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

Jiang Yin, Grace Garen, Craig Garen and Michael N. G. James*

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

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

Received 8 April 2011
Accepted 1 June 2011

Expression, purification and preliminary crystallographic analysis of Rv3002c, the regulatory subunit of acetolactate synthase (IlvH) from *Mycobacterium tuberculosis*

Branched amino-acid biosynthesis is important to bacterial pathogens such as *Mycobacterium tuberculosis* (*Mtb*), a microorganism that presently causes more deaths in humans than any other prokaryotic pathogen (http://www.who.int/tb). In this study, the molecular cloning, expression, purification, crystallization and preliminary crystallographic analysis of recombinant IlvH, the small regulatory subunit of acetohydroxylic acid synthase (AHAS) in *Mtb*, are reported. AHAS carries out the first common reaction in the biosynthesis of valine, leucine and isoleucine. AHAS is an essential enzyme in *Mtb* and its inactivation leads to a lethal phenotype [Sassetti *et al.* (2001), *Proc. Natl Acad. Sci. USA*, **98**, 12712–12717]. Thus, inhibitors of AHAS could potentially be developed into novel anti-*Mtb* therapies.

1. Introduction

The biosynthetic pathway for the production of branched amino acids, *i.e.* valine, leucine and isoleucine, is unique to bacteria, fungi, algae and plants. This makes the key enzymes of this pathway very attractive targets for the design and development of herbicides, fungicides and antibiotics. One such enzyme, acetohydroxylate synthase (AHAS), catalyzes the first step in the biosynthesis of all branched amino acids (Friden *et al.*, 1985; Jackson & Henderson, 1975).

AHAS holoenzymes are assembled from large catalytic subunits and small regulatory subunits. In many enterobacteria there are three AHAS enzymes: AHAS I, II and III. The three AHAS isozymes in *Escherichia coli* are named IlvB, IlvG and IlvI, respectively, for the large subunits, and IlvN, IlvM and IlvH, respectively, for the corresponding small subunits. Although such genetic redundancy/variety is not always observed in other bacteria, AHAS III homologues exist ubiquitously among bacteria, especially in genomes encoding a single acetohydroxylate synthase (Fig. 1).

In the absence of the small subunits, the large catalytic subunits of AHAS enzymes exhibit only partial activity *in vitro* and are insensitive to product feedback inhibition (Vyazmensky *et al.*, 2009). In the presence of their cognate small subunits, however, the holoenzymes gain full catalytic activity that is negatively regulated by the product valine (IlvB and IlvI). Interestingly, AHAS II, *i.e.* the complex of IlvG and IlvM, is resistant to inhibition by valine.

The crystal structure of *E. coli* IlvH shows two domains, each with a canonical $\beta\alpha\beta\beta\alpha\beta$ ferrodoxin fold, connected by a short linker (Kaplun *et al.*, 2006). The N-terminal domain is arranged at almost a right angle relative to the C-terminal domain, thereby creating an overall L-shaped structure. A higher-ordered structure of IlvH is also evident in the crystal: there are two IlvH homodimers in the asymmetric unit. In the IlvH dimer, the N-terminal domains of two IlvH molecules are juxtaposed with each other in an antiparallel fashion, with their β-sheets forming a contiguous and larger sheet. The two C-terminal domains are packed back-to-back against each other and are perpendicular to the extended β-sheet of the two N-terminal domains. Several lines of evidence lend support to the biological significance of the dimeric assembly of IlvH. Firstly, the dimerization interface is very large (4322 Å^2) and involves many highly conserved residues. Secondly, the IlvH dimer is structurally close to other

members of the ACT-domain family, *e.g.* 3-phosphoglycerate dehydrogenase (3PGDH), for which the dimerization of functional domains is well known (Schuller *et al.*, 1995; Dey *et al.*, 2005). Thirdly, the valine-binding site has been predicted to lie between the two N-terminal domains in the IlvH dimer based on the interactions observed between 3PGDH and its feedback inhibitor L-serine (Dey *et al.*, 2008). Analysis of IlvH variants resulting from mutagenesis of the residues in the predicted valine-binding site seems to support this hypothesis (Jackson & Henderson, 1975).

In the present study, we report the molecular cloning, protein purification and preliminary X-ray crystallographic analysis of Rv3002c from *Mycobacterium tuberculosis* (*Mtb*). The open reading frame of Rv3002c encodes the small regulatory subunit (IlvH) of *Mtb* AHAS. There are presently two large subunits (Rv3003c and Rv3470c) and only one small regulatory subunit (Rv3002c) annotated in the sequenced *Mtb* genome (TB Structural Genomics Consortium; http://www.webtb.org/). Based on the identical orientation and close proximity of the Rv3002c and Rv3003c coding sequences in the *Mtb* genome, as well as their similarity to *E. coli* AHAS III, it is likely that Rv3002c is the small subunit that regulates the catalytic activity of Rv3003c. On the other hand, no known or predicted small regulatory subunit seems to exist for Rv3470. Interestingly, it has been shown that IlvM and some C-terminally truncated forms of IlvH are capable of cross-activating noncognate *E. coli* isozymes *in vitro*. However, these heterologously assembled AHAS holoenzymes are not sensitive to product feedback inhibition (Vyazmensky *et al.*, 2009). It would be interesting to find out whether *Mtb* IlvH can activate both Rv3470c and Rv3003c and/or confer product feedback regulation on their catalysis. Because the branched amino-acid biosynthetic pathway is absent in humans, the key enzymes of this pathway are potential targets for the design of novel antituberculosis therapies.

2. Experimental methods

2.1. Cloning, expression and purification of recombinant *Mtb* Rv3002c

The gene for Rv3002c was amplified by polymerase chain reaction (PCR) using a bacterial artificial chromosome (BAC) genomic library generously provided by L'Institut Pasteur (Brosch *et al.*, 1998; Gordon *et al.*, 1999). The sequences of the PCR primers (Integrated DNA Technologies) used in directional cloning of the amplified product into the Gateway cloning system (Invitrogen) were 3002f, ggg aca agt ttg tac aaa aaa gca ggc tcg ggc agc **CTG GTG CCG CGC GGC** ATG AGC CCG AAG ACG CAC ACG (bold letters, thrombin protease cleavage site; upper case, Rv3002c coding region), and 3002r, gg gac cac ttt gta caa gaa agc tgg gtc ctc gag CTA CTT GGC GGT GCC GAT GCC.

The full open reading frame encoding residues 1–168 of *Mtb* Rv3002c was gel-purified and inserted into the pDONR221 plasmid (Invitrogen) to generate an entry clone plasmid pDONR3002c. Subsequently, the coding region of Rv3002c in pDONR3002c was cloned into an expression plasmid pVP-16 (Novagen) with an amino-terminal hexahistidine maltose-binding protein (hMBP) fusion partner. The coding sequence of Rv3002c in this expression plasmid (pMBP-3002c) was verified by DNA-sequencing analysis (DNA Core Facility, Department of Biochemistry, University of Alberta, Canada).

Recombinant *Mtb* Rv3002c (IlvH) protein expression and purification were carried out using procedures previously described for other *Mtb* proteins apart from a few minor modifications (Yin *et al.*, 2008). Briefly, clarified bacterial cell lysate containing recombinant IlvH protein was first passed through a 5 ml MBPTrap column (GE Healthcare). Recombinant hMBP-Rv3002c fusion protein was eluted using a buffer consisting of 50 m*M* Tris–HCl pH 8.5, 1 m*M* DTT, 20 m*M* maltose and 0.02% NaN_3. Recombinant thrombin protease (GE Healthcare) was used to proteolytically release the hMBP tag from the hMBP-Rv3002c fusion protein. After overnight dialysis to remove DTT, the digested protein mixture was loaded onto a HisTrap column (GE Healthcare). The flowthrough fractions that contained liberated Rv3002c were concentrated and exchanged into buffer *C* (10 m*M* Tris–HCl pH 7.5, 50 m*M* NaCl, 50 m*M* KCl, 5 m*M* $MgCl_2$, 5 m*M* β-mercaptoethanol, 0.5 m*M* DTT and 0.02% NaN_3) before being loaded onto a size-exclusion column (Sephadex G-75). The eluted fractions were sampled for the presence of purified Rv3002c and the peak fractions were pooled and concentrated to $\sim$5 mg ml^{-1} using an Amicon Ultra filtration unit (5 kDa cutoff,

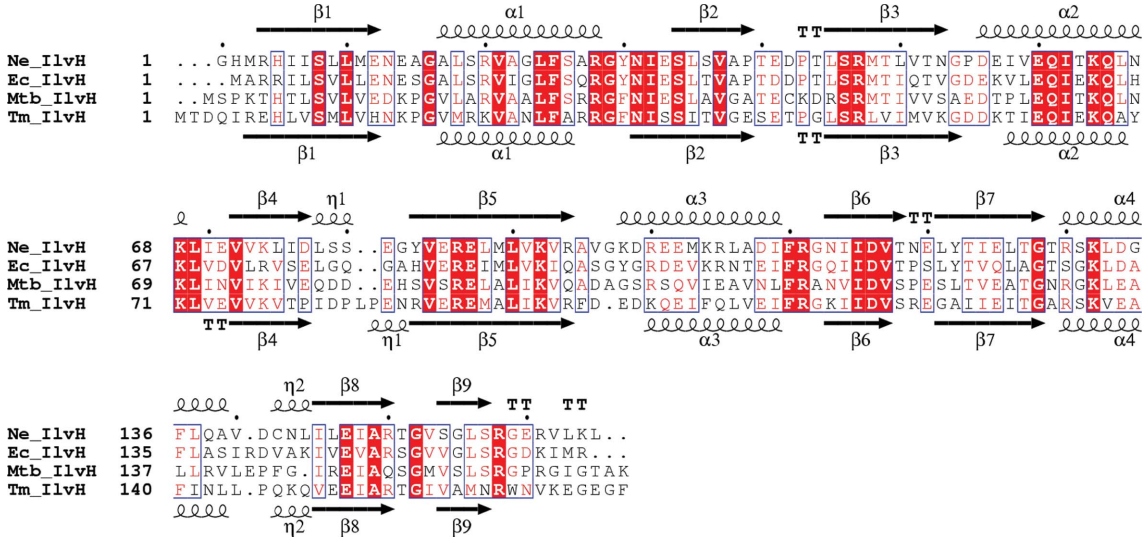

Figure 1
Sequence alignment of *Mtb* IlvH and its homologues from *E. coli*, *Nitrosomonas europaea* and *Thermotoga maritima*. The secondary structures corresponding to *N. europaea* IlvH (PDB entry 2pc6; Petkowski *et al.*, 2007) are drawn at the top of the sequence alignment. Helices, β-strands and turns are represented by coils, arrows and letter 'T's, respectively. Conserved residues are highlighted in red, whereas residues with similar chemical properties are colored red. The 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).

Millipore); its purity (>95%) was confirmed on 15% SDS–PAGE (data not shown).

2.2. Crystallization

Crystallization conditions were robotically (Hydra, Art Robbins Instruments) screened using the sitting-drop method, 96-well Intelli-Plates and commercial sparse-matrix-based solution sets at room temperature (Hampton Research Index HT, Crystal Screen and Crystal Screen 2). The protein solution and reservoir solution were mixed in a 1:1 ratio (0.2 + 0.2 µl). Multiple hits were obtained using the Index HT screen and were further optimized *via* the hanging-drop method. While varying the pH and PEG molecular weight could lead to thicker or larger crystals, crystals with the best diffraction quality were obtained by mixing equal volumes (0.5–1 µl) of *Mtb* IlvH and reservoir solution [0.2–0.35 *M* ammonium acetate, 0.1 *M* Bis-Tris pH 6.0, 20–28%(*w*/*v*) PEG 3350] and equilibrating the mixture over 1 ml reservoir solution. Thin plate-shaped crystals with dimensions of ∼10 × 50 × 200 µm were obtained in 2–4 d (Fig. 2).

2.3. Data collection

Mtb IlvH crystals were first cryoprotected in reservoir solution containing 30% glycerol and were then flash-cooled by directly submerging the looped crystals in liquid nitrogen. Native X-ray diffraction data sets were collected using a Rigaku R-AXIS IV^{++} image-plate detector with a Rigaku rotating-anode X-ray generator and at Stanford Synchrotron Radiation Lightsource (SSRL beamline 9-1). Raw data were indexed, integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

3. Results

We have cloned, expressed, purified and crystallized *Mtb* Rv3002c, a putative small regulatory subunit of acetolactate synthase (ALS) from *M. tuberculosis*. X-ray diffraction data were collected both at the home source and on a synchrotron beamline. The crystals diffracted to 2.9 Å resolution at the synchrotron-radiation source (Fig. 3). Preliminary X-ray crystallographic analysis (Table 1) indicated that the IlvH crystals belonged to space group $P2_12_12$, with unit-cell parameters $a = 78.2$, $b = 86.0$, $c = 116.4$ Å. As the calculated molecular weight of *Mtb* IlvH is 18.2 kDa, we concluded that there are four molecules of IlvH in the asymmetric unit of the crystal lattice. This calculation gives a Matthews coefficient of 2.69 Å^3 Da^{-1} and a corresponding solvent content of 54.24% (Matthews, 1968).

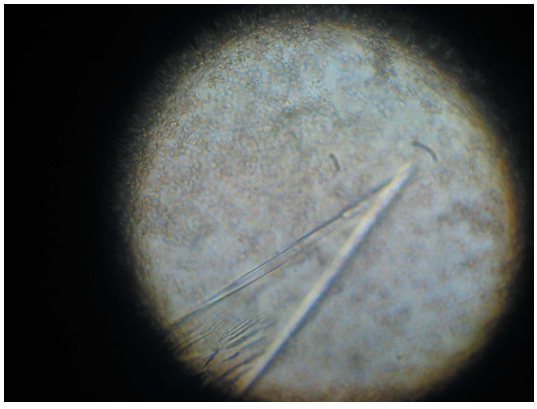

Figure 2
Morphology of Rv3002c crystals. The approximate dimensions of Rv3002c crystals of diffraction quality grown over a period of 3–4 d were 10 × 50 × 200 µm.

Table 1
Crystallographic statistics of the native Rv3002c data set.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12$
Unit-cell parameters (Å)	
a	78.22
b	86.01
c	116.38
No. of molecules in asymmetric unit	4
Matthews coefficient (Å^3 Da^{-1})	2.68
Solvent content (%)	54.09
Data-collection temperature (K)	100
Detector	MAR CCD 325
Wavelength (Å)	0.98
Resolution (Å)	50.0–2.85 (2.95–2.85)
Unique reflections	17906
Multiplicity	6.5 (3.8)
Mosaicity (°)	0.838
$\langle I/\sigma(I)\rangle$	12.7 (1.4)
Completeness (%)	96.2 (74.7)
$R_{\rm merge}$† (%)	13.9 (72.2)

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

IlvH is the small regulatory subunit of the AHAS III enzyme, the full catalytic activity of which requires the formation of a quaternary complex consisting of both large catalytic subunits (∼60 kDa) and small subunits. In enterobacteria there are three isozymes of AHAS, two of which (I and III) are sensitive to product (valine) feedback regulation. The catalytic activity of AHAS II is known to be unaffected by valine. In contrast, there are two annotated large catalytic subunits (Rv3003c and Rv3470c) in the *Mtb* H37Rv genome. However, only one small regulatory subunit (Rv3002c) has been identified *via* sequence analysis. This scenario opens up a number of interesting questions. Does *Mtb* possess a large subunit (Rv3470c) that does not require interaction with small subunits for full activity? If so, is Rv3470c naturally resistant to regulation by valine? Is Rv3002c capable of activating both Rv3003c and Rv3470c? Does Rv3002c

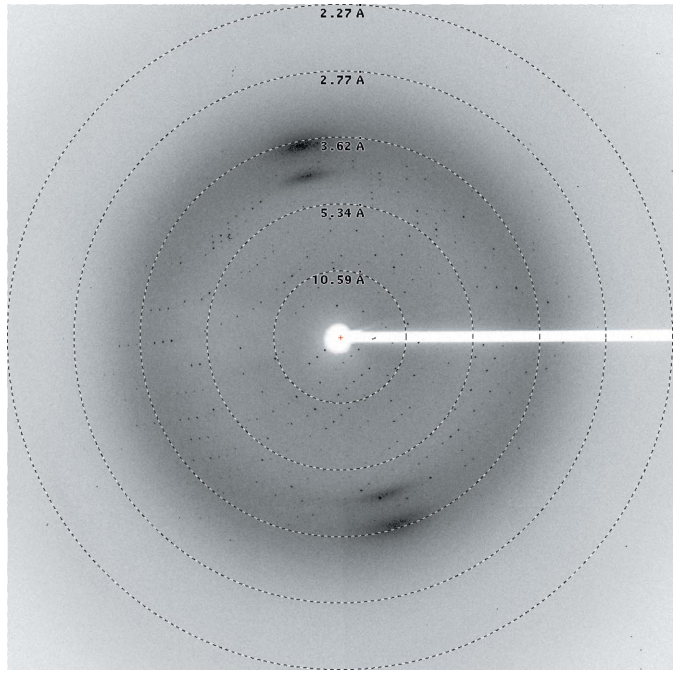

Figure 3
A representative frame of the Rv3002c data set showing the diffraction quality. Rings indicate resolutions.

confer valine sensitivity to both Rv3003c and Rv3470c? While the structure of *Mtb* IlvH will not provide us with the answers to all of these questions, it is an important first step in understanding the biosynthesis of branched amino acids in *Mtb*.

X-ray diffraction data were collected on beamline 9-1 of the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program and the National Institute of General Medical Sciences. Research in the laboratory of MNGJ is funded by the Alberta Heritage Foundation for Medical Research (AHFMR) and by the Canadian Institute of Health Research (CIHR). We thank Sheraz Khan and Maia Cherney for their excellent technical assistance. MNGJ held a Canada Research Chair in Protein Structure and Function during the course of this study. 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.

Dey, S., Burton, R. L., Grant, G. A. & Sacchettini, J. C. (2008). *Biochemistry*, **47**, 8271–8282.

Dey, S., Grant, G. A. & Sacchettini, J. C. (2005). *J. Biol. Chem.* **280**, 14892–14899.

Friden, P., Donegan, J., Mullen, J., Tsui, P., Freundlich, M., Eoyang, L., Weber, R. & Silverman, P. M. (1985). *Nucleic Acids Res.* **13**, 3979–3993.

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. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Jackson, J. H. & Henderson, E. K. (1975). *J. Bacteriol.* **121**, 504–510.

Kaplun, A., Vyazmensky, M., Zherdev, Y., Belenky, I., Slutzker, A., Mendel, S., Barak, Z., Chipman, D. M. & Shaanan, B. (2006). *J. Mol. Biol.* **357**, 951–963.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Petkowski, J. J., Chruszcz, M., Zimmerman, M. D., Zheng, H., Skarina, T., Onopriyenko, O., Cymborowski, M. T., Koclega, K. D., Savchenko, A., Edwards, A. & Minor, W. (2007). *Protein Sci.* **16**, 1360–1367.

Sassetti, C. M., Boyd, D. H. & Rubin, E. J. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12712–12717.

Schuller, D. J., Grant, G. A. & Banaszak, L. J. (1995). *Nature Struct. Biol.* **2**, 69–76.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.

Vyazmensky, M., Zherdev, Y., Slutzker, A., Belenky, I., Kryukov, O., Barak, Z. & Chipman, D. M. (2009). *Biochemistry*, **48**, 8731–8737.

Yin, J., Garen, C. R., Cherney, M. M., Cherney, L. T. & James, M. N. G. (2008). *Acta Cryst.* F**64**, 805–808.