

Bharat G. Reddy,^a Derek B. Moates,^b Heung-Bok Kim,^c Todd J. Green,^d Chang-Yub Kim,^c Thomas C. Terwilliger^c and Lawrence J. DeLucas^{e*}

^aDepartment of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, 1025 18th Street South, Birmingham, AL 35233, USA, ^bDepartment of Biology, University of Alabama at Birmingham, 1025 18th Street South, Birmingham, AL 35233, USA, ^cBioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, ^dDepartment of Microbiology, University of Alabama at Birmingham, 1025 18th Street South, Birmingham, AL 35233, USA, and ^eDepartment of Optometry, University of Alabama at Birmingham, 1025 18th Street South, Birmingham, AL 35233, USA

Correspondence e-mail: duke2@uab.edu

Received 27 January 2014
Accepted 18 February 2014

PDB reference: Rv3902c, 4o6g



1.55 Å resolution X-ray crystal structure of Rv3902c from *Mycobacterium tuberculosis*

The crystallographic structure of the *Mycobacterium tuberculosis* (TB) protein Rv3902c (176 residues; molecular mass of 19.8 kDa) was determined at 1.55 Å resolution. The function of Rv3902c is unknown, although several TB genes involved in bacterial pathogenesis are expressed from the operon containing the Rv3902c gene. The unique structural fold of Rv3902c contains two domains, each consisting of antiparallel β -sheets and α -helices, creating a hand-like binding motif with a small binding pocket in the palm. Structural homology searches reveal that Rv3902c has an overall structure similar to that of the *Salmonella* virulence-factor chaperone InvB, with an r.m.s.d. for main-chain atoms of 2.3 Å along an aligned domain.

1. Introduction

Mycobacterium tuberculosis (TB) is an aerobic acid-fast Gram-positive bacterium. A waxy coating on the cell surface of the TB organism makes it exceedingly resistant to host defense mechanisms, particularly when present in the alveoli of lung tissue. Despite the wide use of antibiotics and attenuated vaccines, TB is one of the leading causes of death from bacterial infections, killing an estimated 1.3 million people each year (World Health Organization, 2012). Today, the occurrence of multidrug-resistant tuberculosis (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) is becoming more prevalent in both developing and industrialized nations, and has been reported in 84 countries (World Health Organization, 2012). There is a vital need for new vaccines and small-molecule therapeutics to combat these drug-resistant strains of TB. In 1998, researchers sequenced the complete genome of TB and revealed approximately 4000 genes (Cole *et al.*, 1998), which was followed by the establishment of the TB Structural Genomics Consortium under the NIH Protein Structure Initiative in 2000. The TB Structural Genomics Consortium is an effort to expedite the structure determination of all TB proteins in an effort to further the understanding of TB biology (Chim *et al.*, 2011). It is hoped that this structural knowledge will provide a basis for the identification of new protein targets and drug-discovery strategies with which to treat infections caused by TB.

Rv3902c is a TB protein with a molecular weight of 19.8 kDa comprised of 176 amino acids expressed on the same operon as EsxF, EsxE and Rv3903c (<http://www.tbdb.org>). The function of Rv3902c is unknown; however, the products of two genes transcribed along with Rv3902c, EsxE and EsxF, are paralogs to early secreted antigenic target 6 (Esat-6) proteins found in TB (Agarwal *et al.*, 2007). Esat-6 proteins are potent T-cell antigens and play a role in TB pathogenesis (Smith, 2003). Since Rv3902c is expressed on the same operon as Esat-6 proteins, this suggests that Rv3902c may play a role in TB virulence. Here, we present the crystal structure of the TB protein Rv3902c.

2. Materials and methods

2.1. Protein expression and purification

Rv3902c was subcloned into a pVP16 vector consisting of an N-terminal fusion with a 6 \times His tag, maltose-binding protein (MBP) and a *Tobacco etch virus protease* (TEV) cleavage site via the

Table 1
Summary of Rv3902c crystallographic data.

Values in parentheses are for the highest resolution shell.

	Native	Br-soaked
Experiment	Native	SAD
Wavelength (Å)	1.0000	0.9198
Unit-cell parameters (Å)	$a = b = 91.81, c = 54.23$	$a = b = 92.08, c = 54.03$
Space group	$P6_1$	$P6_1$
Resolution (Å)	39.76–1.55 (1.58–1.55)	32.08–1.62 (1.65–1.62)
No. of unique reflections	37938 (1875)	33300 (1646)
Completeness (%)	100 (100)	100 (100)
Multiplicity	14.7 (11.5)	7.5 (6.5)
Mean $I/\sigma(I)$	18.2 (2.1)	11.8 (2.0)
Molecules in asymmetric unit	1	1
Matthews coefficient (Å ³ Da ⁻¹)	3.32	3.32
Solvent content (%)	62.90	62.98
R_{merge}	0.087 (1.373)	0.096 (0.972)
R_{meas}	0.094 (1.510)	0.110 (1.146)
$R_{\text{p.i.m.}}$	0.034 (0.617)	0.054 (0.603)
$CC_{1/2}$	0.999 (0.609)	0.997 (0.597)
Anomalous completeness (%)		91.7 (88.8)
Anomalous multiplicity		3.5 (2.9)
Refinement statistics		
$R_{\text{work}}/R_{\text{free}}$ (%)	16.5/17.7	
No. of atoms		
Protein	1595	
Water	221	
Heavy atoms		18
R.m.s. deviations		
Bond lengths (Å)	0.006	
Bond angles (°)	0.98	
Average B factor (Å ²)		
Protein	22.1	
Water	37.3	
Ramachandran plot statistic, residues in		
Most favored regions	144	
Additionally allowed regions	8	
Generously allowed regions	0	
Disallowed regions	0	
$MolProbity$ statistics		
Score	1.08	
Clashscore	2.95	
Poor rotamers (%)	0	

Gateway cloning method (Invitrogen). It should be noted that owing to the TEV construct used, an additional serine remains at the N-terminus of the protein after TEV cleavage. The resulting construct N-His₈-MBP-linker-TEV+Rv3902c/pVP16 was transformed into *Escherichia coli* BL21(DE3) cells. A seed culture was created *via* the selection of a single ampicillin-resistant colony and inoculation into 25 ml Luria broth (LB) containing 100 µg ml⁻¹ ampicillin (cultured overnight in an incubator shaker set at 250 rev min⁻¹ and 37°C). On the following day, a 1:100 dilution of the seed culture was placed into two 11 LB flasks each containing 100 µg ml⁻¹ ampicillin and grown in an incubator/shaker set at 250 rev min⁻¹ and 37°C for ~4 h. When an OD₆₀₀ of ~0.6 was reached, the flasks were removed from the incubator, cooled on ice to 16°C, and isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM to induce expression. Induced cultures were allowed to grow overnight in an incubator/shaker set at 250 rev min⁻¹ and 16°C.

The following morning, the incubated cells were centrifuged at 8000g and resuspended in 70 ml buffer A (20 mM Tris-HCl pH 8.0, 200 mM KCl, 5% glycerol, 1.4 mM β-mercaptoethanol) with 1× cOmplete Protease Inhibitor (Roche). All subsequent purification steps were performed either on ice or at 4°C. The cells were sonicated and the cellular debris was spun down at 25 000g for 30 min. The supernatant was filtered using a 0.45 µm filter, supplemented with imidazole to a final concentration of 20 mM and batch bound to 5 ml Ni-NTA Superflow beads (Qiagen) pre-equilibrated with buffer A and rocked for 1 h. The Ni-NTA beads were packed into a column

and attached to an ÄKTA FPLC system (GE Healthcare). The Ni-NTA column was washed with five column volumes (CV) of 95% buffer A and 5% buffer B (buffer A plus 1 M imidazole) and eluted at 2 ml min⁻¹ in a gradient from 5 to 100% buffer B over 10 CV. The purity of the peak fractions was analysed *via* SDS-PAGE followed by concentration of the peak fractions to a final volume of 5 ml with Amicon 10 000 NMWL (Millipore) spin concentrators. TEV protease was added to the pooled and concentrated fractions in a 1:100 (protease:protein) ratio to cleave the His-tagged MBP followed by a gentle overnight rocking. The sample was filtered with a 0.45 µm filter and applied onto a Superdex 75 size-exclusion column (26 mm diameter × 60 cm length; flow rate of 2.5 ml min⁻¹) attached to an ÄKTA FPLC system pre-equilibrated with buffer A. The purity of the peak fractions was analyzed *via* SDS-PAGE. Since further purification was required to remove the His-tagged MBP, the peak fractions were applied onto a Ni-NTA column pre-equilibrated with buffer A. Because the purified and cleaved Rv3902c did not bind to the Ni-NTA column, the flowthrough was collected, concentrated to 25 mg ml⁻¹ and then analysed *via* SDS-PAGE. The purified protein was snap-frozen in liquid nitrogen and stored at -80°C.

2.2. Crystallization, data collection and structure determination

Nanocrystallization trials of purified Rv3902c were conducted with kits from Hampton Research, Emerald Bio and Qiagen. Each trial utilized a Gryphon (Art Robbins Instruments) nanodispensing robot with 200 nl drops (1:1 protein:well solution volume ratio) in 96-well sitting-drop plates. Several hits were found and optimized in 1 µl vapour-diffusion hanging drops (1:1 protein:well solution volume ratio), with the final optimized crystallization condition being 1.5 M ammonium sulfate, 200 mM sodium cacodylate pH 6.5. Clusters of hexagonal crystals often appeared 16–48 h later at room temperature. The crystals were manually separated with the Hampton Research Micro-Tools set, with some of the larger crystals reaching 0.2 × 0.2 × 0.5 mm.

Crystals were cryoprotected with a final concentration of 25% glycerol followed by flash-cooling in a nitrogen cryo-stream. Phases were obtained by soaking some of the crystals in cryosolution supplemented with 1 M sodium bromide for 30–60 s and immediately flash-cooling them. Diffraction data were collected on the SER-CAT 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. The bromide-soaked crystals were exposed to X-ray radiation (wavelength tuned to the bromide absorption edge at 0.9198 Å). A single-wavelength anomalous difference (SAD) data set was then collected to 1.62 Å resolution. Additionally, a native data set was collected at a wavelength of 1 Å to a resolution of 1.55 Å. Both data sets were indexed and scaled with *iMosflm* (Battye *et al.*, 2011) and *AIMLESS* (Winn *et al.*, 2011), respectively, and resolution limits were determined using the $CC_{1/2}$ criterion (Karplus & Diederichs, 2012). The space group was determined to be $P6_1$ using *POINTLESS* (Evans, 2006, 2011).

With one subunit in the asymmetric unit, the Matthews coefficient (V_M) was calculated to be 3.32 Å³ Da⁻¹, with an estimated solvent content of 63% (Kantardjiev & Rupp, 2003). The Br-atom positions, phasing and initial model were determined using the *AutoSol* module of *PHENIX* (Adams *et al.*, 2010). The model was further refined utilizing the native data set with iterative rounds of *PHENIX* automated refinement and manual refinement using *Coot* (Emsley & Cowtan, 2004). Residues 175–178 were not traced owing to missing electron density. Ramachandran plots revealed 98.3% of residues in favored regions and 1.7% in additionally allowed regions (Adams *et al.*, 2010). All structural figures, including the electrostatic map, were

produced using *PyMOL* (<http://www.pymol.org>). The topology diagram was constructed with *Pro-origami* (Stivala *et al.*, 2011), with additional modifications made with *Inkscape* (<http://inkscape.org/>). Atomic coordinates and structure factors have been deposited in the Protein Data Bank (Berman *et al.*, 2004) with accession code 4o6g. Crystallographic data are summarized in Table 1.

3. Results and discussion

3.1. General features of the Rv3902c structure

Initial phases were determined *via* bromine-soaked crystals utilizing the single-wavelength anomalous dispersion (SAD) method (Dauter *et al.*, 2000). An additional high-resolution native crystal data set was collected to 1.55 Å resolution. The final structure consists of 174 residues and 216 waters (Fig. 1*a*). Electron density was not observed for residues 175–178. The core of Rv3902c consists of two main structural domains. The first domain is composed of two anti-

parallel β -sheets containing β -strands 1–5 and 7–9 (Fig. 1*b*) as well as two β - α - β motifs with 3_{10} -helices (B and C) positioned between β -strands 1 and 2 and between β -strands 2 and 3, and a third 3_{10} helix (E) immediately preceding β -strand 7. The second domain consists of α -helices A, D, F and G and two short antiparallel β -strands 6 and 10. A feature of notable functional interest is the creation of a hydrophobic pocket with an acidic entrance between α -helices D and F that is ~ 7 Å in diameter and ~ 7 Å deep (Fig. 2). The interior surface of this pocket is lined with the side chains of residues Tyr80, Tyr84, Leu145, Tyr148, Arg141 (the aliphatic portion) and Ile156 as well as the main chains of residues Lys79 and Gly83. The carboxylic acid moiety of Glu144 and the hydroxyl of Tyr148 are major acid charge contributors located at the entrance to the pocket. This small pocket is located in the center (palm) of a hand-like binding motif, with the bottom of the palm and thumbs made of the antiparallel β -strands 1–5 and 3_{10} -helices B and C, and with the fingers made up of α -helices D and F (Figs. 1*a* and 2). The surface of Rv3902c is highly charged, with an estimated pI of 4.77 (Gasteiger *et al.*, 2005).

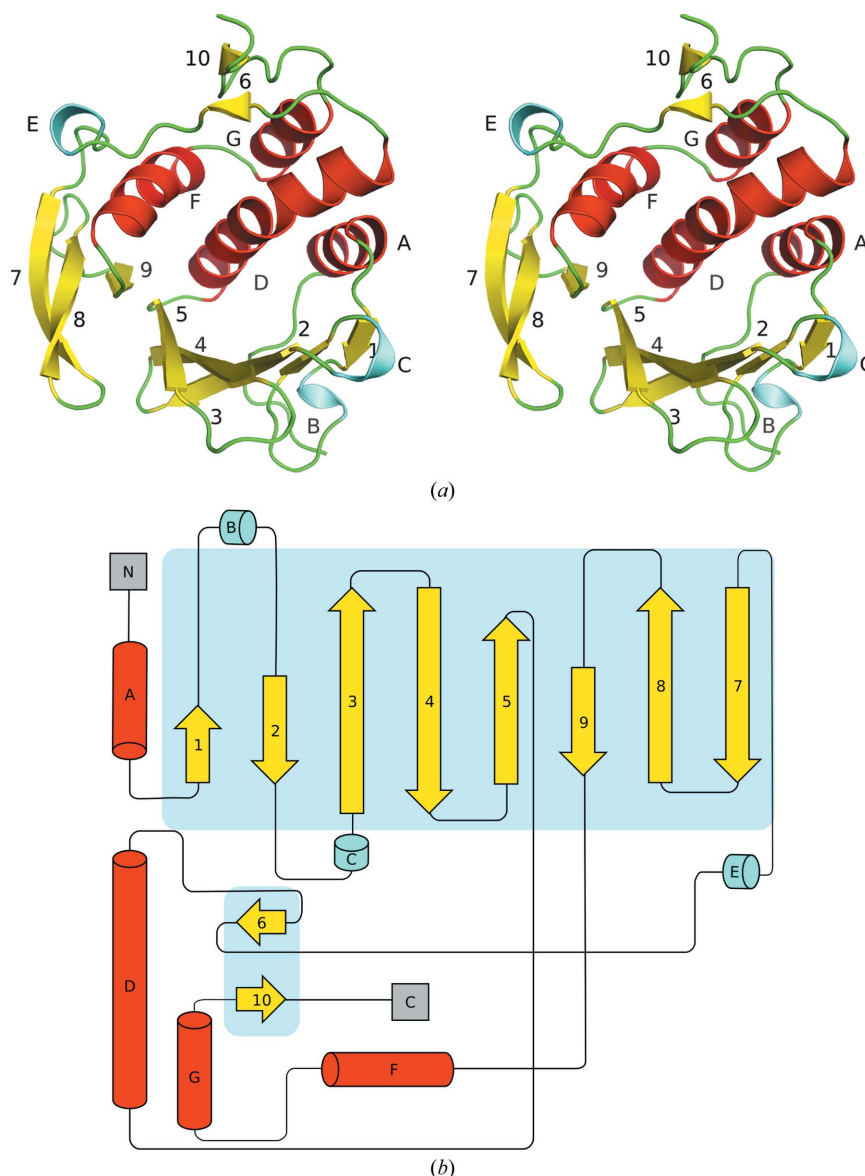


Figure 1

(*a*) Stereoview of the crystal structure of Rv3902c. (*b*) Secondary-structure diagram of Rv3902c. The structure is colored according to secondary-structure elements: loops, green; 3_{10} -helices, cyan; α -helices, red; β -strands, yellow. Each β -strand (1–10) is numbered and each helix is lettered (A–G).

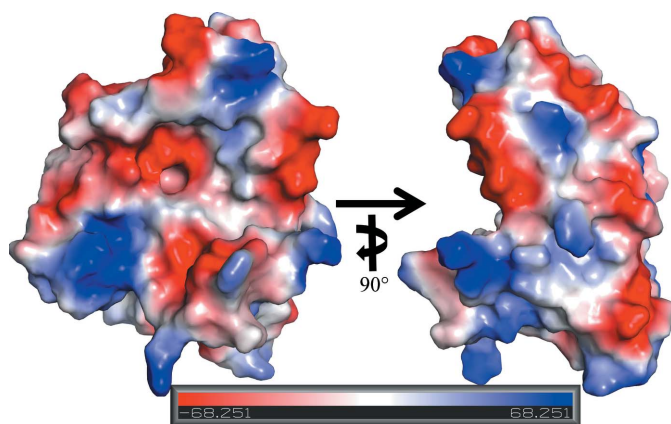


Figure 2
Surface electrostatic potential map of Rv3902c generated by PyMOL, with basic and acidic regions in blue and red, respectively. The views differ by a 90° rotation about the vertical axis.

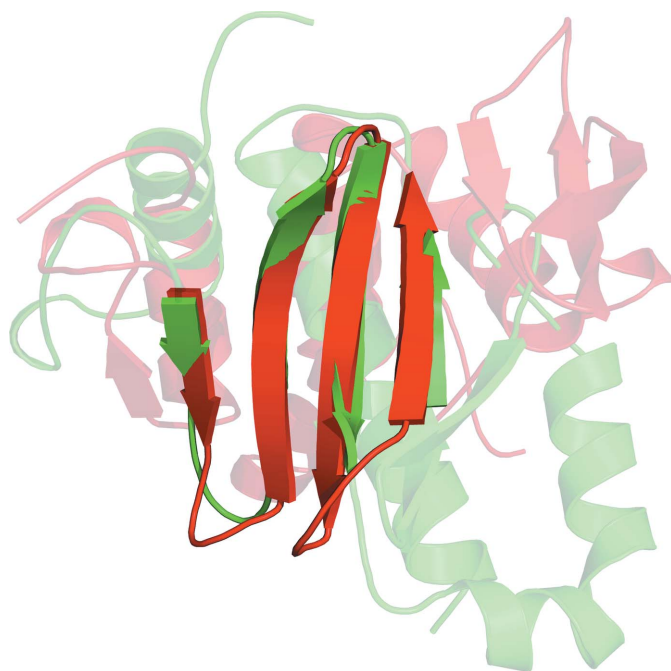


Figure 3
Alignment of Rv3902c (red) with InvB (green). The regions that are not transparent are the aligned homologous domains consisting of antiparallel β -strands 2–5 of Rv3902c.

3.2. Homology search and analysis

Three-dimensional structural homology searches utilizing the *iPBA* web server (Gelly *et al.*, 2011) did not identify structures with reasonable homology to Rv3902c. However, a type III virulence-factor chaperone, InvB, from *Salmonella* exhibited an r.m.s.d. of 2.3 Å along the aligned homologous domains comprising 24.2% of the main-chain structure (60 residues) of Rv3902c. Virulence-factor chaperones vary greatly in structural homology and sequence similarity and are typically small acidic proteins without an ATP-binding site or hydrolytic function that are involved in the secretion and translocation of bacterial virulence proteins (Lilic *et al.*, 2006). Rv3902c has some of the physical characteristics of these chaperones. Fig. 3 aligns the structures of Rv3902c and InvB along the aligned

homologous domains. The aligned domain consists of the antiparallel β -strands 2–5 of Rv3902c, while much of the α -helical domains are not structurally homologous. If Rv3902c is a virulence-factor chaperone, it would represent the first known TB virulence-factor chaperone to be crystallized. Several other virulence-factor chaperones have been crystallized in other species and these virulence-factor chaperones are often in complex with their respective virulence factors (Lilic *et al.*, 2006; Phan *et al.*, 2004). The only other low-homology match to Rv3902c suggested by *iPBA* was a biotin protein ligase with a score of 3.2 Å along aligned regions comprising 10.5% of the structure of Rv3902c (Gelly *et al.*, 2011). An additional homology search was performed using *DALI* (Holm & Rosenström, 2010), but did not yield any clear matches among aligned domains. Finally, a functional search for Rv3902c conducted using the *ProFunc* server (Laskowski *et al.*, 2005) yielded no significant hits. We conclude that it is possible that there is a set of secreted virulence factors including EsxF, EsxE and Rv3903c, and that the Rv3902c protein may be a chaperone involved in the secretion of these proteins.

This research was supported in part by NASA grant NNJ12GA74G, NIH grant 5P30CA13148-37 and NIH grant P50 GM62410. Data were collected on the Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at <http://www.ser-cat.org/members.html>. We thank the staff of SER-CAT for their help during data collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
 Agarwal, N., Woolwine, S. C., Tyagi, S. & Bishai, W. R. (2007). *Infect. Immun.* **75**, 452–461.
 Batty, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
 Berman, H., Nakamura, H. & Henrick, K. (2004). *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*, edited by M. J. Dunn, L. B. Jorde, P. F. R. Little & S. Subramaniam. Hoboken: John Wiley & Sons. doi:10.1002/047001153X.g408315.
 Chim, N. *et al.* (2011). *Tuberculosis*, **91**, 155–172.
 Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
 Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
 Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
 Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
 Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.
 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
 Gelly, J. C., Joseph, A. P., Srinivasan, N. & de Brevern, A. G. (2011). *Nucleic Acids Res.* **39**, W18–W23.
 Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
 Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
 Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
 Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). *J. Mol. Biol.* **351**, 614–626.
 Lilic, M., Vujanac, M. & Stebbins, C. E. (2006). *Mol. Cell*, **21**, 653–664.
 Phan, J., Tropea, J. E. & Waugh, D. S. (2004). *Acta Cryst.* **D60**, 1591–1599.
 Smith, I. (2003). *Clin. Microbiol. Rev.* **16**, 463–496.
 Stivala, A., Wybrow, M., Wirth, A., Whisstock, J. C. & Stuckey, P. J. (2011). *Bioinformatics*, **27**, 3315–3316.
 World Health Organization (2012). *Global Tuberculosis Report 2012*. Geneva: World Health Organization. http://www.who.int/tb/publications/global_report/gtbr12_main.pdf.
 Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.