Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, purification and preliminary crystallographic analysis of *N*-acetylglucosamine-1-phosphate uridylyltransferase from *Mycobacterium tuberculosis*

Jiang Yin,‡ Craig R. Garen,‡ Maia M. Cherney, Leonid T. Cherney and Michael N. G. James*

Protein Structure and Function Group, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

‡ These authors contributed equally.

Correspondence e-mail: michael.james@ualberta.ca

Received 20 April 2008
Accepted 31 July 2008

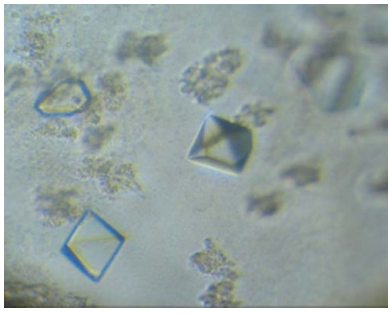
The gene product of open reading frame Rv1018c from *Mycobacterium tuberculosis* is annotated as encoding a probable *N*-acetylglucosamine 1-phosphate uridylyltransferase (*Mtb*GlmU), an enzyme that catalyzes the biosynthesis of UDP-*N*-acetylglucosamine, a precursor common to lipopolysaccharide and peptidoglycan biosynthesis. Following overexpression in *Escherichia coli*, the enzyme was purified and crystallized using the hanging-drop vapour-diffusion method. Native diffraction data were collected from crystals belonging to space group *R*32 and processed to a resolution of 2.2 Å.

1. Introduction

Discovered more than a century ago by Robert Koch, *Mycobacterium tuberculosis* is still a leading cause of human casualties resulting from infectious diseases. More than two billion people are currently infected with *M. tuberculosis*, with new cases accumulating at a rate of nine million per annum. In 2006, 1.7 million people died from tuberculosis alone or in combination with AIDS, according to a report by the World Health Organization (WHO; http://www.who.int/tb).

Despite an initial period of decline in the incidence of tuberculosis (TB) since the introduction of the BCG vaccine and anti-TB drugs, the disease has been resurgent in the past few decades. In 1993, tuberculosis was declared to be a global emergency by the WHO. This resurgence of TB can be linked to several factors. Firstly, HIV infection has created a large highly susceptible population of individuals for whom TB is the most common cause of death owing to their compromised immune systems (Nunn *et al.*, 2005). Furthermore, noncompliance with existing anti-TB therapies has given rise to multi-drug-resistant and extensively drug-resistant *M. tuberculosis* strains, further compounding the already grave situation (Matteelli *et al.*, 2007). The *M. tuberculosis* Structural Genomics Consortium was formed in 2000 (http://www.doe-mbi.ucla.edu/TB) with the goal of providing structural bases for the development of novel effective therapeutics for tuberculosis (Goulding *et al.*, 2002; Smith & Sacchettini, 2003; Terwilliger *et al.*, 2003; Murillo *et al.*, 2007). Such a systematic approach is possible as the complete genome of the best-characterized virulent strain of *M. tuberculosis*, H37Rv, has been sequenced and many of its open reading frames have been annotated (Camus *et al.*, 2002; Cole *et al.*, 1998).

In this study, we report the molecular cloning, protein purification and preliminary X-ray crystallographic analysis of the Rv1018c gene product, which has been annotated as a probable *N*-acetylglucosamine 1-phosphate uridyltransferase (EC 2.7.7.23; *Mtb*GlmU). GlmU is a bifunctional enzyme that catalyzes two consecutive reactions leading to the biosynthesis of UDP-*N*-acetylglucosamine (UDP-GlcNAc). In the first reaction, GlmU catalyzes the CoA-dependent acetylation of glucosamine-1-phosphate (Glc-1-P) to afford *N*-acetylglucosamine-1-phosphate (GlcNAc-1-P). The acetyltransferase domain resides in the C-terminal ∼230 residues of GlmU (253–456 in *Escherichia coli* GlmU numbering), which form a very long left-handed β-helix (LβH) domain. In the second reaction, the UMP moiety from a UTP molecule is transferred to GlcNAc-1-P to give the final product UDP-GlcNAc (Gehring *et al.*, 1996). The N-



© 2008 International Union of Crystallography
All rights reserved

terminal uridyltransferase domain (residues 1–250 in *E. coli* GlmU numbering) displays a Rossmann-like fold with a central seven-stranded β-sheet surrounded by six helices (Brown *et al.*, 1999; Gehring *et al.*, 1996). The acetyltransferase and the uridyltransferase domains have been separately expressed and purified; they both showed their corresponding catalytic activities in the absence of the other domain of the full-length enzyme (Olsen & Roderick, 2001).

2. Experimental methods

2.1. Cloning, expression and purification

The template DNA for Rv1018c amplification by the polymerase chain reaction (PCR) was obtained from a bacterial artificial chromosome (BAC) genomic library of the *M. tuberculosis* H37Rv strain from L'Institut Pasteur (Brosch *et al.*, 1998; Gordon *et al.*, 1999). The PCR primer sequences (Integrated DNA Technologies) for directional cloning of the amplified product into the Gateway cloning system (Invitrogen) were Rv1018cF (5′-GGGACAAGTTTGTACAAAAAAGCAGGCTCC**GAAAACCTGTATTTTCAGGG**ATGACGTTTCCTGGTGACACCGCG-3′) and Rv1018cR (5′-GGGACCACTTTGTACAAGAAAGCTGGGTC-3′). The bold type in the forward primer indicates the nucleotides encoding the cleavage site (ENLYFQ^G) for tobacco etch virus protease (TEV^{pro}). Complete proteolysis by TEV^{pro} leaves a single glycine residue at the amino-terminus as indicated. The PCR product containing the full open reading frame encoding residues 1–495 of Rv1018c was gel-purified and inserted into the pDONR221 plasmid (Invitrogen) to generate an entry clone plasmid pDONR1018c. Subsequently, the coding region of Rv1018c in pDONR1018c was cloned into expression plasmid pDEST-15 (Invitrogen) with an amino-terminal glutathione *S*-transferase fusion partner. The coding sequence of Rv1018c in this expression plasmid (pGST-1018c) was confirmed by DNA-sequencing analysis (DNA Core Facility, Department of Biochemistry, University of Alberta, Canada).

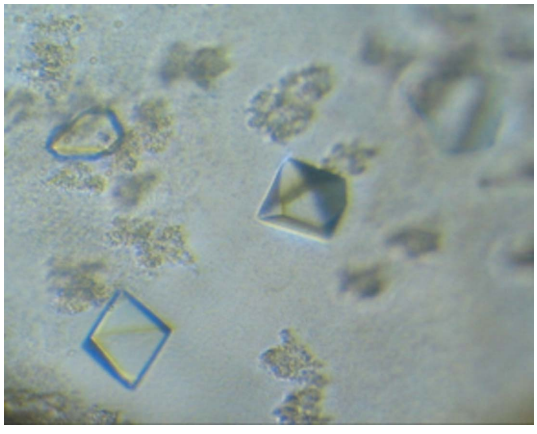


Figure 1
Morphology of Rv1018c crystals. The approximate dimensions of Rv1018c crystals of diffraction quality grown in 20% PEG 3350, 0.15 *M* DL-malate pH 7.0 over a period of two weeks were 40 × 50 × 50 µm.

The expression strain *E. coli* BL21 (DE3) (Novagen) was transformed to ampicillin resistance by pGST-1018c. An overnight culture from a single colony was used to inoculate 2 l Terrific Broth (TB) supplemented with 100 µg ml^{-1} ampicillin. Shaking of the culture at 310 K was continued until its $OD_{600\,nm}$ reached 0.8. Subsequently, the temperature of the culture was shifted to 295 K and protein overexpression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 m*M*. After overnight incubation, the cells were harvested by centrifugation at 7000*g* for 15 min. Bacterial pellets were resuspended in 40 ml buffer *A* (20 m*M* NaH_2PO_4/Na_2HPO_4 pH 7.4, 300 m*M* NaCl, 1 m*M* DTT) supplemented with a tablet of Complete protease-inhibitor cocktail (Roche). For protein purification, the cells were lysed by freeze–thawing and then subjected to ultrasonication in buffer *A*. The lysate was cleared by centrifugation (30 min, 20 000*g*) and the supernatant was loaded onto a 5 ml GSTrap glutathione Sepharose column (GE Healthcare) pre-equilibrated with phosphate-buffered saline (PBS). The GST-Rv1018c fusion protein was competitively eluted from the column with 10 m*M* reduced glutathione (Sigma) in 50 m*M* Tris–HCl pH 8.0, 2 m*M* DTT and 0.02% NaN_3. The GST tag was removed by proteolytic cleavage using recombinant tobacco etch virus protease (AcTEV; Invitrogen). The site recognized by the AcTEV protease is encoded in the forward primer (shown in bold in the sequence above) and leaves a single glycine residue in the P1′ position following complete proteolysis. After overnight dialysis against PBS, the cleaved protein mixture was once again loaded onto a GSTrap column and the flowthrough fractions containing liberated Rv1018c were dialyzed against 10 m*M* Tris–HCl pH 7.4 and 100 m*M* NaCl. The resulting solution was concentrated to ∼12 mg ml^{-1} using an Amicon Ultra filtration unit (5 kDa cutoff; Millipore). All steps of protein purification subsequent to the initial thawing of cells were performed

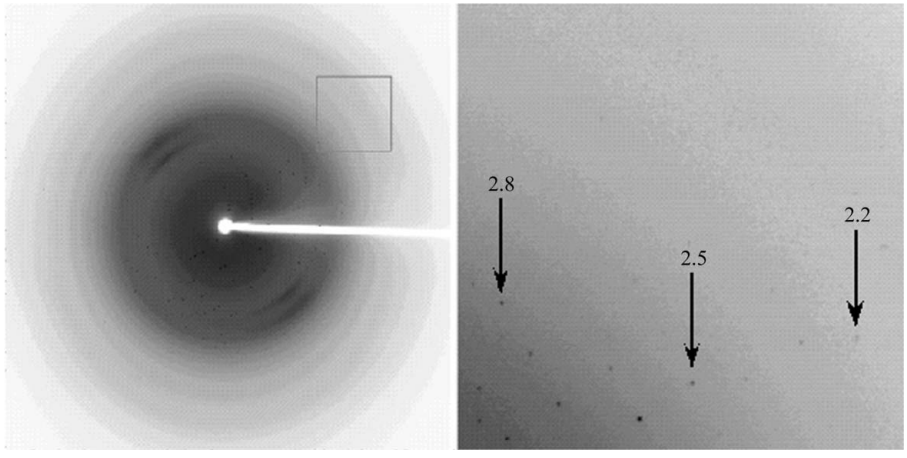


Figure 2
A representative frame showing the diffraction quality of Rv1018c crystals. The boxed area is enlarged and shown on the right; numbers indicate resolutions in Å.

at 277 K and the result of each step was monitored by 15% SDS–PAGE.

2.2. Crystallization

Initial screening of commercially available crystallization conditions was performed on native full-length Rv1018c at 295 K using the sitting-drop vapour-diffusion technique in 96-well Intelliplates (Hampton Research). Crystal Screens I and II and Index Screen (Hampton Research) were applied using equal volumes (0.2 µl) of protein solution and reservoir solution. Periodic checking of the crystallization setups over four months indicated that Rv1018c crystals grew under five conditions, with the largest crystals observed in

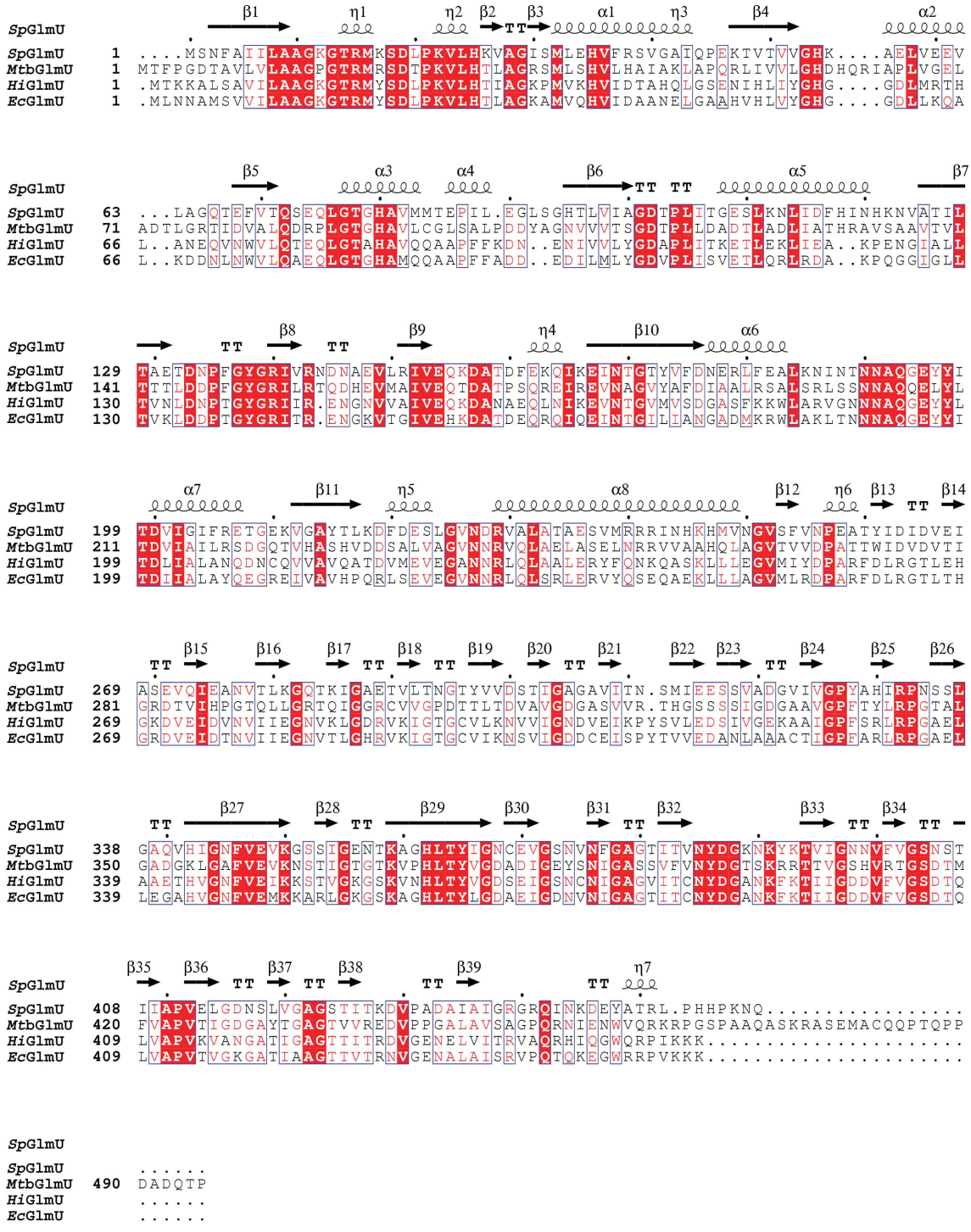


Figure 3
Sequence alignment of *Mtb*GlmU with GlmU enzymes from *Streptococcus pneumonia* (*Sp*GlmU), *Haemophilus influenza* (*Hi*GlmU) and *E. coli* (*Ec*GlmU). The numbering of the alignment corresponds to the sequence of Rv1018c from *M. tuberculosis*. The secondary structures corresponding to *Sp*GlmU (PDB code 1g95) are shown above the sequence alignment. Helices, β-strands and turns are represented by coils, arrows; T represents turns. Conserved residues are shown on a red background, whereas residues with similar chemical properties are shown in red. The identities between *Mtb*GlmU and *Sp*GlmU, *Hi*GlmU and *Ec*GlmU are 40, 39 and 38%, respectively. Sequence alignment was performed using the program *ClustalW* (Thompson *et al.*, 1994) and the figure was generated with the program *ESPript* (Gouet *et al.*, 1999).

Table 1
Crystal parameters and data-collection statistics for native Rv1018c.

Values in parentheses are for the highest resolution shell.

Space group	*R*32
Unit-cell parameters (hexagonal setting)	
$a = b$ (Å)	94.69
c (Å)	285.8
$\alpha = \beta$ (°)	90
γ (°)	120
No. of molecules per ASU	1
Matthews coefficient (Å^3 Da^{-1})	2.4
Solvent content (%)	48.1
Data collection	
Temperature (K)	100
Detector	MAR 225 CCD
Wavelength (Å)	0.979338
Resolution (Å)	50.0–2.2 (2.28–2.20)
Unique reflections	23513
Multiplicity	6.3 (4.0)
$\langle I/\sigma(I)\rangle$	13.8 (2.3)
Completeness (%)	92.8 (74.0)
$R_{\rm merge}$† (%)	10.4 (41.8)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.

drops containing Index Screen condition No. 91. To obtain crystals suitable for X-ray diffraction, equal volumes (1 µl) of protein solution at ∼10 mg ml^{-1} and reservoir solution (20% PEG 3350, 0.15 *M* DL-malate pH 7.0) were mixed and equilibrated in hanging drops against 1 ml reservoir solution on a 24-well VDX plate (Hampton Research). Crystals grew to maximal dimensions of 40 × 50 × 50 µm over two weeks (Fig. 1).

2.3. Data collection

Crystals for synchrotron data collection were first rinsed in cryoprotectant (30% glycerol in reservoir solution) and then flash-cooled in liquid nitrogen. Native data sets were collected on beamline CMCF1 at the Canadian Light Source (CLS). Raw data were indexed, integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

3. Results

Native data sets (Fig. 2) were collected from Rv1018c crystals on beamline CMCF1 of the Canadian Light Source. Crystal parameters and crystallographic statistics of the data sets are summarized in Table 1. According to the calculated molecular weight of Rv1018c including the N-terminal glycine as the result of AcTEV cleavage, 51 641.05 Da, the Matthews coefficient (Matthews, 1968) is calculated to be 2.4 Å^3 Da^{-1} with a solvent content of 48.1%. This result corresponds to the presence of one *Mtb*GlmU molecule per asymmetric unit. The biological assembly of GlmU enzymes from other bacterial species is thought to consist of a homotrimer related by threefold symmetry (Sulzenbacher *et al.*, 2001; Brown *et al.*, 1999; Mochalkin *et al.*, 2007). The available crystal structures of GlmU enzymes show only one (Kostrewa *et al.*, 2001; Mochalkin *et al.*, 2007) or two (Sulzenbacher *et al.*, 2001; Olsen & Roderick, 2001) molecules of GlmU in the asymmetric unit. However, the crystal symmetry of these structures always allows the reconstruction of a GlmU homotrimer *via* the symmetry operations. Most of the inter-protomer interactions within the GlmU trimer are provided by the C-terminal LβH domain. Based on the significant sequence conservation in the C-terminal acetyltransferase domains of *Mtb*GlmU and other bacterial GlmU enzymes (Fig. 3) and the finding that the *Mtb*GlmU crystals belong to space group *R*32, we posit that *Mtb*GlmU also assembles into a crystallographic homotrimer.

X-ray diffraction data were collected on beamline CMCF1 at the Canadian Light Source. Research in the laboratory of MNGJ is funded by the Alberta Heritage Foundation for Medical Research (AHFMR) and the Canadian Institute of Health Research (CIHR). MNGJ holds a Canada Research Chair in Protein Structure and Function. JY is grateful for the support of a postdoctoral fellowship from AHFMR.

References

Brosch, R., Gordon, S. V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B. G. & Cole, S. T. (1998). *Infect. Immun.* **66**, 2221–2229.
Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C. & Bourne, Y. (1999). *EMBO J.* **18**, 4096–4107.
Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). *Microbiology*, **148**, 2967–2973.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Gehring, A. M., Lees, W. J., Mindiola, D. J., Walsh, C. T. & Brown, E. D. (1996). *Biochemistry*, **35**, 579–585.
Gordon, S. V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K. & Cole, S. T. (1999). *Mol. Microbiol.* **32**, 643–655.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Goulding, C. W. *et al.* (2002). *Curr. Drug Targets*, **2**, 121–141.
Kostrewa, D., D'Arcy, A., Takacs, B. & Kamber, M. (2001). *J. Mol. Biol.* **305**, 279–289.
Matteelli, A., Migliori, G. B., Cirillo, D., Centis, R., Girard, E. & Raviglione, M. (2007). *Expert Rev. Anti Infect. Ther.* **5**, 857–871.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mochalkin, I., Lightle, S., Zhu, Y., Ohren, J. F., Spessard, C., Chirgadze, N. Y., Banotai, C., Melnick, M. & McDowell, L. (2007). *Protein Sci.* **16**, 2657–2666.
Murillo, A. C. *et al.* (2007). *Infect. Dis. Drug Targets*, **7**, 127–139.
Nunn, P., Williams, B., Floyd, K., Dye, C., Elzinga, G. & Raviglione, M. (2005). *Nature Rev.* **5**, 819–826.
Olsen, L. R. & Roderick, S. L. (2001). *Biochemistry*, **40**, 1913–1921.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Smith, C. V. & Sacchettini, J. C. (2003). *Curr. Opin. Struct. Biol.* **13**, 658–664.
Sulzenbacher, G., Gal, L., Peneff, C., Fassy, F. & Bourne, Y. (2001). *J. Biol. Chem.* **276**, 11844–11851.
Terwilliger, T. C. *et al.* (2003). *Tuberculosis (Edinb.)*, **83**, 223–249.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.