Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Overexpression, purification, crystallization and data collection on the *Bordetella pertussis wlb*D gene product, a putative UDP-GlcNAc 2′-epimerase

Velupillai Sri Kannathasan,[a]† Adam G. Staines,[b]† Chang Jiang Dong,[b] Robert A. Field,[a] Andrew G. Preston,[c] Duncan J. Maskell[c] and James H. Naismith[b]*

[a]Centre for Carbohydrate Chemistry, School of Chemical Sciences, University of East Anglia, Norwich NR4 7TJ, England, [b]The Centre for Biomolecular Sciences, The University, St Andrews KY16 9ST, Scotland, and [c]Department of Clinical Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, England

† These authors made an equal contribution.

Correspondence e-mail: naismith@st-and.ac.uk

Received 12 May 2001
Accepted 28 June 2001

The *Boredetella pertussis wlb*D gene product is a putative uridine-5-diphosphate *N*-acetylglucosamine (UDP-GlcNAc) 2′-epimerase involved in Band A lipopolysaccharide biosynthesis. The *wlb*D gene is homologous to *Escherichia coli rff*E (32% identical), an established UDP-GlcNAc 2′-epimerase that is involved in enterobacterial common antigen (ECA) formation. The structure of the *rff*E protein reveals an unexpected role for a bound sodium ion in orientating a substrate-binding α-helix in the enzyme active site. Whilst key active-site residues in rffE are present in the *wlb*D sequence, the sodium-binding residues outside the active site are absent. This raises questions about the modulation of enzyme activity in these two enzymes. The *wlb*D gene from *B. pertussis* has been cloned and overexpressed in *E. coli* and the resulting protein has been purified to homogeneity. In the current study, crystals of the mutant Gln339Arg *wlb*D enzyme have been obtained by sitting-drop vapour diffusion. Uncomplexed Gln339Arg and UDP-GlcNAc complex data sets have been collected in-house on a rotating-anode generator to 2.1 Å. Combined, the data sets identify the space group as $P2_12_12_1$, with unit-cell parameters $a = 78$, $b = 91$, $c = 125$ Å, $\alpha = \beta = \gamma = 90^\circ$. The asymmetric unit contains two monomers and 53% solvent.

1. Introduction

B. pertussis is the causative agent of whooping cough in children and, as is increasingly recognized, chronic cough in adults (Cherry, 1996; Nennig *et al.*, 1996; Hewlett, 1997). *Bordetella* lipopolysaccharide (LPS) has been implicated in the infection process by research that shows that it is highly immunogenic, acts as an immunological adjuvant and displays the properties expected of an endotoxin (Amano *et al.*, 1990; Watanabe *et al.*, 1990; Cherry, 1996). Defined *B. pertussis* mutants, differing from parental strains solely in the lack of the Band A trisaccharide from the LPS, are severely compromised in their ability to infect mice, resulting in a more rapid clearance of the mutants from the host. This is the first direct evidence that *B. pertussis* LPS is required for virulence (Harvill *et al.*, 2000).

Tentative gene-function assignments have been made for the *B. pertussis* LPS Band A trisaccharide biosynthesis locus (*wlb*), albeit based on rather weak sequence similarities (Allen & Maskell, 1996; Allen *et al.*, 1998). The assignment of the *wlb*D protein as a UDP-GlcNAc 2′-epimerase arises from its similarity to *rff*E (32% identity), the established *E. coli* UDP-GlcNAc 2′-epimerase involved in enterobacterial common antigen (ECA) formation (Meier-Dieter *et al.*, 1992; Morgan *et al.*, 1997 and references cited therein). However, genetic studies highlight that *Staphylococcus aureus* cap5G (30% identity to *rff*E) fails to complement *E. coli rff*E mutants, whereas cap5P (52% identity to *rff*E) rescues ECA formation in these same mutants (Kiser *et al.*, 1999). In addition, complementation of mutation in *Pseudomonas aeruginosa wbp*I (33% identical to *rff*E but 77% identical to *wlb*) with *E. coli rff*E was unsuccessful (Burrows *et al.*, 2000). Sequence-database searching with *wlb*D yields similarity to 33 proteins identified as putative UDP-GlcNAc 2′-epimerases, with an overall identity of 26% and 27 absolutely conserved residues. Recently, the *rff*E structure has been determined (Campbell *et al.*, 2000) and of the residues identified as important in either UDP binding (Arg10, Ser290 and Glu296) or in the active site (Lys15, Asp95, Glu117, Glu131, Arg135 and His213) only Lys15 is not absolutely conserved in all 33 sequences (it being substituted for an Arg in three of the 33 sequences). Only *rff*E has been definitively established as a UDP-GlcNAc 2′-epimerase. Campbell *et al.* (2000) identified a sodium-binding motif in *rff*E which orientates a substrate-binding helix within the active site. None of these three sodium-binding residues is conserved in *wlb*D and are all absent in a number of other presumed UDP-GlcNAc

2′-epimerases. Taken together with the complementation studies it suggests the precise function of *wbl*D may need to be revised. This is consistent with our inability so far to detect UDP-GlcNAc 2′-epimerase activity in any *wbl*D by assay. Structural biology can play an important role in assigning function to proteins of unknown or unclear function and by solving the structure of *wlb*D we will clarify its function.

We report here the overexpression, crystallization and data collection for crystals of the mutant Q339R *wlb*D enzyme, cocrystals obtained in the presence of the putative substrate UDP-GlcNAc and a selenomethionine form of the protein.

2. Materials and methods

2.1. Expression and purification

The *B. pertussis wlb*D gene was cloned by PCR amplification of genomic DNA and overexpressed in *E. coli* using the IMPACT expression system (New England Biolabs), giving rise to high-level expression (>3 mg l^{-1}). The high GC content of *Bordetella* DNA resulted in random point mutants in the *wlb*D sequence even when high-fidelity polymerases were used. A selection of mutant *wlb*D proteins were thus generated; for the present study, the Gln339Arg mutant was selected on the basis of crystallizability. The selection of such mutants on this basis is now routine and is probably important for high-throughput studies. Residue 339 is remote from the active site and seems unlikely to affect the structure. Studies on the native protein and other mutants are ongoing.

Individual colonies of the transformed *E. coli* strain XL1-Blue (PCYB2 construct) were used to inoculate 1 l of LB broth containing 100 µg ml^{-1} ampicillin. After growing at 310 K to an A_{600} of 0.6–0.8, the culture was induced by addition of 1 m*M* IPTG and allowed to grow for a further 8 h. Cells were harvested by centrifugation (8000*g* for 15 min) and stored at 193 K. Analysis of cell extracts by SDS–PAGE confirmed the high expression level of *wlb*D protein.

All subsequent steps were carried out at 277 K. The frozen cell pellet was thawed rapidly and resuspended in 50 ml lysis buffer (20 m*M* HEPES pH 8.0, 0.5 *M* NaCl, 1 m*M* EDTA, 0.1% Triton X-100) and the cell suspension was broken by sonication with three 40 s bursts with 2 min intervals between bursts. The cell-free extract was obtained by centrifugation (12 000*g* for 20 min). The supernatant was collected and filtered using a 0.2 µm filter. A 15 ml slurry of chitin-coated beads was poured into a column (2.5 × 10 cm) and equilibrated with 50 ml of column buffer (20 m*M* HEPES pH 8.0, 0.5 *M* NaCl, 1 m*M* EDTA, 0.1% Triton X-100). Cell-free extract was slowly loaded onto the chitin column at a flow rate of 0.5 ml min^{-1} and the flowthrough was collected. The column was then washed with column buffer (without Triton X-100) at a flow rate of 1 ml min^{-1} until the protein content of the eluate reached a minimum (OD_{595} < 0.05 as measured by Bradford assay). Cleavage of the *wlb*D protein from the column-bound tag and elution was achieved with the column buffer containing 30 m*M* DTT. The effluent was fractionated and analysed by SDS–PAGE (10% gel); a single major band migrating at an apparent mass of 40 kDa was observed. MALDI-TOF mass spectrometry on a tryptic digest of this material confirmed the identity of the protein. Relevant fractions from the chitin column were pooled and concentrated at 193 K to 5 mg ml^{-1} with an Amicon concentrator. The *wlb*D Gln339Arg protein proved inactive as a UDP-GlcNAc 2′-epimerase, as judged by HPLC and NMR assays (Morgan *et al.*, 1997).

N-terminal sequencing indicated that the terminal methionine residue had been removed during expression (observed sequence: PKKIL). Protease digestion and MALDI-TOF analysis indicated the incorporation of a Pro-Gly C-terminal extension derived from the PCYB2 vector (calculated mass = 39 368 Da; observed mass = 39 377 Da).

Selenomethionine-enriched *wlb*D protein was produced by inhibiting methionine biosynthesis and feeding selenomethionine (Doublie, 1997) in *E. coli* BL21 plysS. The protein was purified using the same protocol as used for Gln339Arg, except that all buffers were rigorously degassed and contained 5 m*M* DTT in order to prevent selenomethionine oxidation. The mass spectrum of selenomethionine-enriched protein showed only one major peak with an observed increase of 369 Da, consistent with a full incorporation of eight selenomethionine residues (calculated mass = 376 Da).

2.2. Crystallization

Crystallization trials of *wlb*D protein were conducted with 96-condition Hampton Screens 1 and 2 (Jancarik & Kim, 1991; Cudney *et al.*, 1994) at 294 K. Crystals were obtained by vapour equilibration of sitting drops (Ducruix & Giegé, 1992) of protein (5 µl at 5 mg ml^{-1}) and precipitant (5 µl) within 5 d. Conditions were optimized to give crystals suitable for diffraction from 0.1 *M* Tris–HCl pH 7.5, 8.0 and 8.5, 4.0 *M* sodium formate. The *wlb*D protein was also cocrystallized with UDP-GlcNAc (10 m*M*), following overnight pre-incubation at 277 K, using the sitting-vapour diffusion method. Crystals were obtained within a week in the presence of 0.1 *M* Tris–HCl pH 8.6, 3.5 *M* sodium formate. Crystals of selenomethionine-enriched *wlb*D protein grew in 0.1 *M* Tris–HCl pH 8.5, 4.0 *M* sodium formate within 10 d.

Table 1
Data collection of Gln339Arg mutant *B. pertussis* UDP-GlcNAc 2′-epimerase.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.54	1.54
Temperature (K)	130	298
Detector	DIP2000	R-AXIS IV++
Resolution	20–2.2 (2.24–2.20)	20–2.7 (2.75–2.70)
Space group	$P2_12_12_1$	
Unit-cell parameters (Å)	a = 77.8, b = 90.4, c = 123.5	a = 79.0, b = 91.1, c = 125.3
Unique reflections	44535	25807
$I/\sigma(I)$	43.4 (4.3)	26.9 (3.2)
Completeness (%)	99.6 (99.9)	96.3 (81.3)
R_{merge} (%)	5.1 (32.6)	9.0 (26.3)

2.3. Data collection

A 2.7 Å resolution data set was collected in-house from a single crystal mounted in a glass capillary. The data were collected on an R-AXIS IV++ detector with a Rigaku rotating-anode generator (5 kW, 100 mA) as 1° 5 min exposures totalling 140° of non-overlapping data. Additionally, a 2.3 Å data set was collected in-house from a single crystal cocrystallized with UDP-GlcNAc and mounted in a glass capillary. As before, data were collected on an R-AXIS IV++ detector with a Rigaku rotating-anode generator as 1° 5 min exposures totalling 150° of non-overlapping data. Both sets of oscillation images were indexed and integrated with the program *DENZO* (Otwinowski & Minor, 1996) and scaled with *SCALEPACK* (Otwinowski & Minor, 1996). A 2.2 Å resolution data set from a flash-frozen crystal of the protein was measured in-house. Crystals were cryoprotected by immersing a cryoloop (containing the crystal in the mother liquor) in paraffin oil for 10 s prior to rapid cooling to 120 K for data collection. The data were collected on a Nonius DIP2000 image-plate rotating-anode system as 1° 7 min exposures totalling 275° of non-overlapping data. A crystal cocrystallized with UDP-GlcNAc was frozen as above. The data were collected

Table 2
Data collection of *B. pertussis* UDP-GlcNAc 2′-epimerase cocrystallized with UDP-GlcNAc.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.54	1.54
Temperature (K)	130	298
Detector	DIP2000	R-AXIS IV++
Resolution	25–2.1 (2.14–2.1)	24–2.3 (2.34–2.3)
Space group	$P2_12_12_1$	
Unit-cell parameters (Å)	a = 77.7, b = 90.5, c = 123.2	a = 78.90, b = 91.1, c = 125.3
Unique reflections	51182	41485
$I/\sigma(I)$	15.0 (2.8)	13.7 (1.9)
Completeness (%)	98.6 (97.2)	85.4 (65.1)
$R_{\rm merge}$ (%)	8.6 (49.3)	12.6 (29.8)

on a Nonius DIP2000 image-plate rotating-anode system as 1° 7 min exposures totalling 123° of non-overlapping data.

Full details of data collection are given in Tables 1 and 2. Together, the four data sets measure more than 20 reflections along each of the three axis. The pattern of absences shows that the space group is $P2_12_12_1$. So far, all our attempts at molecular replacement using *rff*E as a model have failed. We will use multiwavelength methods to solve the structure.

This work was supported by the Wellcome Trust and the BBSRC. We are grateful to Professors M. E. Tanner and N. C. J. Strynadka (University of British Columbia) for providing a clone of *rff*E and for the co-ordinates of the protein crystal structure.

References

Allen, A. G. & Maskell, D. J. (1996). *Mol. Microbiol.* **19**, 37–52.
Allen, A. G., Thomas, R. M., Cadisch, J. T. & Maskell, D. J. (1998). *Mol. Microbiol.* **29**, 27–38.
Amano, K., Fukushi, K. & Watanabe, M. (1990). *J. Gen. Microbiol.* **136**, 481–486.
Burrows, L. L., Pigeon, K. E. & Lam, J. S. (2000). *FEMS Microbiol. Lett.* **189**, 135–141.
Campbell, R. E., Mosimann, S. C., Tanner, M. E. & Strynadka, N. C. J. (2000). *Biochemistry*, **49**, 14993–15001.
Cherry, J. D. (1996). *J. Infect. Dis.* **174**, S259–S263.
Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.
Doublie, S. (1997). *Methods Enzymol.* **276**, 523–530.
Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins, a Practical Approach.* Oxford: IRL Press.
Harvill, E. T., Preston, A., Allen, A. G., Cotter, P. A., Maskell, D. J. & Miller, J. F. (2000). *Infect. Immun.* **68**, 6720–6728.
Hewlett, E. L. (1997). *Pediatr. Infect. Dis. J.* **16** (Suppl.), S78–S84.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kiser, K. B., Bhasin, N., Deng, L. Y. & Lee, J. C. (1999). *J. Bacteriol.* **181**, 4818–4824.
Meier-Dieter, U., Barr, K., Starman, R., Hatch, L. & Rick, P. D. (1992). *J. Biol. Chem.* **267**, 746–753.
Morgan, P. M., Sala, R. F. & Tanner, M. E. (1997). *J. Am. Chem. Soc.* **119**, 10269–10277.
Nennig, M. E., Shinefield, H. R., Edwards, K. M., Black, S. M. & Fireman, B. H. (1996). *J. Am. Med. Assoc.* **275**, 1672–1674.
Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.
Watanabe, M., Takimoto, H., Kumazawa, Y. & Amano, K. (1990). *J. Gen. Microbiol.* **136**, 489–493.