands and turn residues, respectively (Yoder *et al.*, 1993). The conserved, hydrophobic residues at position *i* are boxed. Residues in left-handed conformation at position *i* + 3 are displayed in bold. This alignment is derived from the present structure for residues Val252–Pro328, and from a model for residues Pro328–Gly424. L1 denotes the loop region Arg333–Lys352 and L2 a second loop region, Asn375–Lys394, in which the hexapeptide repeat sequence is not found.

acetyltransferases are known, but none contain a PPase domain. Moreover, the three-dimensional structures of the sugar-phosphate transferring enzymes, galactose 1-phosphate uridylyltransferase (Wedekind *et al.*, 1995) and kanamycin nucleotidyltransferase (Pedersen *et al.*, 1995), are known but these enzymes do not activate sugars and differ structurally from GlmU. To understand the catalytic mechanism of GlmU and provide a scaffold for the development of new bacterial antibiotics, we have determined the crystal structures of a truncated form of GlmU (GlmU-Tr) at 2.25 Å resolution and of GlmU-Tr bound to UDP-GlcNAc at 2.3 Å resolution. Coupling of the structural data with mutagenesis studies reveals the structural and functional duality of the enzyme and the structural determinants responsible for the uridylyltransferase activity.

Results and discussion

Structure determination

Conditions were found that yielded crystals of the recombinant entire 147 kDa trimeric GlmU, but these crystals did not diffract to a resolution sufficient for determination of a structure. Because of the occurrence of a stable spontaneous proteolytic fragment of GlmU-Tr, consisting of residues Met1–Arg331 (Gehring *et al.*, 1996), we next reproduced this truncated form of GlmU, which yielded an apparent molecular mass of 36 kDa on SDS–polyacrylamide gel and from which large, well-ordered crystals diffracting up to 2.0 Å resolution were obtained (cf. Materials and methods). The crystals which belonged to the rhombohedral space group R32 ($a = b = 142.7$ Å and $c = 248.1$ Å) with two GlmU-Tr molecules per asymmetric unit were used to solve the structure by the multiple wavelength anomalous dispersion (MAD) method

from a mercury derivative. A representative portion of the experimental electron density map is shown (Figure 2A). The GlmU-Tr and GlmU–UDP-GlcNAc complex structures have a crystallographic *R*-factor of 23.4 and 22.3% ($R_{\text{free}} = 27.4$ and 26.3%) at 2.25 and 2.3 Å resolution, respectively, and have good stereochemistry (Tables I and II). In each of the two structures, the two independent molecules present in the asymmetric unit are very similar within the two domains, but show a different overall bend as a consequence of different packing forces in the crystal. The first two N-terminal and last five C-terminal residues are ill-defined in the two molecules and were not incorporated into the models. In addition, side chains of the surface loop Gly140–His155, located in the N-terminal domain of one molecule, are poorly defined in the electron density maps.

Overall structure of the molecule

GlmU-Tr consists of two separate domains, formed from contiguous segments in the amino acid sequence and linked by a long α -helical arm, and has overall dimensions of 45×58×66 Å (Figure 3). The N-terminal domain comprises residues Asn3–Asn227 and consists of a central, seven-stranded mixed β -sheet (β 1– β 7), of which six are parallel, surrounded by six helices (α 1– α 6), a fold reminiscent of the dinucleotide-binding Rossmann fold (Rossmann *et al.*, 1975). At one of its ends, the central β -sheet is topped by a two-stranded β -sheet (β 5a and 5b) that participates in a 30 residue connection between strands β 5 and β 6. Comparison of this N-terminal domain with the DALI database of protein structures (Holm and Sander, 1995) did not reveal any homologous structures. The C-terminal domain comprises residues Gly251–Arg331 and adopts the left-handed parallel β -helix (L β H)

Fig. 3. Ribbon diagram of the overall view of GlmU-Tr showing the N-terminal domain (Asn3–Asn227) with the central β -sheet displayed in magenta, the additional two-stranded β -sheet in green, and α -helices in blue; the α -helical arm (Asn228–Ala250) is displayed in green and the C-terminal L β H domain (Gly251–Ala330) in magenta.

Table I. Data collection statistics

	f'' max	f'' min	Remote II	Remote I	Native	UDP-GlcNAc
Beamline	DESY BW7A			ESRF ID14-EH3		ESRF ID14-EH4
R_{sym} (%) ^a	6.4 (34)	6.6 (34)	6.8 (25.7)	6.4 (45.6)	3.9 (31)	7.3 (41)
R_{anom} (%) ^b	6.1 (27.3)	5.3 (24.4)	5.4 (20.3)	5.6 (25.6)	–	–
Resolution (Å)	2.8	2.8	2.8	2.8	2.25	2.2
$I/\sigma(I)$	9.8 (2.2)	4.5 (2.2)	6.6 (2.5)	4.9 (2.2)	14 (2.3)	5.1 (1.6)
Redundancy	3.6	3.5	3.6	3.5	2.7	3.6
Completeness (%)	99.7 (99.7)	99.5 (99.8)	99.6 (99.4)	99.6 (100)	97 (76.8)	97.3 (98.1)
Anom completeness (%)	97.6 (100)	95 (99.8)	82.9 (75.3)	96.2 (100)	–	–

Values of R_{sym} and R_{anom} in parentheses are for the outer resolution shell.

^a $R_{\text{sym}} = \sum |I| - \langle I \rangle / \sum \langle I \rangle$, where I is intensity and $\langle I \rangle$ is the average I for all equivalent reflections.

^b $R_{\text{anom}} = \sum |\langle I^+ \rangle - \langle I^- \rangle| / \sum (\langle I^+ \rangle + \langle I^- \rangle)$, where I^+ and I^- are the intensities of Bijvoet positive and negative reflections, respectively.

Table II. Refinement statistics

	Apo GlmU-Tr	UDP-GlcNAc
Resolution (Å)	20–2.25	30–2.3
R -factor ^a – R_{free} ^b (%)	23.4–27.4	22.3–26.2
No. of reflections (no σ cutoff)	44 974	40 231
R.m.s. deviations		
Bond length (Å)	0.013	0.013
Bond angles (°)	1.67	1.67
Dihedral angles (°)	24.8	24.7
Improper angles (°)	0.9	0.96
Mean B -factors (Å ²)		
Main chain	53	57
Side chain	56	59
Solvent	55	58
Cofactor ^c	69/82	53/56/79
R.m.s. deviations on B -factors (Å ²)		
Main chain	1.6	1.4
Side chain	2.3	2.3

^a R -factor = $\sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$.

^b R_{free} is the same as the R -factor for 3% of the data omitted from the refinement.

^cCofactors are respectively MES/sulfate and UDP-GlcNAc/sulfate/ethylene glycol.

domain fold found in the structures of other bacterial acetyltransferases, which include *E. coli* UDP-*N*-acetylglucosamine acyltransferase (LpxA) (Raetz and Roderick, 1995), *Methanosarcina thermophila* carbonic anhydrase (Cam) (Kisker *et al.*, 1996), tetrahydrodipicolinate *N*-succinyltransferase (DapD) (Beaman *et al.*, 1997) and *Pseudomonas aeruginosa* hexapeptide xenobiotic acetyltransferase (PaXAT) (Beaman *et al.*, 1998). The L β H domain resembles a prism, each turn or coil containing three short β -strands (PB1–PB3) of nearly equal length and no insertion loop at the corners (Figure 2). The first coil is unusual in that a histidine residue (His268) is present at the *i* position of the hexapeptide repeat sequence, a position which ordinarily is occupied by aliphatic residues (Leu, Ile and Val), and the T2 turn of coil C1 contains two additional residues compared with homologous structures. In addition, an unprecedented disulfide bridge is formed between Cys307 and Cys324 both located at the *i* + 4 position of β -strand PB2 in two adjacent coils (Figure 2). The N- and C-terminal domains are connected by a 21 residue α -helical arm which is 33 Å long, lies perpendicular to the L β H domain axis and projects the N-terminal domain away from the axis of the

Table III. Enzymatic activities of wild-type and mutant GlmU^a

Protein	K_m (mM)		k_{cat} (/s)		Complementation UGS 83
	GlcNAc-1-P	UTP	Acetyltransferase	Uridyltransferase	
Wild-type GlmU	0.07	0.1	1500	350	++
His ₆ -GlmU	0.07	0.1	1350	330	++
His ₆ -GlmU-Tr	0.08	0.1	0	136	—
His ₆ -GlmU-G14A	0.1	0.08	811	40	++
His ₆ -GlmU-R18A	0.06	0.26	1045	0.06	+
His ₆ -GlmU-K25A	0.075	0.17	1167	46	+

^aSee Materials and methods for experimental details.

LβH domain. As a result, the only contacts between the N- and C-terminal domains involve van der Waals interactions between the surface loop Ala31–Gly32 in the N-terminal domain and the Arg263 side chain in the C-terminal domain. In contrast, residues within the long α-helical arm establish numerous interactions with residues in the two domains, owing to the protruding loop T2 that emanates from the LβH domain coil C1 and interacts with the helical linker (Figures 2 and 3).

Biochemical experiments, showing that the present truncated form of GlmU displays normal uridyltransferase activity but has lost acetyltransferase activity (Table III), suggest that GlmU has two separate active sites which can function independently of each other. Moreover, complementation with GlmU-Tr using an *E.coli* strain containing a chromosomal disruption of the *glmU* gene reveals that the bacteria are not viable (Table III); hence the two distinct GlmU activities are required for bacterial viability.

The pyrophosphorylase domain and UDP-GlcNAc-binding site

The N-terminal domain of GlmU-Tr consists of a mixed seven-stranded twisted β-sheet with strand order 7564123, surrounded by three α-helices on each side and an additional two-stranded β-sheet on one edge (Figure 3). The uridyltransferase active site, identified by the presence of soaked UDP-GlcNAc, is a large open pocket, 21 Å long and 13 Å deep, which delineates two lobes of approximately equal length and separated by only 7 Å at the pocket entrance. The first lobe consists of residues Asn3–Val111 and surface loop His216–Asn227, and encompasses residues interacting with the nucleotide. The second lobe contains the remaining residues of the N-terminal domain and encompasses residues interacting with the sugar moiety (Figure 4). The consensus sequence motif G-X-G-T(R/S)-(X)₄-P-K, which is located in surface loop β1–α1 and is a signature of PPases, exposes Leu11, Gly14, Gly16, Arg18 and Lys25, along with the region Gln76–Thr82, towards the pocket, while the region Gly138–Asp157 and the sequence-conserved motif Gln193–Tyr197 are located in the second lobe. Two distinct regions, Tyr103–Asp105 and Val223–Asn227, form the floor of the pocket. The importance of the pocket for substrate binding is also apparent from the complete exposure of the sequence-conserved residues in or at its rim in addition to its electrostatic properties (Figure 4B).

GlmU binds UTP with high specificity, utilizing neither ATP, CTP nor TTP as substrates, although CTP is a weak

inhibitor (data not shown). The exocyclic N3 and O4 ring atoms of uracil are anchored against the sequence-conserved Gln76 side chain, while the exocyclic O4 position involves an additional hydrogen bond with the nitrogen backbone atom of Gly81 (Figures 4C and 5A). In the second lobe, all the hydroxyl groups, O2, O3, O4 and O5, of the GlcNAc moiety are directly hydrogen-bonded to protein residues, with the exception of the O6 atom which is bound to the protein via a water molecule. The N-acetyl arm, which distinguishes GlcNAc from Glc, is hydrogen-bonded to both the Glu154 and Thr82 side chains, with the methyl group in van der Waals contacts with the Tyr197 side chain. Unlike the nucleotide and sugar moieties, which are bound to the protein, the phosphate groups are totally solvent accessible (Figures 4 and 5).

Site-directed mutagenesis shows that substitution of an Ala residue for Arg18 within the consensus sequence motif dramatically impairs the uridyltransferase activity, whereas substitution of Ala residues for Gly14 and Lys25 induces only an 8-fold decrease in this activity (Table III). In contrast, none of these three mutations alter the acetyltransferase activity, a result consistent with the two distinct activities of GlmU belonging to separate domains (Mengin-Lecreulx and van Heijenoort, 1994; Gehring *et al.*, 1996). Arg18 is located at the tip of loop β1–α1 near the pocket entrance and is within hydrogen bonding distance of a sulfate molecule, present in both the free and complex structures and separated by only 9 Å from the P_α of UDP-GlcNAc; this suggests that this sulfate ion mimicks the P_γ position in UTP. Most importantly, Arg18 is located on top of the positively charged N-cap of the long α-helical arm, which provides a remarkable charge complementarity for phosphate binding. In contrast, Lys25 is located deep in the bottom of the pocket and is salt-bridged to Asp105 and hydrogen-bonded to Asn227, suggesting that Lys25 could stabilize the correct orientation of the Asp105 and Asn227 side chains for UTP recognition, whilst it may not be directly involved in the catalytic reaction. Based on our structure, the Gly14 to Ala mutation would create a steric conflict with the ribose moiety, associated with a higher K_m value for UTP, a hypothesis which, however, is not supported by biochemical data (Table III); further study will therefore be needed to confirm the role of this residue unambiguously.

Superimposition of all C_α carbons of the UDP-GlcNAc-bound GlmU-Tr onto those of the free structure results in an average r.m.s. deviation of 0.6 and 0.8 Å for the two molecules, respectively, indicating that

Fig. 4. (A) Ribbon diagram of the N-terminal domain showing the first lobe (Asn3–Val111) in blue and the second lobe (Glu112–Gln231) in magenta, the bound UDP-GlcNAc as orange bonds, and the consensus sequence motif Gly14–Lys25 (with Arg18 and Lys25 side chains) in red. (B) Electrostatic potential mapped onto the molecular surface of the N-terminal domain of GlnU (residues Asn3–Asn227) from -7 kT (red) to $+7$ kT (blue) with bound UDP-GlcNAc within the surface pocket. The bottom of the pocket has the most electronegative surface potential (centre), with two discrete patches of electropositive potential [Lys156 and Arg153 (centre left) and Lys15 and Arg18 (centre right)] surrounding the pocket entrance. The UDP-GlcNAc molecule is displayed as white bonds with red oxygen, blue nitrogen and magenta phosphorous as spheres, as well as a sulfate molecule (green bonds). (C) A close-up stereo view of the N-terminal domain, viewed in the same orientation as (A), with UDP-GlcNAc (orange bonds), sulfate (green bonds) and side chain residues Arg18, Lys25 (red), Leu11, Ala13, Gln76 and Asp105 (nucleotide-binding site, pink), Gly81, Asp105, Tyr139, Glu154, Asn169, Tyr197 and Thr199 (sugar-binding site, green), and Gln193, Glu195, Asn227 and Gln231 (yellow). (D) Superposition of the N-terminal domain of apo GlnU and the GlnU–UDP-GlcNAc complex. Backbone regions that deviate significantly are highlighted with their associated side chains (yellow for apo and green for the complex). A rigid-body motion of regions Val131–Gln166 and Val187–Tyr197 is observed, along with significant movements of the Tyr103, Tyr139 and Gln193 side chains within the active site.

significant structural changes occur upon UDP-GlcNAc binding (Figure 4D). The side chain of the sequence-conserved Gln193 is moved by 8 Å even though this residue is not directly involved in UDP-GlcNAc binding.

The Gln193 and Glu195 side chains are located in the second lobe, 5 Å away from the P_{α} phosphate group of UDP-GlcNAc, and face Arg18, suggesting that they are not necessary for UTP recognition but could be

Fig. 5. (A) Schematic figure showing the main interactions between GlmU and UDP-GlcNAc. (B) Amino acid alignment of residues in the regions Ala13–Lys25, Gln76–Thr82 and Met101–Val106 (nucleotide binding) and regions Gly138–Arg141, Val153–Lys156 and Ile168–Gly171 (sugar binding) of *E. coli* GlmU (accession No. P17114) compared with eight XDP-sugar PPases: UDP-Glc (P25520), dTDP-Glc (P55253), GDP-Man (P37741), ADP-Glc (P00584), CMP-NeuAc (P13266) and CMP-KDO (P42216). (+) denotes the residues critical for UDP-GlcNAc binding and (*) those mutated in the present study. A dash means that no significant homology was found in the corresponding region.

involved in magnesium binding, which is known to be critical for the uridylyltransferase activity of GlmU and other XDP-sugar PPases (Fukui *et al.*, 1993; Mengin-Lecreux and van Heijenoort, 1993). We conclude from these data that Lys25 and Asn227 stabilize the UTP substrate through hydrogen bonding, and that Arg18 could be the catalytic residue. Thus, GlmU appears to utilize the α -helical arm as an active site cap, as a helix dipole and as an oligomerization domain with a probable role for sequence-conserved Asn227 and possibly Gln231 for recognition of the P_{β} and P_{γ} phosphate

groups of UTP. The topology of the active site is consistent with the proposed mechanism of stereochemical inversion at the P_{α} position for the homologous UDP-Glc PPase, based on NMR spectra of thio-substituted nucleotides (Sheu and Frey, 1978; Sheu *et al.*, 1979), indicating that a similar mechanism may occur for GlmU. Based on our crystal structure, GlmU contains two close binding sites allowing the ternary GlmU–GlcNAc-1-P-UTP complex to form, resulting in the formation of a P–O bond and departure of a pyrophosphate from each side of the central P_{α} phosphate group.

Structural conservation in other XDP-sugar pyrophosphorylases

Sequence comparison and hydrophobic cluster analysis of bacterial XDP-sugar PPases, such as dTDP-Glc, GDP-Man, UDP-Glc, ADP-Glc and CDP-Glc, and to some extent CMP-KDO and CMP-NeuAc PPases, reveal close similarities with the N-terminal domain of GlmU, with sequence identity ranging from 24 to 12%, which suggests that these proteins may share a similar fold (B.Henrissat and Y.Bourne, unpublished results). Large differences are observed in loops connecting secondary structure elements, indicating that subtle amino acid changes within the active site pocket are responsible for differences in substrate specificity. The functionally important residues, such as GlmU Gly81, Asp105, Glu154, Gly140 and Gly171, are remarkably conserved in most related bacterial XDP-sugar PPases, as shown for both the nucleotide- and sugar-binding sites (Figure 5); hence, the structure of the N-terminal domain of GlmU represents a prototypic structure for studying the large family of bacterial XDP-sugar PPases. A single amino acid substitution at position 82, Thr→Leu, would be predicted to alter the specificity of UDP-GlcNAc PPase from GlcNAc into Glc, while a Glu76→Glu substitution in GlmU would be required for GTP binding. The recently determined crystal structure of *E.coli* CMP-KDO synthetase shows a similar phosphate-binding site, also located in surface loop β 1– α 1 (Jelakovic *et al.*, 1996). However, in this class of PPases that comprise CMP-NeuAc synthase, Arg10 and Lys19 are catalytically important residues as shown by chemical modification and site-directed mutagenesis. Whereas CMP-KDO Lys19 corresponds to GlmU Lys25 for UTP recognition, CMP-KDO Arg10 is directed towards the pocket, whereas the equivalent residue in GlmU, Ala13, points toward the protein and is involved in pyrimidine recognition through its backbone. Therefore, slight differences occur in the catalytic site of these two distinct classes of enzymes.

Saccharomyces cerevisiae and human UDP-GlcNAc PPases are larger than GlmU (up to 500 residues) but they have conserved most of these functionally important residues, indicating that evolution has favoured the conservation of this fold. Mutagenesis data on Uap1p, which shows Gly112, Arg116 and Lys123 (GlmU Gly14, Arg18 and Lys25) as possible catalytic residues (Mio *et al.*, 1998), are consistent with this observation. The structure of the consensus sequence motif is probably conserved in human UDP-Glc PPase, where a Gly115→Asp (GlmU Gly14) mutation dramatically impairs its activity. This mutation may affect UDP-Glc binding, but the consensus sequence motif of GlmU does not share the structural homology with the short active site peptide NGGLG of glycogen phosphorylase that was proposed for human UDP-Glc PPase (Flores-Diaz *et al.*, 1997). We can anticipate that the eukaryotic enzymes possess similar nucleotide- and sugar-binding sites located in distinct protein regions. The fact that a small insertion in the C-terminal region of human UDP-Glc PPase alters the sugar specificity is consistent with the present GlmU structure. However, the bifunctionality of the bacterial GlmU enzyme is not conserved in eukaryotes, which evolved with two distinct enzymes: a PPase enzyme, and an acetyltransferase resembling the GCN5-related *N*-acetyltransferases fold (Neuwald and Landsman, 1997)

and not the L β H domain. Clearly, further mutagenesis, kinetic and structural studies will be necessary to assess the roles of the functionally important residues in the catalytic mechanism and the specificity of XDP-sugar PPases.

The L β H trimeric domain fold and comparison with homologous structures

The oligomeric structure of most enzymes containing the L β H domain fold is trimeric; the elution volume for either full-length GlmU or GlmU-Tr from a gel filtration column is consistent with this structural assignment (data not shown). The crystal structure of GlmU-Tr also shows a trimeric arrangement of subunits situated around a crystallographic 3-fold rotation axis (Figure 6A). The overall dimensions of the trimer are $91 \times 94 \times 68$ Å, with the shortest dimension being parallel to the 3-fold axis. An extensive area of the surface of each monomer participates in monomer–monomer interactions, with 1800 Å² of each monomer, or 13% of the entire GlmU-Tr molecular surface per subunit, buried to a 1.6 Å radius probe upon trimer formation. Critical to the trimer assembly of GlmU are the protruding α -helical arms, which cement the trimer by forming a tight interface with the three C1 coils and thus, overall, form an equilateral triangle seated on top of the L β H trimeric structure (Figure 6A). The long α -helical arms project each N-terminal domain of the trimer away from the 3-fold axis, with their centres of gravity 55 Å apart, thus preventing any interaction between the three N-terminal domains.

The trimer interface involves 24 residues, three of which are in the N-terminal domain, seven in the α -helical arm and the remainder in the C-terminal L β H domain, and is stabilized by hydrogen bonds, salt bridges and hydrophobic interactions. A key salt bridge forms between the side chains of the sequence-conserved Arg229 and Asp271, located at the tip of the α -helical arm, and holds the GlmU trimer assembly. The side chain of L β H domain Arg259, located near the 3-fold axis, forms van der Waals interactions with its symmetry-related homologues within the trimer, and is salt-bridged to Asp261 and hydrogen-bonded to Asn277 of a neighbouring subunit. Similarly, residues Tyr312, located at the tips of loops T2, pack on each other within the trimer and contribute to the trimer interface. Residues that participate in stabilizing the L β H trimer are type-conserved substituted within known amino acid sequences of GlmU, suggesting that the present trimeric arrangement may be conserved for other GlmU proteins (Figure 1).

The trimeric association of this particular L β H domain is highly conserved between GlmU-Tr and LpxA, PaXAT, DapD and Cam, with r.m.s. deviation values of 1.28, 1.65, 2.08 and 2.23 Å for 143, 107, 167 and 118 C α carbon atoms, respectively. Compared with these structures, the L β H domains of GlmU-Tr adopt a nearly parallel arrangement (within 1 – 2°), as observed in LpxA and DapD structures, and, in its truncated version, does not contain any insertion loops. These five structures contain a similar number of coils (from seven to 10), with the exception of Cam which contains only four coils, resulting in slight differences in the parallel packing of the three subunits (Figure 6B). Sequence analysis of GlmU reveals that the

Fig. 6. (A) Ribbon diagram of the GlmU trimer viewed along (left) and perpendicular to (right) the LβH axis. (B) Comparison of GlmU trimeric structure with the four known structures containing a trimeric LβH domain viewed parallel to the LβH axis: DapD with bound CoA (2tdt) (Beaman *et al.*, 1997), Cam with bound Zn ion (1thj) (Kisker *et al.*, 1996), LpxA (1lxa) (Raetz and Roderick, 1995) and PaXAT with bound desulfo CoA (2xat) (Beaman *et al.*, 1998) (clockwise from top left). The LβH domain is coloured in magenta, the linker in green and the insertion loops in blue.

hexapeptide repeat motif could span over residues 251–424 with two insertion loops that disobey the hexapeptide repeat sequence rule. We constructed a model of a complete LβH domain of GlmU based on the existing structure. In this model, 2100 Å² of each monomer are buried to a 1.6 Å radius probe, a value in agreement with that found for other acetyltransferase trimeric structures. The two insertion loops, which were not modelled, are made up of segments Leu332–Ala353 and Asp374–Lys394 and are inserted in turn T3 and T1 of coils C5 and C7, respectively; each contains a sequence-conserved motif Gly345–Asn–

Phe–Val–Glu349 and Asn386–Tyr–Asp–Gly389 which could be important for acetyltransferase activity. Comparison of these homologous structures also reveals that the localization of the substrate acetyl-CoA is positioned similarly in DapD (Beaman *et al.*, 1997) and PaXAT (Beaman *et al.*, 1998), where the substrate is located between two subunits on the exterior face of the trimeric LβH domains, a positioning that could also be adopted for GlmU. In these structures, acetyl-CoA binds in a tunnel created by the presence of Gly residues at the *i* + 1 position of the β-sheet PB2, a residue also found in the

L β H domain of GlmU. In addition, the two insertion loops could complete the acetyl-CoA-binding site of GlmU, with loop I emerging from one subunit and loop II from a second subunit, as observed in DapD (Beaman *et al.*, 1997). Together with the fusion of an N-terminal domain with an unrelated fold, these loops may represent an important means by which hexapeptide proteins have attained their structural and functional diversity. The acetyl-CoA molecule is oriented with the direction of the long acyl chain roughly parallel to the L β H axis, with the nucleotide moiety directed towards the C-terminal end while the *N*-acetyl group is directed towards the top, 30 Å away from the N-terminal domain, an orientation that could allow GlcNAc-1-P to reach the pocket within a small channel. The localization of the putative acetyl-CoA-binding site in GlmU is consistent with recent mutagenesis data showing that mutations of the Cys307 and Cys324 residues, which are located only 12 Å away from the acetyl-CoA-binding site, dramatically affect the acetyltransferase activity (Pompeo *et al.*, 1998). Yet, these two cysteine residues are directed inwards, towards the axis of the prism, and are disulfide-bridged; a direct explanation of their role for the acetyltransferase activity of GlmU will only be achieved using full-length protein from other bacterial species, or engineered domains designed from the present structure.

Conclusion

The structure of a truncated form of GlmU, a bifunctional enzyme, shows how the enzyme binds UDP-GlcNAc and reveals precisely the nature of the two domains which are responsible for the acetyltransferase and uridylyltransferase activities. Importantly, the structure provides a prototypic template for structure–function analysis of the catalytic domains of a superfamily of pyrophosphorylase enzymes and suggests an explanation for the mutagenesis data on other XDP-sugar enzymes, whose catalytic domains are likely to be structurally related to that of GlmU. Finally, the structure of GlmU opens the way for future studies on its function and the development of new antibiotics.

Materials and methods

Site-directed mutagenesis

The pFP3 plasmid allowing high level overexpression of the *E. coli* *glmU* gene product (N-terminal His₆-tagged form) under the control of the isopropyl- β -D-galactopyranoside (IPTG)-inducible *trc* promoter has been described previously (Pompeo *et al.*, 1998). pFP3 derivative plasmids for the expression of mutant G14A, R18A and K25A GlmU proteins were constructed using the Transformer™ site-directed mutagenesis kit (Clontech), based on the method of Deng and Nickoloff (Deng and Nickoloff, 1992; Pompeo *et al.*, 1998). The sequences of the oligonucleotides chosen were 5'-GATCCTTGCCGCGGCCAAAGGCA-CGCGCATG-3' (G14A), 5'-GGCAAAGGCACGGCAATGTATTCCG-ATCTT-3' (R18A) and 5'-ATGTATTCCGATCTTCCTGCAGTGCT-GCATACCTTGCCGGG-3' (K25A), for the replacement of amino acids by alanine residues (codons in bold) at the positions indicated, and 5'-TGGTTGAGTATTCACCATC-3' for suppression of the unique *ScaI* site lying within the ampicillin resistance gene of the target plasmid pFP3. Plasmid pFP3-Tr331 for expression of the truncated GlmU protein carrying only the first 331 amino acids was constructed as follows: two oligonucleotide primers 5'-GGACGGGATCCTTGAATAATGCTAT-GAGCGTAGTGA3' and 5'-CTCAGCTGCAGGACGCAATCAGGCA-AACCG-3' were used to amplify by PCR the truncated form of the gene from the chromosome (new stop codon underlined) and the resulting DNA was cut with *Bam*HI and *Pst*II (sites in bold) and inserted between the corresponding sites of vector pTrc His30 (Pompeo *et al.*, 1998).

Expression and purification

Full-length GlmU was purified from the overproducing strain [BL21(DE3)/pET22b-*glmU*] according to Gehring *et al.* (1996) with modifications. After centrifugation of the cell debris, the supernatant was chromatographed on a Pharmacia Resource Q column (6 ml) using a 0–1 M KCl gradient in 50 mM Tris–HCl pH 8.0, 2 mM EDTA, 1 mM dithiothreitol (DTT). The GlmU-containing fractions (as attested by SDS–PAGE) were pooled, concentrated on an Amicon YM10 membrane and loaded on a Pharmacia Hiload 26/60 Superdex 200 column equilibrated in 10 mM Tris–HCl pH 8.0, 150 mM KCl. The fractions were analysed by SDS–PAGE, N-terminal sequencing and mass spectrometry. Purification typically yielded 12.5 mg/500 ml culture medium. A truncated form of GlmU (Δ 332–456) was obtained by allowing proteolysis to occur at 20°C for 2 weeks; it was purified by gel filtration only.

For the mutants, *E. coli* cells [DH5 α or JM83 *glmU::kan* (Pompeo *et al.*, 1998)] carrying plasmids described in the present work were grown exponentially up to OD₆₀₀ = 0.1 at 37°C in 2YT-ampicillin medium. IPTG was added to a final concentration of 1 mM, and growth was continued for 3 h. Cells were harvested on ice and washed with 40 ml of ice-cold 20 mM potassium phosphate pH 7.2, 0.5 mM MgCl₂, 0.1% β -mercaptoethanol. The cell pellet was suspended in 5 ml of the same buffer and supplemented with: 1 μ M leupeptin, 1 mM benzamide, 1 mM phenylmethylsulfonyl fluoride and 20 μ g/ml of trypsin inhibitor. Cells were disrupted by sonication on ice and the resulting suspension centrifuged at 4°C for 30 min at 200 000 g. The supernatant was dialysed finally against 100 vol of the same buffer. Proteins in these crude extracts were analysed by SDS–PAGE and quantified using bovine serum albumin (BSA) as a standard (Bradford, 1976). The wild-type and mutant His₆-tagged enzymes were purified in a single step and under native conditions according to the manufacturer's recommendation (Qiagen). The His₆-tagged wild-type and mutant GlmU enzymes were in all cases 90% pure, as estimated by SDS–PAGE.

Enzymatic assays

Glucosamine-1-phosphate acetyltransferase activity. The standard assay mixture (100 μ l) contained 50 mM Tris–HCl pH 8.2, 1.2 mM GlcN-1-P, 0.5 mM [¹⁴C]acetyl-CoA (1 kBq), 3 mM MgCl₂ and enzyme (0.5 ng to 1 μ g of protein, depending on overexpression or purification yields).

N-Acetylglucosamine-1-phosphate uridylyltransferase activity. The standard assay mixture (100 μ l) contained 50 mM Tris–HCl pH 8.2, 2 mM UTP, 0.2 mM [¹⁴C]GpcNac-1-P (500 Bq), 3 mM MgCl₂ and enzyme (0.01–0.1 μ g of protein). [¹⁴C]GpcNac-1-P (1.9 GBq/mmol) used in this assay was synthesized from [¹⁴C]acetyl-CoA using the acetyltransferase activity of pure GlmU. In both cases, appropriate dilutions of the enzyme were performed in 20 mM potassium phosphate pH 7.2, 0.5 mM MgCl₂, 1 mg/ml BSA, 0.1% β -mercaptoethanol. The samples were incubated at 37°C for 30 min and the reaction was stopped by the addition of 10 μ l of glacial acetic acid. Reaction products were separated by high-voltage electrophoresis on Whatman 3MM filter paper in 2% formic acid pH 1.9 for 2 h at 40 V/cm with an LT36 apparatus (Savant Instruments, Hicksville, NY). The radioactive spots were located by overnight autoradiography using type R2 films (3M, St Paul, MN) or with a radioactivity scanner (Multi-Tracemaster LB285); they were cut out and counted in a Betamatic IV liquid scintillation spectrometer (Kontron Instruments) with a solvent system consisting of 2 ml of water and 13 ml of Aqualyte mixture (J.T. Baker Chemicals). One unit of enzyme activity was defined as the amount which catalysed the synthesis of 1 μ mol of product in 1 min.

Complementation of the *glmU* mutation

The thermosensitive mutant UGS83 (Mengin-Lecreulx and van Heijenoort, 1993) was transformed by the plasmids to be tested. Competent cells (300 μ l) mixed with plasmid DNA were kept on ice for 3 h before heating for 3 min at 42°C. 2YT medium (400 μ l) was then added and cells were incubated at 30°C for >2 h to allow expression of plasmid-coded genes. Aliquots (100 μ l) of the cell suspensions were plated onto two pre-warmed 2YT-ampicillin plates, one incubated at 30°C and the other at 42°C. Growth was observed after 24 h of incubation.

Crystallization

Crystals of the spontaneously proteolysed GlmU-Tr were obtained at 20°C using the vapour diffusion technique. Typically, 2 μ l of the protein solution (25 mg/ml) were mixed with 2 μ l of the reservoir solution made up of 1.3–1.5 M ammonium sulfate, 0.1 M MES pH 6.0, 4% acetone (v/v). Crystals usually appeared within 3–7 days with typical dimensions of 0.3×0.3×0.5 μ m³. The GlmU-Tr–mercury derivative was obtained

by the addition of mercury II acetate to the protein solution to a final concentration of 2 mM. The complex was incubated overnight on ice before crystallization and crystals were observed after 7 days. The GlmU–UDP–GlcNAc complex was obtained after soaking apo GlmU–Tr crystals in the reservoir solution supplemented with 10 mM UDP–GlcNAc.

Data collection and processing

The GlmU–Tr crystals belong to the space group R32 with unit cell dimensions $a = b = 142.68$ Å and $c = 248.13$ Å. They contain two GlmU–Tr molecules per asymmetric unit, giving a V_m value of 3.38 Å³/Da and a solvent content of ~63% (Matthews, 1968). Crystals selected for data collection were transferred in reservoir solution containing increasing amounts of ethylene glycol, flash cooled at 100 K in the nitrogen gas stream and stored in liquid nitrogen. The data collection statistics are shown in Table I. The X-ray fluorescence and transmission from the GlmU–Tr–Hg co-crystal were measured as functions of incident X-ray energy in the vicinity of the mercury L_{III} edge. Two energies were chosen near the absorption edge: 12 288 eV ($\lambda = 1.009$ Å) and 12 335 eV ($\lambda = 1.005$ Å), corresponding to the minimum f' and maximum f'' , respectively. A third, remote energy (remote II) was selected at 15 400 eV ($\lambda = 0.805$ Å). Data were indexed and integrated by DENZO (Otwinowski, 1997). A fourth data set (referred to as remote I), collected on beamline ID14-EH3 (ESRF, Grenoble) at an energy of 13 240 eV ($\lambda = 0.932$ Å), was included in the scaling procedure. The two near-edge and the remote I data sets were scaled to the remote II data set using SCALA and reduced by TRUNCATE. Estimates of FM, the optimized value of the normalized anomalous scattering, were calculated by the program REVISE; two mercury sites were obtained by SHELX (Sheldrick, 1990) from the corresponding Patterson.

Phasing, model building and refinement

Phases were calculated from these two mercury sites by MLPHARE (CCP4, 1994) and improved by the techniques of solvent flattening as implemented in the program DM. The correct hand was identified by inspection of the electron density maps, obtained within 15 min of completing the data collection. A preliminary C_α model was constructed in the solvent-flattened maps using the graphics program TURBO-FRODO (Roussel and Cambillau, 1991). Multi-domain, 2-fold non-crystallographic symmetry (NCS) averaging was employed by DM, taking only the two N- and C-terminal subunits, separately, allowing residues Asn4–Gly138 and Ala155–Pro328 to be built. Preliminary refinement was performed by X-PLOR (Brünger *et al.*, 1987) against data collected from the native crystal at 2.25 Å resolution using NCS restraints. After several iterative cycles of refinement and model building, water molecules were placed automatically using the REFMAC/ARP procedure (CCP4, 1994; Perrakis *et al.*, 1997). X-PLOR omit maps from the final model were used to check the N-terminal domain; 10 residues were deleted systematically in each calculation, simulated annealing then being used to reduce model bias. The model was then refined using CNS version 0.5 (Brünger *et al.*, 1998), giving final R_{work} and R_{free} values of 23 and 27.5%, respectively. This model comprises residues Asn3–Ile326 and Asn3–Phe330 for the two molecules, respectively, a MES molecule, two sulfate ions and 360 solvent molecules. The two molecules in the asymmetric unit are essentially identical, with an r.m.s. deviation of 0.47, 0.08 and 0.85 Å for C_α atoms of residues 3–227, 251–326 and 3–326, respectively. High temperature factors and weak electron density are associated with residues Tyr139–Gln166 in both molecules. The apo GlmU–Tr structure, without solvent and cofactors, was used as a starting model for the GlmU–Tr–UDP–GlcNAc complex structure. Rigid-body refinement decreased the R -factor to 28.7% in the 10–3.5 Å resolution range. Fourier difference maps clearly revealed the location of the bound UDP–GlcNAc in the two molecules. Subsequent refinement, alternated with graphic inspection, gave R -factor and R_{free} values of 22.3 and 26.1%, respectively. This model comprises Asn3–Pro328 and Asn3–Ile326 for the two molecules, respectively, four sulfate ions, two ethylene glycol molecules and 233 solvent molecules. The stereochemistry of the two models was analysed with PROCHECK (Laskowski *et al.*, 1993) and WHATIF (Hooft *et al.*, 1996); 90% of the polypeptide backbone dihedral angles were found to lie in the most favourable regions of the Ramachandran plot, with the remainder in allowed regions. The coordinates of the free GlmU and GlmU–UDP–GlcNAc complex have been deposited with the Protein Data Bank. Figure 1 was generated by Alscript (Barton, 1993), Figure 2A with TURBO-FRODO (Roussel and Cambillau, 1991), Figures 3–5 with Molscript (Kraulis, 1991) and Raster3D (Meritt and Murphy, 1994),

Figure 4B with GRASP (Nicholls *et al.*, 1991) and Figure 5A with Ligplot (Wallace *et al.*, 1995).

Acknowledgements

We are grateful to Eric Brown for providing us with the pET22b-*glmU* plasmid and Ana Gonzalez for expert assistance during data collection at the BW7A beamline of DESY. We thank Veronique Zamboni, Sabine Leydier, Nicolas Maguet and Sonia Longhi who initiated this work, and Louis Gastinel, Anne Belaich and Bernard Henrissat for helpful discussions. We thank Pascale Marchot for assistance in data collection and critical reading of the manuscript, and Pascal Arnoux for assistance in preparation of figures. This work was supported in part by the Centre National de la Recherche Scientifique to the UPR-9039 (Marseille) and the EP-1088 (Orsay).

References

- Barton, G.J. (1993) Alscript: a tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.
- Beaman, T., Binder, D., Blanchard, J. and Roderick, S. (1997) Three-dimensional structure of tetrahydrodipicolinate *N*-succinyltransferase. *Biochemistry*, **36**, 489–494.
- Beaman, T., Sugantino, M. and Roderick, S. (1998) Structure of the hexapeptide xenobiotic acetyltransferase from *Pseudomonas aeruginosa*. *Biochemistry*, **37**, 6689–6696.
- Bradford, M. (1976) A rapid and sensitive method for the quantification of the microgram quantities of protein using the principles of protein–dye binding. *Anal. Biochem.*, **72**, 248–254.
- Brünger, A., Kuriyan, J. and Karplus, M. (1987) Crystallographic R -factor refinement by molecular dynamics. *Science*, **235**, 458–460.
- Brünger, A.T. *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905–921.
- Cabib, E., Roberts, R. and Bowers, B. (1982) Synthesis of the yeast cell wall and its regulation. *Annu. Rev. Biochem.*, **51**, 763–793.
- CCP4 (1994) The CCP4 suite: programs for crystallography. *Acta Crystallogr.*, **D50**, 760.
- Deng, W. and Nickoloff, J. (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal. Biochem.*, **200**, 81–88.
- Diekmann, A.B. and Goldberg, E. (1994) Characterization of a human antigen with sera from infertile patients. *Biol. Reprod.*, **50**, 1087–1093.
- Dutka-Malen, S., Mazodier, P. and Badet, B. (1988) Molecular cloning and overexpression of the glucosamine synthetase gene from *Escherichia coli*. *Biochimie*, **70**, 287–290.
- Fisher, W. (1990) Bacterial phosphoglycolipids and lipoteichoic acids. In Kates, M. (ed.), *Glycolipids, Phosphoglycolipids and Sulfoglycolipids*. Plenum Press, New York, NY, pp. 123–234.
- Fleischmann, F. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* RD. *Science*, **269**, 496–512.
- Flores-Diaz, M., Alape-Giron, A., Persson, B., Pollesello, P., Moos, M., von Eichel-Streiber, C., Thelestam, M. and Florin, I. (1997) Cellular UDP-glucose deficiency caused by a single point mutation in the UDP-glucose pyrophosphorylase gene. *J. Biol. Chem.*, **272**, 23784–23791.
- Fukui, T., Kazuta, Y., Katsube, T., Tagaya, M. and Tanizawa, K. (1993) Exploring the active site in UDP-glucose pyrophosphorylase by affinity labelling and site-directed mutagenesis. *Biotechnol. Appl. Biochem.*, **18**, 209–216.
- Gehring, A.M., Lees, W.J., Mindiola, D.J., Walsh, C.T. and Brown, E.D. (1996) Acetyltransfer precedes uridylyltransfer in the formation of UDP-*N*-acetylglucosamine in separable active sites of the bifunctional GlmU protein of *Escherichia coli*. *Biochemistry*, **35**, 579–585.
- Hammond, C. and Helenius, A. (1995) Quality control in the secretory pathway. *Curr. Opin. Cell Biol.*, **7**, 523–529.
- Herscovics, A. and Orlean, P. (1993) Glycoprotein biosynthesis in yeast. *FASEB J.*, **7**, 540–550.
- Holm, L. and Sander, C. (1995) DALI: a network tool for protein structure comparison. *Trends Biochem. Sci.*, **20**, 478–480.
- Hooft, R., Friend, G., Sander, C. and Abola, E. (1996) Errors in protein structures. *Nature*, **381**, 271–271.
- Hove-Jensen, B. (1992) Identification of *tms-26* as an allele of the *gcd* gene, which encodes *N*-acetylglucosamine 1-phosphate uridylyltransferase in *Bacillus subtilis*. *J. Bacteriol.*, **174**, 6852–6856.
- Jelakovic, S., Jann, K. and Schultz, G.E. (1996) The three-dimensional structure of capsule-specific CMP:2-keto-deoxy-manno-octonic acid synthetase from *Escherichia coli*. *FEBS Lett.*, **391**, 157–161.

- Kisker, C., Schindelin, H., Alber, B., Ferry, J. and Rees, D. (1996) A left-handed β -helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J.*, **15**, 2323–2330.
- Kraulis, P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Laskowski, R., MacArthur, M., Moss, D. and Thornton, J. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 91–97.
- Leiting, B., Pryor, K.D., Eveland, S.S. and Anderson, M.S. (1998) One-day enzymatic synthesis and purification of UDP-N-[1-¹⁴C]acetylglucosamine. *Anal. Biochem.*, **15**, 185–191.
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
- Mengin-Lecreulx, D. and van Heijenoort, J. (1993) Identification of the *glmU* gene encoding N-acetylglucosamine-1-phosphate uridylyltransferase in *Escherichia coli*. *J. Bacteriol.*, **175**, 6150–6157.
- Mengin-Lecreulx, D. and van Heijenoort, J. (1994) Copurification of glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphate uridylyltransferase activities of *Escherichia coli*: characterization of the *glmU* gene product as a bifunctional enzyme catalyzing two subsequent steps in the pathway for UDP-N-acetylglucosamine synthesis. *J. Bacteriol.*, **176**, 5788–5795.
- Mengin-Lecreulx, D. and van Heijenoort, J. (1996) Characterization of the essential gene *glmM* encoding phosphoglucosamine mutase in *Escherichia coli*. *J. Bacteriol.*, **271**, 32–39.
- Meritt, E. and Murphy, M. (1994) Raster3D version 2.0—a program for photorealistic molecular graphics. *J. Appl. Crystallogr.*, **D50**, 869–873.
- Mio, T., Yabe, T., Arisawa, M. and Yamada-Okabe, H. (1998) The eukaryotic UDP-N-acetylglucosamine pyrophosphorylases. Gene cloning, protein expression and catalytic mechanism. *J. Biol. Chem.*, **273**, 14392–14397.
- Mio, T., Yamada-Okabe, H., Arisawa, M. and Yamada-Okabe, H. (1999) *Saccharomyces cerevisiae* *gnal*, an essential gene encoding a novel acetyltransferase involved in UDP-N-acetylglucosamine synthesis. *J. Biol. Chem.*, **274**, 424–429.
- Neuwald, A. and Landsman, D. (1997) GCN5-related N-acetyltransferases belong to a diverse superfamily that includes the yeast Spt10 proteins. *Trends Biochem. Sci.*, **22**, 154–155.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Pedersen, L., Benning, M. and Holden, H. (1995) Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. *Biochemistry*, **34**, 13305–13311.
- Perrakis, A., Sixma, T.K., Wilson, K.S. and Lamzin, V.S. (1997) Warp: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Crystallogr.*, **D52**, 448–455.
- Plumbridge, J. and Vimr, E. (1999) Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine and N-acetylneuraminic acid by *Escherichia coli*. *J. Bacteriol.*, **181**, 47–54.
- Pompeo, F., van Heijenoort, J. and Mengin-Lecreulx, D. (1998) Probing the role of cysteine residues in glucosamine-1-phosphate acetyltransferase activity of the bifunctional GlmU protein from *Escherichia coli*: site-directed mutagenesis and characterization of the mutant enzymes. *J. Bacteriol.*, **180**, 4799–4803.
- Raetz, C.R.H. and Roderick, S. (1995) A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science*, **270**, 997–1000.
- Raetz, C.R.H. (1996) Bacterial lipopolysaccharides: a remarkable family of bioactive macroamphiphiles. In Neidhardt, F.C., Lin, E., Low, K., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M. and Umberger, H. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2nd edn. ASM Press, Washington, DC, pp. 104–122.
- Rick, P.D. and Silver, R.P. (1996) Enterobacterial common antigen and capsular polysaccharides. In Neidhardt, F.C., Lin, E., Low, K., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M. and Umberger, H. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2nd edn. ASM Press, Washington, DC, pp. 1035–1063.
- Rossmann, M.G., Liljas, A., Branden, C.-I. and Bansazak, L.J. (1975) Evolutionary and structural relationships among dehydrogenases. In Boyer, I.P.D. (ed.), *The Enzymes*. Academic Press, New York, NY, Vol. 11, pp. 61–102.
- Roussel, A. and Cambillau, C. (1991) The TURBO-FRODO graphics package. In *Silicon Graphics Geometry Partners Directory*. Silicon Graphics Corp., p. 81.
- Sarvas, M. (1971) Mutant of *Escherichia coli* K-12 defective in D-glucosamine biosynthesis. *J. Bacteriol.*, **105**, 467–471.
- Sheldrick, G.M. (1990) Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr.*, **A46**, 467.
- Sheu, K.F. and Frey, P.A. (1978) UDP-glucose pyrophosphorylase. Stereochemical course of the reaction of glucose 1-phosphate with uridine-5′[1-thiotriphosphate]. *J. Biol. Chem.*, **253**, 3378–3380.
- Sheu, K.F., Richard, J.P. and Frey, P.A. (1979) Stereochemical courses of nucleotidyltransferase and phosphotransferase action. Uridine diphosphate glucose pyrophosphorylase, galactose-1-phosphate uridylyltransferase, adenylate kinase and nucleoside diphosphate kinase. *Biochemistry*, **18**, 5548–5556.
- Suntharankar, P., Pastuszak, I., Rooke, A., Elbein, A.D., van de Rijn, I., Canfield, W.M. and Drake, R.R. (1998) Synthesis of 5-azido-UDP-N-acetylhexosamine photoaffinity analogs and radiolabelled UDP-N-acetylhexosamines. *Anal. Biochem.*, **258**, 195–201.
- Ullrich, J. and van Putten, J.P. (1995) Identification of the gonococcal *glmU* gene encoding the enzyme N-acetylglucosamine 1-phosphate uridylyltransferase involved in the synthesis of UDP-GlcNAc. *J. Bacteriol.*, **177**, 6902–6909.
- Vaara, M. (1992) Eight bacterial proteins, including UDP-N-acetylglucosamine acyltransferase (LpxA) and three other transferases of *Escherichia coli*, consist of a six-residue periodicity theme. *FEMS Microbiol. Lett.*, **76**, 249–254.
- van Heijenoort, J. (1996) Murein synthesis. In Neidhardt, F.C., Lin, E., Low, K., Magasanik, B., Reznikoff, W., Riley, M., Schaechter, M. and Umberger, H. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2nd edn. ASM Press, Washington, DC, pp. 1025–1034.
- Wallace, A., Laskowski, R. and Thornton, J. (1995) Ligplot: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.*, **8**, 127–134.
- Wang-Gillam, A., Pastuszak, I. and Elbein, A. (1998) A 17-amino acid insert changes UDP-N-acetylhexosamine pyrophosphorylase specificity from UDP-GalNAc to UDP-GlcNAc. *J. Biol. Chem.*, **273**, 27055–27057.
- Wedekind, J., Frey, P. and Rayment, I. (1995) Three-dimensional structure of galactose-1-phosphate uridylyltransferase from *Escherichia coli* at 1.8 Å resolution. *Biochemistry*, **34**, 11049–11061.
- White, R.J. (1968) Control of amino sugar metabolism in *Escherichia coli* and isolation of mutants unable to degrade amino sugars. *Biochem. J.*, **106**, 847–858.
- Wu, H.C. and Wu, T.C. (1971) Isolation and characterization of a glucosamine-requiring mutant in *Escherichia coli* K-12 defective in glucosamine-6-phosphate synthetase. *J. Bacteriol.*, **105**, 455–466.
- Yoder, M.D., Lietzke, S.E. and Jurnak, F. (1993) Unusual structural features in the parallel β -helix in pectate lyases. *Structure*, **1**, 241–251.

Received April 13, 1999; revised June 11, 1999;
accepted June 15, 1999