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

HisB from *Mycobacterium tuberculosis*: cloning, overexpression in *Mycobacterium smegmatis*, purification, crystallization and preliminary X-ray crystallographic analysis

Mohammad Syed Ahangar,‡ Yogesh Khandokar,‡ Nazia Nasir, Rajan Vyas and Bichitra K. Biswal*

Protein Crystallography Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110 067, India

‡ These authors contributed equally to this work.

Correspondence e-mail: bbiswal@nii.res.in

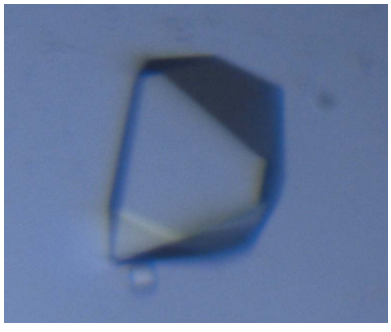
Received 8 June 2011
Accepted 13 September 2011

HisB, encoded by open reading frame Rv1601, possesses enzymatic activity as an imidazoleglycerol-phosphate dehydratase in the histidine-biosynthetic pathway of *Mycobacterium tuberculosis*. A recombinant form of HisB was crystallized in three crystal forms: crystals grown using 20% PEG 1500 as a precipitant belonged to either the cubic space group *P*432 or the tetragonal space group *P*4, while an orthorhombic crystal form belonging to space group $P2_12_12$ was obtained using 15% PEG 5000 and 10 m*M* $MnCl_2$ as precipitant. The structure of HisB in the orthorhombic crystal form was solved by the molecular-replacement method using the crystal structure of its *Arabidopsis thaliana* counterpart, which shares 47% sequence identity with Rv1601, as the search model.

1. Introduction

The histidine-biosynthetic pathway, involving the enzymatic synthesis of histidine from 5-phosphoribosyl-1-pyrophosphate in ten enzymatic reactions carried out by ten enzymes, is conserved among bacteria, lower eukaryotes and plants, but is absent in mammals (Ames *et al.*, 1960; Alifano *et al.*, 1996; Stepansky & Leustek, 2006). Over the past seven decades, extensive genetic and biochemical studies of this pathway have contributed significantly to the understanding of many fundamental mechanisms of biology, particularly the development of the operon theory (Alifano *et al.*, 1996). In the last decade, the three-dimensional structures of enzymes of this pathway determined using X-ray crystallography from various organisms such as *Escherichia coli* (Sivaraman *et al.*, 2001; Barbosa *et al.*, 2002; Rangarajan *et al.*, 2006), *Methanobacterium thermoautotrophicum* (Sivaraman *et al.*, 2005), *Thermotoga maritima* (Schwarzenbacher *et al.*, 2004), *Mycobacterium tuberculosis* (Cho *et al.*, 2003; Javid-Majd *et al.*, 2008; Due *et al.*, 2011), *Filobasidiella neoformans* (Sinha *et al.*, 2004), *Arabidopsis thaliana* (Glynn *et al.*, 2005) and *Saccharomyces cerevisiae* (Quevillon-Cheruel *et al.*, 2006) have led to the elucidation of the molecular mechanism of the formation of various intermediates.

As analysis of the complete genome sequence of *M. tuberculosis* (*Mtb*), the debilitating human pathogen that causes the chronic infectious disease tuberculosis (TB), revealed that the histidine-metabolic pathway is conserved in *Mtb* (Cole *et al.*, 1998) and subsequent transposon-site hybridization studies suggested that this pathway is essential for optimal growth of *Mtb* (Sassetti *et al.*, 2003), it is important to carry out structure–function relationship studies of the enzymes of this pathway from this pathogen that have not yet been structurally characterized. The three-dimensional structures of these enzymes will be central to complete understanding of histidine synthesis in *Mtb* and may also aid in rational design of inhibitors for these enzymes using a structure-based approach. Here, we report the cloning, overexpression, purification, crystallization and preliminary X-ray studies of HisB (Rv1601), which catalyses the conversion of imidazole glycerol phosphate to imidazole acetol phosphate.



2. Experimental methods

2.1. Cloning

The open reading frame corresponding to Rv1601 was amplified by polymerase chain reaction (PCR) with *Mtb* H37Rv genomic DNA as the template, gene-specific primers (Table 1), deoxynucleotides (dNTPs), $MgCl_2$ and Phusion polymerase (Finnzymes, Finland). The reaction was carried out for 30 cycles, with each cycle consisting of denaturation at 371 K for 2 min, annealing at 344 K for 1 min and extension at 345 K for 1 min. The amplified PCR product was purified using a Miniprep kit (Qiagen, Germany) and inserted directionally into the entry vector pENTR following the manufacturer's protocol (Invitrogen, USA). The entry clone was isolated, digested with *Nde*I and *Hin*dIII and purified using a gel-extraction kit (Qiagen). The purified insert was ligated into the *Nde*I- and *Hin*dIII-digested *M. smegmatis* (*Msg*)–*E. coli* shuttle expression vector pYUB1062 using T4 DNA ligase. Successful integration of the insert into the expression vector and its directionality were confirmed by DNA sequencing (Ocimum Biosolutions, Hyderabad, India).

The construct containing the following expression sequence [an N-terminal methionine and hexahistidine tag (bold) followed by the Rv1601 specific amino acids] was overexpressed in *Msg*: **MHHHHHH**TTTQTAKASRRARIERRTRESDIVIELDLDGTGQVAVDTGVPFYDHMLTALGSHASFDLTVRATGDVEIEAHHTIEDTAIALGTALGQALGDKRGIRRFGDAFIPMDETLAHAAVDLSGRPYCVHTGEPDHLQHTTIAGSSVPYHTVINRHVFESLAANARIALHVRVLYGRDPHHITEAQYKAVARALRQAVEPDPRVSGVPSTKGAL.

2.2. Overexpression

The Rv1601-pYUB1062 plasmid was transformed into *Msg* strain mc^24517 by electroporation and the cells were plated on Difco Middlebrook 7H10 agar plates supplemented with 10% OADC (oleic acid–albumin–dextrose–catalase), 0.5% Tween 80 and the antibiotics kanamycin (25 µg ml^{-1}) and hygromycin B (100 µg ml^{-1}). Transformed colonies appeared after incubation for 72 h at 310 K. A primary culture was prepared by inoculating a single colony in 10 ml Difco Middlebrook 7H9 broth supplemented with 0.2% Tween 80, 0.05% glycerol, 10% ADC and the antibiotics kanamycin (25 µg ml^{-1}) and hygromycin B (100 µg ml^{-1}); the culture was grown at 310 K and 200 rev min^{-1} for 24 h. 1 ml of the primary culture was inoculated into 50 ml of the same medium and was grown at 310 K and 200 rev min^{-1} until the optical density at 600 nm ($OD_{600\,nm}$) reached 0.6–0.8. Subsequently, the culture was diluted 1:32 into fresh Middlebrook 7H9 broth and grown to mid-exponential phase (an

Table 1
Primers used for PCR amplification.

The forward primer is comprised of complementary nucleotides for directional cloning into the entry vector pENTR (first four nucleotides), an *Nde*I restriction site (underlined), the DNA sequence for a six-histidine tag (italics) to facilitate purification of the recombinant protein using nickel–nitrilotriacetic acid (Ni–NTA) metal-affinity chromatography and the gene-specific nucleotides (bold). The reverse primer was designed with a spacer (first three nucleotides) for efficient cleavage of the *Hin*dIII restriction site (underlined) followed by the gene-specific nucleotides (bold).

Forward primer	5′-CACCCATATG*CACCATCATCATCATCAT***ACAACCACCCAGACAGCCAA**-3′
Reverse primer	5′-TATAAGCTT**TCACAGAGCACCTTTGGTG**-3′

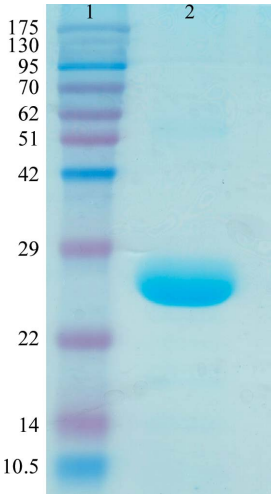
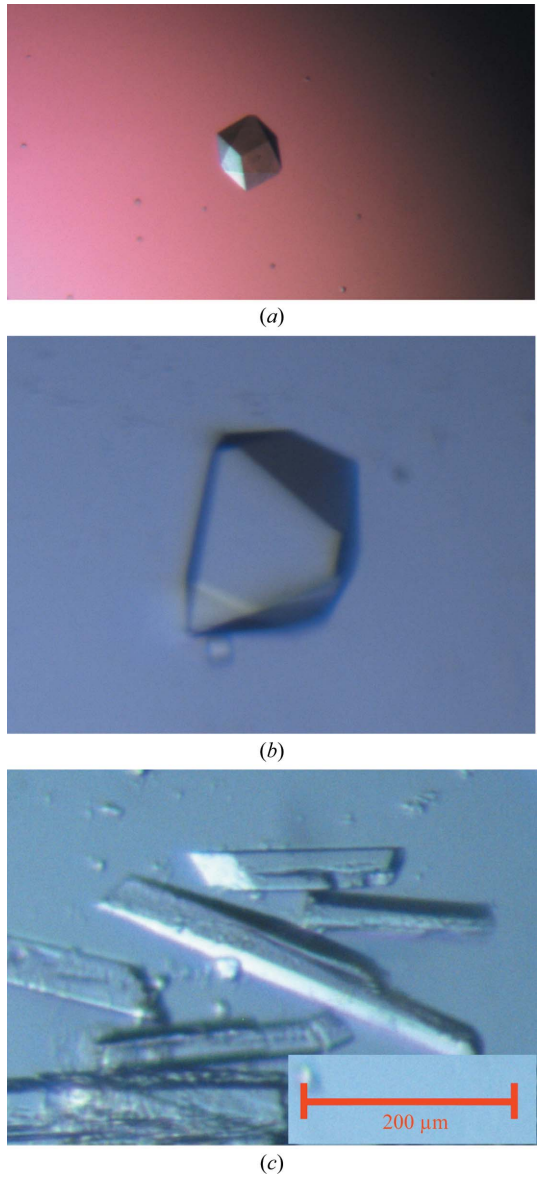


Figure 1
Electrophoretic profile of Rv1601. The protein was resolved by SDS–PAGE and stained using Coomassie Brilliant Blue R-250. Lane 1, molecular-mass marker (labelled in kDa); lane 2, purified Rv1601.



(*a*) (*b*) (*c*)

Figure 2
Morphology of Rv1601 crystals. The approximate dimensions of the cubic (*a*), tetragonal (*b*) and orthorhombic (*c*) crystals were 50 × 50 × 50, 200 × 150 × 150 and 300 × 50 × 50 µm, respectively.

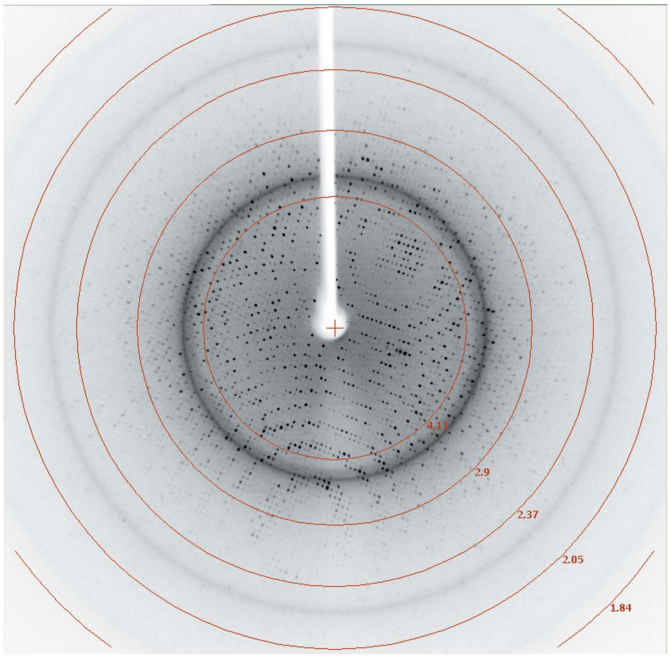
$OD_{600\,nm}$ of ∼0.7) at 310 K and 200 rev min^{-1} and then induced with 0.2% acetamide. After 12 h of induction, the cells were harvested by centrifugation at 10 000*g* for 25 min. The cell pellet was washed three times with 50 m*M* Tris pH 8.5, 200 m*M* NaCl and 5% glycerol to remove the supplementary proteins in the growth medium and was resuspended in 50 ml of the same buffer with one Complete Mini EDTA-free protease-inhibitor tablet (Roche, Germany).

2.3. Purification

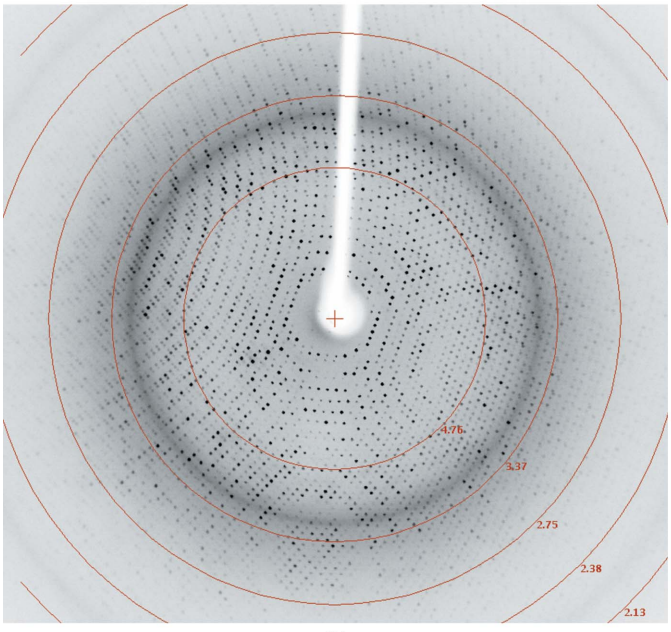
The cells were lysed at 277 K at high pressure (138 MPa) using a cell disrupter (Constant Systems Ltd, UK). The lysate was centrifuged at 10 000*g* for 45 min at 277 K to remove unbroken cells and inclusion bodies. The supernatant was then loaded onto an Ni–NTA affinity column and the column was washed with 50 m*M* Tris pH 8.5, 200 m*M* NaCl, 5% glycerol and 10 m*M* imidazole to remove non-specifically bound proteins. Subsequently, HisB was eluted from the column using 300 m*M* imidazole. The eluted protein was further purified by gel-filtration chromatography using a HiLoad 16/60 Superdex 200 pg column (GE Healthcare, USA). The purity of the Rv1601 was examined on 15% SDS–PAGE (Fig. 1). The identity of the HisB was confirmed by mass-spectrometric analysis of the gel band (Technoconcept, India).

2.4. Crystallization and data collection

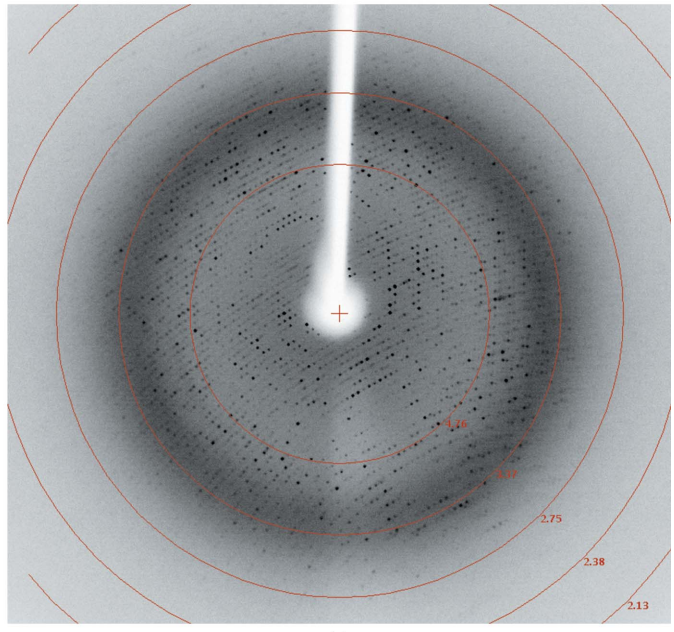
Crystallization screening experiments with a 1:1 ratio of protein (18 mg ml^{-1} in 20 m*M* Tris–HCl pH 8.5 and 50 m*M* NaCl) to precipitant were set up at 293 K by the hanging-drop vapour-diffusion technique in a 24-well plate using Crystal Screen and Crystal Screen 2 from Hampton Research. The well in each case contained 1 ml screen solution. A large number of small crystals appeared in Crystal Screen condition No. 13 [30%(*v*/*v*) polyethylene glycol (PEG) 400, 0.2 *M* sodium citrate tribasic dehydrate, 0.1 *M* Tris–HCl pH 8.5] after 4 d. This condition was optimized by varying the protein concentration and the molecular weight of the PEG and by varying the pH range of the screen buffer from 8 to 9 in steps of 0.25. To understand the effects (if any) of metal ions, particularly manganese, on crystal growth, 10 m*M* $MnCl_2$ was added to each drop in a parallel crystallization setup. $MnCl_2$ was chosen as an additive because the crystal structure of HisB from *A. thaliana* shows manganese binding in the active site (PDB entry 2f1d; Glynn *et al.*, 2005). Diffraction-quality crystals belonging to either a cubic or a tetragonal space group were grown from a condition consisting of 14 mg ml^{-1} Rv1601, 20% PEG 1500, 0.2 *M* sodium citrate tribasic dehydrate and 0.1 *M* Tris–HCl pH 8.5, whereas drops containing the precipitant 15% PEG 5000, 0.2 *M* sodium citrate tribasic dehydrate, 0.1 *M* Tris–HCl pH 8.5 and 10 m*M* $MnCl_2$ produced needle-shaped orthorhombic crystals.



(*a*)



(*b*)



(*c*)

Figure 3
Representative diffraction images of (*a*) cubic, (*b*) tetragonal and (*c*) orthorhombic crystals. The resolution rings are shown by the red concentric circles.

Table 2
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

	Cubic	Tetragonal	Orthorhombic
Space group	*P*432	*P*4	$P2_12_12$
Unit-cell parameters (Å, °)	$a = b = c = 112.54$, $\alpha = \beta = \gamma = 90$	$a = b = 123.76$, $c = 183.39$, $\alpha = \beta = \gamma = 90$	$a = 148.14$, $b = 171.17$, $c = 112.83$, $\alpha = \beta = \gamma = 90$
Matthews coefficient (Å^3 Da^{-1})	2.52	2.48	2.53
Solvent content (%)	51.1	50.5	51.3
Data-collection temperature (K)	100	100	100
Detector	R-AXIS IV^{++}	R-AXIS IV^{++}	R-AXIS IV^{++}
Wavelength (Å)	1.5418	1.5418	1.5418
Resolution (Å)	50.0–2.05 (2.12–2.05)	50.0–2.50 (2.59–2.50)	50.0–2.70 (2.80–2.70)
Unique reflections	15502 (1527)	90879 (9373)	70695 (6947)
Multiplicity	7.9 (8.6)	3.6 (3.6)	4.2 (4.1)
$\langle I/\sigma(I)\rangle$	37.2 (4.8)	10.3 (5.6)	7.6 (2.3)
Completeness (%)	97.2 (98.1)	95.9 (99.4)	89.0 (89.1)
$R_{\rm merge}$† (%)	10.5 (59.0)	15.6 (27.6)	17.0 (57.2)

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

For X-ray data collection, crystals were mounted on CryoLoops (Hampton Research), rinsed with cryoprotectant solution [30%(*v*/*v*) glycerol in reservoir solution] and flash-cooled directly in a nitrogen stream at 100 K. Diffraction data were collected in-house with Cu *Kα* radiation using the in-house X-ray diffraction facility: a Rigaku FR-E+ SuperBright microfocus rotating-anode dual-wavelength (Cu and Cr) X-ray generator equipped with R-AXIS IV^{++} detectors. The data sets were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Data statistics are summarized in Table 2.

3. Results and discussion

HisB was purified to homogeneity by Ni–NTA affinity and gel-filtration chromatography (Fig. 1). Attempts to produce the milligram quantities of soluble recombinant Rv1601 required for structural studies using pDEST17 vector expressed in an *E. coli* BL21 (DE3) expression host were unsuccessful; however, soluble Rv1601 was obtained using an *Msg* expression system with a yield of 35 mg per litre of culture. This finding corroborates previously reported data showing that the expression of *Mtb* proteins in *Msg* resulted in improved solubility (Goldstone *et al.*, 2008). HisB was crystallized in cubic, tetragonal and orthorhombic crystal forms (Fig. 2) and the crystals diffracted to 2.05, 2.50 and 2.70 Å resolution, respectively (Fig. 3). Assuming the presence of one molecule (with a calculated molecular weight of 23 592 Da) per crystal asymmetric unit for the cubic form and 12 molecules per crystal asymmetric unit for both the tetragonal and orthorhombic forms, the calculated Matthews coefficients were 2.52, 2.48 and 2.53 Å^3 Da^{-1}, respectively, corresponding to 51.1, 50.5 and 51.3% solvent content (Matthews, 1968).

The structure of HisB in the orthorhombic crystal form was solved by the molecular-replacement method using the crystal structure of its *A. thaliana* counterpart (PDB entry 2f1d; Glynn *et al.*, 2005), which shares 47% sequence identity with Rv1601, as a search model (Fig. 4). The program *Phaser* from the *CCP*4 suite (Winn *et al.*, 2011) was used

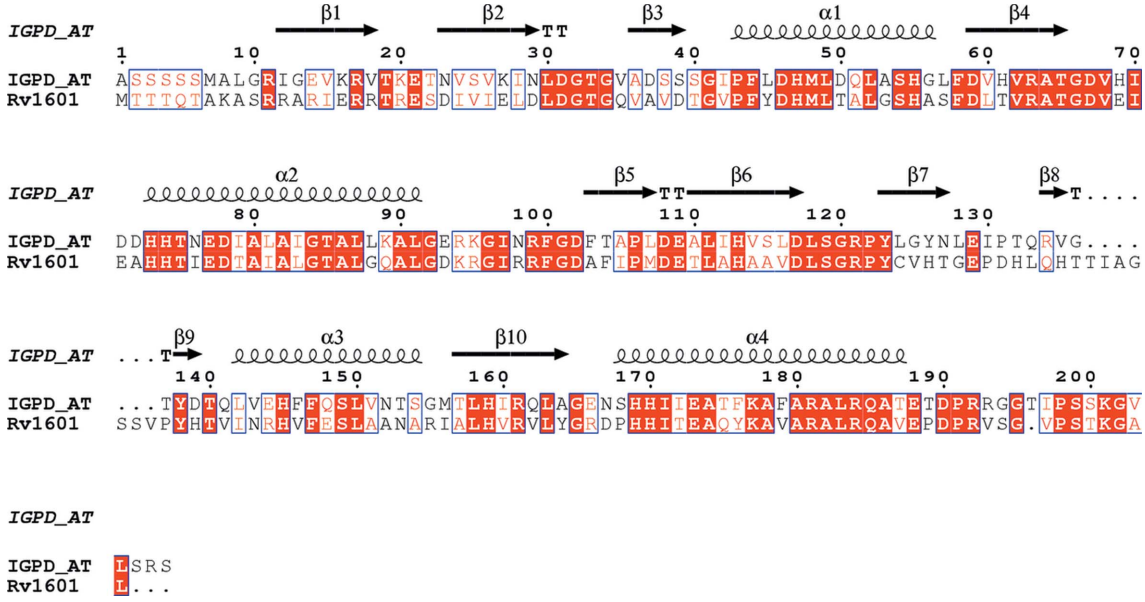


Figure 4
Sequence alignment of *Mtb* HisB and its counterpart from *A. thaliana* (IGPD_AT). The secondary-structural elements corresponding to IGPD_AT are drawn at the top. Helices, *β*-strands and turns are represented by coils, arrows and the letter T, respectively. Conserved residues are highlighted in red boxes. The sequence alignment was produced using the program *ClustalW* (Thompson *et al.*, 1994) and the figure was generated using the program *ESPript* (Gouet *et al.*, 1999).

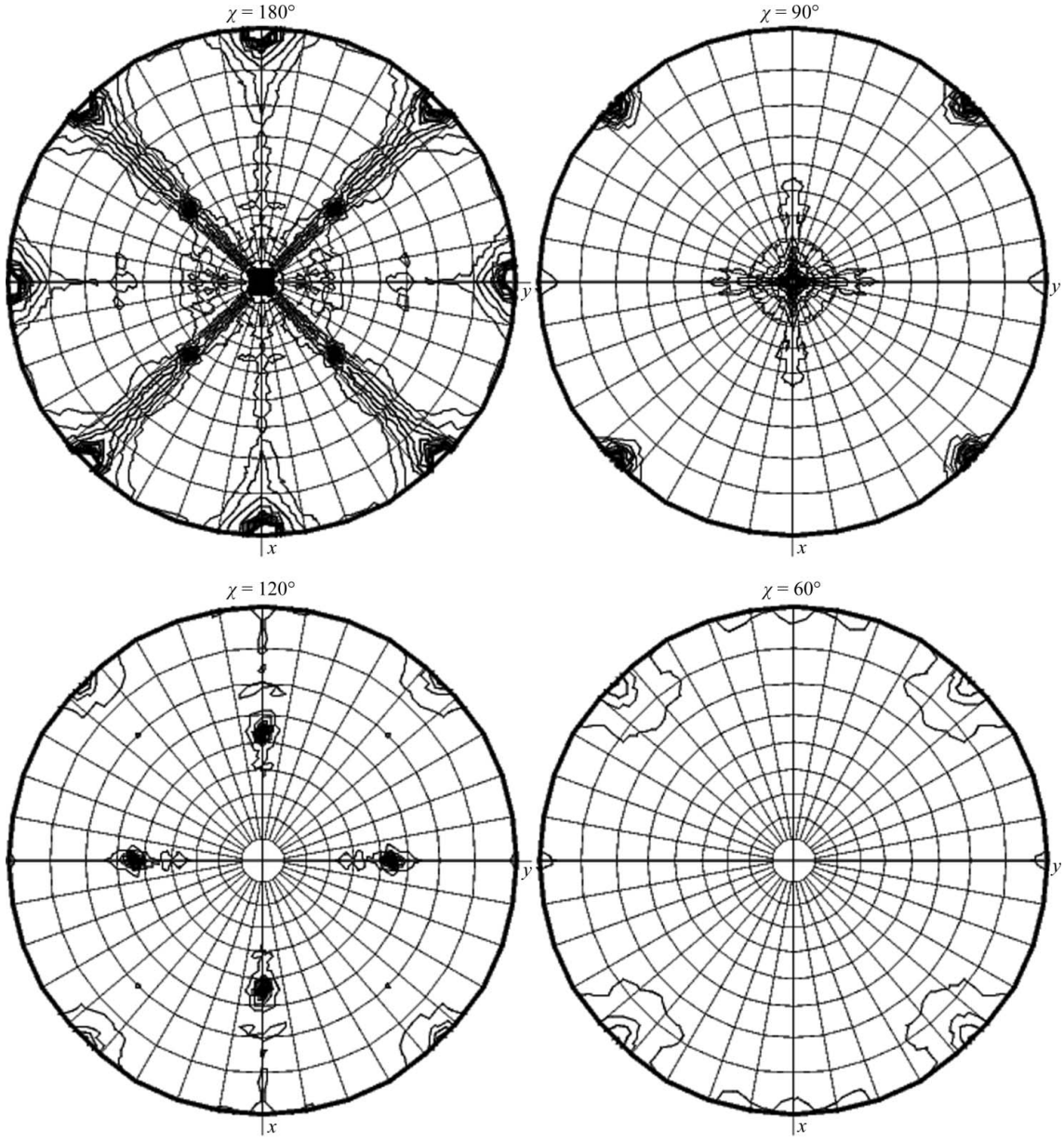


Figure 5
Stereographic projections of various χ sections of the self-rotation function calculated using data for the $P2_12_12$ crystal.

to solve the structure and yielded a model (solution) comprised of 12 molecules. Analysis of the molecular-replacement solution revealed that the 12 molecules are arranged symmetrically in space, forming a hemispherical dome, and that this molecular assembly possesses threefold and fourfold noncrystallographic symmetry (NCS). Self-rotation function computations using the program *MOLREP* from the *CCP*4 suite confirmed the presence of NCS threefold ($\chi = 120^\circ$ section) and fourfold ($\chi = 90^\circ$ section) axes (Fig. 5). The values of R_{work} and R_{free} after rigid-body refinement were 45.6% and 49.0%, respectively. Refinement of the model is in progress. Efforts to solve the structure of Rv1601 in the cubic and tetragonal space groups are being made.

We are grateful to Professor William R. Jacobs of the Department of Microbiology and Immunology and the Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, New York, USA for providing us with the expression vector pYUB1062 and the *M. smegmatis* mc^24517 expression system. *Mtb* H37Rv genomic DNA was obtained from the Biodefense and Emerging Infections Research Resources Repository (BEI Resources, Manassas, Virginia, USA). The work reported in this paper was supported by a start-up grant (to BKB) from the National Institute of Immunology (NII), New Delhi, India. The in-house X-ray diffraction facility was established with financial support from the Department of Biotechnology (DBT), Government of India. Ravikant Pal is acknowledged for his help during data collection. NN is a junior research fellow of the Council of Science and Industrial Research, Government of India.

References

Alifano, P., Fani, R., Liò, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M. S. & Bruni, C. B. (1996). *Microbiol. Rev.* **60**, 44–69.
Ames, B. N., Garry, B. & Herzenberg, L. A. (1960). *J. Gen. Microbiol.* **22**, 369–378.

Barbosa, J. A., Sivaraman, J., Li, Y., Larocque, R., Matte, A., Schrag, J. D. & Cygler, M. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1859–1864.
Cho, Y., Sharma, V. & Sacchettini, J. C. (2003). *J. Biol. Chem.* **278**, 8333–8339.
Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
Due, A. V., Kuper, J., Geerlof, A., von Kries, J. P. & Wilmanns, M. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 3554–3559.
Glynn, S. E., Baker, P. J., Sedelnikova, S. E., Davies, C. L., Eadsforth, T. C., Levy, C. W., Rodgers, H. F., Blackburn, G. M., Hawkes, T. R., Viner, R. & Rice, D. W. (2005). *Structure*, **13**, 1809–1817.
Goldstone, R. M., Moreland, N. J., Bashiri, G., Baker, E. N. & Lott, J. S. (2008). *Protein Expr. Purif.* **57**, 81–87.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Javid-Majd, F., Yang, D., Ioerger, T. R. & Sacchettini, J. C. (2008). *Acta Cryst.* D**64**, 627–635.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Quevillon-Cheruel, S., Leulliot, N., Graille, M., Blondeau, K., Janin, J. & van Tilbeurgh, H. (2006). *Protein Sci.* **15**, 1516–1521.
Rangarajan, E. S., Proteau, A., Wagner, J., Hung, M.-N., Matte, A. & Cygler, M. (2006). *J. Biol. Chem.* **281**, 37930–37941.
Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2003). *Mol. Microbiol.* **48**, 77–84.
Schwarzenbacher, R. *et al.* (2004). *Proteins*, **54**, 801–805.
Sinha, S. C., Chaudhuri, B. N., Burgner, J. W., Yakovleva, G., Davisson, V. J. & Smith, J. L. (2004). *J. Biol. Chem.* **279**, 15491–15498.
Sivaraman, J., Li, Y., Larocque, R., Schrag, J. D., Cygler, M. & Matte, A. (2001). *J. Mol. Biol.* **311**, 761–776.
Sivaraman, J., Myers, R. S., Boju, L., Sulea, T., Cygler, M., Davisson, V. J. & Schrag, J. D. (2005). *Biochemistry*, **44**, 10071–10080.
Stepansky, A. & Leustek, T. (2006). *Amino Acids*, **30**, 127–142.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.