# Structure and function of Rv0130, a conserved hypothetical protein from *Mycobacterium tuberculosis*

PATRIK JOHANSSON, ALINA CASTELL, T. ALWYN JONES, AND KRISTINA BÄCKBRO

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, SE-751 24 Uppsala, Sweden (RECEIVED April 27, 2006; FINAL REVISION July 5, 2006; ACCEPTED July 5, 2006)

# Abstract

A large fraction of the *Mycobacterium tuberculosis* genome codes for proteins of unknown function. We here report the structure of one of these proteins, Rv0130, solved to a resolution of 1.8 Å. The Rv0130 monomer features a single hotdog fold composed of a highly curved  $\beta$ -sheet on top of a long and a short  $\alpha$ -helix. Two monomers in turn pack to form a double-hotdog-folded homodimer, similar to a large group of enzymes that use thiol esters as substrates. Rv0130 was found to contain a highly conserved *R*-specific hydratase motif buried deeply between the two monomers. Our biochemical studies show that the protein is able to hydrate a short *trans*-2-enoyl-coenzyme A moiety with a  $k_{cat}$  of  $1.1 \times 10^2 \text{ sec}^{-1}$ . The importance of the side chains of D40 and H45 for hydratase activity is demonstrated by site-directed mutagenesis. In contrast to many hotdog-folded proteins, a proline residue distorts the central helix of Rv0130. This distortion allows the creation of a long, curved tunnel, similar to the substrate-binding channels of long-chain eukaryotic hydratase 2 enzymes.

Keywords: Rv0130; Mycobacterium tuberculosis; hydratase; hotdog fold; crystal structure

*Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), is responsible for over two million deaths every year and is estimated to infect one-third of the world's population (World Health Organization, www. who.org). Although effective drugs exist, the treatment is long and difficult. The cell wall of *M. tuberculosis* is rich in unusual polysaccharides, lipids, and glycolipids, making it hard to penetrate for drugs and other small molecules (Imaeda et al. 1968; Jarlier and Nikaido 1990). The bacteria can also survive for many years in a dormant state within activated macrophages, and later be reactivated (Dye et al. 1999; O'Regan and Joyce-Brady 2001). This phenomenon has made TB the major cause of death in HIV-infected individuals. In addition,

the recent emergence of multidrug-resistant strains makes the need for new drugs compelling. The availability of the complete genome sequence of *M. tuberculosis* has greatly aided efforts toward understanding the biology of this bacterium (Cole et al. 1998). The ~4000 gene products were reclassified into 11 functional groups in a study by Camus et al. (2002), in which a putative function could be assigned to 52% of the sequences. However, only a small number of these have been experimentally characterized.

Rv0130 is a 16-kDa protein from *M. tuberculosis* H37Rv, listed as a conserved hypothetical protein of unknown function in the TubercuList server of the Pasteur Institute (http://genolist.pasteur.fr/TubercuList/). Castell et al. (2005) recently predicted Rv0130 to share a single hotdog fold (SHD) with a large group of enzymes that use thiol esters as substrates. The hotdog fold encompasses a large, highly curved, antiparallel  $\beta$ -sheet wrapped around a long  $\alpha$ -helix. This fold was originally observed in the *Escherichia coli*  $\beta$ -hydroxydecanoyl thiol ester dehydrase (Leesong et al. 1996) and later in the *Pseudomonas* sp. strain CBS-3 hydroxybenzoyl coenzyme A (CoA) thioesterase (Benning et al. 1998). Many hotdogfolded proteins associate side-by-side to form functional

Reprint requests to: Kristina Bäckbro, Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden; e-mail: nina@xray.bmc.uu.se; fax: +46-18-530396.

Abbreviations: CoA, coenzyme A; DHD, double hotdog; LALI RMSD, root-mean-square distance; SeMet, seleno-methionine; SHD, single hotdog; TB, tuberculosis.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062309306.

homodimers (Leesong et al. 1996; Benning et al. 1998; Hisano et al. 2003). A number of structures with a double hotdog fold (DHD), similar to the quaternary structure of SHD dimers, have also been described: an *E. coli* thioesterase II (Li et al. 2000), a *Candida tropicalis* hydratase 2 (Koski et al. 2004), and a protein of unknown function from *M. tuberculosis* (Castell et al. 2005).

## Results

The structure of Rv0130 was solved by single-wavelength anomalous dispersion (Hendrickson 1991) using selenomethionine-substituted (SeMet) protein. Data collection and refinement statistics are given in Table 1. As previously predicted (Castell et al. 2005), Rv0130 belongs to a family of single hotdog domain proteins (SHDs). The SHD-type structure of Rv0130 is composed of a curved five-stranded antiparallel  $\beta$ -sheet (ordered 1-3-4-5-2) that wraps around the central so called hotdog helix,  $\alpha 4$ (numbering according to Hisano et al. 2003). In addition to the minimal hotdog fold, Rv0130 has a short  $\beta$ -strand and  $\alpha$ -helix at the N terminus,  $\beta 0$  and  $\alpha 0$ , arranged such

**Table 1.** Data collection and refinement statistics

|   | SeMet                  |
|---|------------------------|
| Data collection statistics                    |                        |
| Beamline                                      | ESRF ID14:3            |
| Wavelength (Å)                                | 0.931                  |
| Resolution range (Å) <sup>a</sup>             | 54.4-1.7               |
| Unit cell dimensions (Å)                      | 58.7, 108.7, 50.4      |
| Number of unique reflections                  | 38,906                 |
| Redundancy                                    | 7.3 (6.7) <sup>a</sup> |
| Completeness (%)                              | 98.8 (98.1)            |
| $\langle I/\sigma(I) \rangle$                 | 20.5 (4.0)             |
| R <sub>merge</sub> (%)                        | 7.9 (30.1)             |
| Phasing & Refinement statistics               |                        |
| Phasing resolution (Å)                        | 25-2.0                 |
| Number of sites                               | 4                      |
| Phasing power                                 | 2.17 (0.91)            |
| Anomalous R <sub>Cullis</sub>                 | 0.61 (0.81)            |
| Refinement resolution (Å)                     | 25-1.8                 |
| Number of used reflections                    | 36,072                 |
| R <sub>work</sub> (%)                         | 18.5                   |
| R <sub>free</sub> (%)                         | 22.6                   |
| Number of atoms                               | 2619                   |
| Protein                                       | 2226                   |
| Solvent                                       | 393                    |
| $\langle B \rangle (A^2)$                     | 17.3                   |
| $\langle B \rangle$ Protein (Å <sup>2</sup> ) | 15.6                   |
| $\langle B \rangle$ Solvent (Å <sup>2</sup> ) | 25.8                   |
| Bond RMSD from ideal values <sup>b</sup> (Å)  | 0.02                   |
| Angle RMSD from ideal values <sup>b</sup> (°) | 1.38                   |
| Ramachandran plot outliers (%) <sup>c</sup>   | 0.74                   |

<sup>a</sup> Values in parentheses refer to the outer resolution shell.

<sup>b</sup>Ideal values from Engh and Huber (1991).

 $^{\rm c} {\rm Calculated}$  using a strict-boundary Ramachandran plot (Kleywegt and Jones 1996).

that the strand is aligned parallel to the loop connecting the hotdog helix and  $\beta 2$ . The asymmetric unit contains two Rv0130 molecules packed around a noncrystallographic twofold axis, creating a homodimer with an extended 10-stranded sheet (Fig. 1). The high curvature of the sheet is achieved by  $\beta$ -bulges at residues V18/D22, Y89/N92/P97, and V113 in  $\beta$ 1,  $\beta$ 2, and  $\beta$ 3, respectively. In  $\beta$ 4 and  $\beta$ 5, gradual changes in ( $\phi$ ,  $\psi$ ) angles of successive residues result in a similar deformation (Chothia and Janin 1982). Four of the strands are connected via type I ( $\beta$ 4– $\beta$ 5), type I' ( $\beta$ 3– $\beta$ 4), and type II ( $\beta$ 2– $\beta$ 3) β-turns (Venkatachalam 1968; Laskowski et al. 2005). Rv0130 has a well defined lid (Castell et al. 2005) formed by the loop linking  $\beta 1$  and the N terminus of the hotdog helix. The lid consists of a three-turn  $\alpha$ -helix,  $\alpha$ 1, that helps to form the dimer interface, and a pair of single-turn  $\alpha$ helices,  $\alpha 2$  and  $\alpha 3$ . Each of the two lids of the dimer packs on the N-terminal region of the respective neighboring hotdog helix. A number of electrostatic interactions stabilize these sections; E30-R72 connects  $\alpha 1$  and the hotdog helix, D36-H41 supports the first loop segment of the lid, while D47-R50 stabilizes  $\alpha$ 3. The 4.5-turn hotdog helix is distorted with a distinct 25° kink at its center, P71. This change of direction enables the formation of an additional strand,  $\beta$ 1b, on the concave side of the molecule, creating a two-stranded sheet with the N-terminal strand  $\beta$ 0. Two salt bridges support the sheet; E5-K81 connects the strands together, whereas R2-E16 links this smaller parallel sheet to the large antiparallel sheet.

The two chains of the Rv0130 dimer are very similar and can be superimposed with a root-mean-square distance (RMSD) of 0.45 Å over 149 CA atoms. The subunits have a highly complementary shape, burying a total solvent accessible area of 3300 Å<sup>2</sup>. Three major regions of quaternary interactions govern this interaction:  $\alpha 1$  and  $\alpha 1'$  pack side-by-side and share mainly hydrophobic interactions (the prime indicating an element of the adjacent polypeptide chain). The hotdog helices interact via a hydrophobic patch created by F64/L67/ L70 and F64'/L67'/L70' whereas  $\beta$ 2 and  $\beta$ 2' shares a number of hydrogen bonds on the convex side, as well as some hydrophobic interactions on the concave side of the extended sheet. The bulge on this strand, N92, is at the dimer interface, forcing a local breakdown of the noncrystallographic symmetry at the carbonyl oxygen. For similar reasons, the neighboring side chain of L91 takes on differing conformations in the two subunits of the dimer.

#### Similarity to other structures

The DALI server (Holm and Sander 1994) identified a number of structures similar to Rv0130 (Table 2). The most closely related enzymes are an SHD (*R*)-specific



**Figure 1.** Overall structure and topology. (*A*) The Rv0130 monomer, colored red to blue from N- to C-terminal, forms a homodimer with a neighboring molecule (gold). (*B*) Topology diagram. The secondary structure elements are numbered according to Hisano et al. (2003). The additional  $\beta$ -strand and  $\alpha$ -helix at the N terminus of Rv0130 are called  $\beta$ 0 and  $\alpha$ 0, respectively, and the additional  $\beta$ -strand between the hotdog helix ( $\alpha$ 4) and  $\beta$ 2 is called  $\beta$ 1b.

enoyl-CoA hydratase (PhaJ) (PDB code 1IQ6; Hisano et al. 2003), an SHD protein of unknown function from *Archaeoglobus fulgius* (PDB code 1Q6W; A.A. Fedorov, E.V. Fedorov, R. Thirumuruhan, and S.C. Almo, unpubl.), and Rv0216, a DHD of unknown function from *M. tuberculosis* (PDB code 2BI0; Castell et al. 2005). The structures of two DHD hydratase 2 enzymes (PDB code 1PN2; PDB code 1S9C; Koski et al. 2004, 2005), two SHD dehydratases (FabZ) (PDB code 1U1Z; PDB code 1Z6B; Kimber et al. 2004; Kostrewa et al. 2005), and a SHD dehydratase-isomerase (FabA) (PDB code 1MKA; Leesong et al. 1996) are more distantly related. The extended  $\beta$ -sheet as well as the first part of the hotdog helix of the Rv0130 dimer are both well conserved in all these proteins. However, due to the broken Rv0130 hotdog helix, the last part of the helix and the small, twostranded sheet do not have any counterparts in any of the known structures (Fig. 2A). The N-terminal domains of the two eukaryotic DHD enzymes, though, feature a similar distortion where the hotdog helix is broken into two by a short random coil. The change of direction of Rv0130s hotdog helix also makes it more solvent-exposed compared with some of the other structures. The solvent-exposed residues at the C terminus of the hotdog helix are polar, in contrast to the hotdog helix of the dehydratase-isomerase 1MKA, which only features two polar residues over a

Table 2. Structure alignment, DALI, and DALILite statistics (Holm and Sander 1994)

| PDB  | Name  | Ζ    | RMSD | L.ALI. | %IDE |  |
|------|---|------|------|--------|------|--|
| 1Q6W | A. fulgius Unknown function                               | 16.9 | 2.4  | 142    | 23   |  |
| 1IQ6 | A. caviae (R)-specific enoyl-CoA hydratase                | 16.2 | 2.2  | 129    | 26   |  |
| 2BI0 | M. tuberculosis Rv0216, unknown function                  | 15.1 | 2    | 130    | 13   |  |
| 1PN2 | C. tropicalis 2-enoyl-CoA hydratase 2                     | 12.6 | 2.2  | 112    | 22   |  |
| 1S9C | H. sapiens 2-enoyl-CoA hydratase 2                        | 10.7 | 2.3  | 109    | 12   |  |
| 1U1Z | P. aeruginosa (3R)-hydroxyacyl-ACP dehydratase            | 8.8  | 2.9  | 106    | 22   |  |
| 1MKA | E. coli β-hydroxydecanoyl thiol ester dehydrase-isomerase | 7.8  | 2.8  | 109    | 15   |  |



**Figure 2.** Close-up views of Rv0130. (*A*) A number of amino acid changes, including a proline residue (P71), create a kink in the hotdog helix of Rv0130 (*middle*) and enable the formation of a two-stranded parallel  $\beta$ -sheet, not present in the *A. caviae* SHD *R*-hydratase (*left*), or the *C. tropicalis* DHD hydratase 2 (*right*). (*B*) A comparison of Rv0130 and the *R*-specific enoyl hydratase 1IQ6 (gray) reveals two well conserved active site residues, D40 and H45. The hydrogen bonding network surrounding these is also well conserved, with one exception; a hydrogen bond between the catalytic Asp of 1IQ6 and a residue of the neighboring subunit (S62') is missing in Rv0130. (*C*) The broken hotdog helix creates a long, curved substrate-binding tunnel housing the active site residues D40 and H45.

19-residue stretch. The unusual N-terminal strand and helix of Rv0130 are both absent in the SHDs 1IQ6 and 1Q6W. In 1MKA, 1U1Z, and 1Z6B, however, the N-terminal segment is significantly longer and folds over to create two strands and a short  $\alpha$ -helix. The lid section is conserved in the SHD hydratase as well as in the *M. tuberculosis* Rv0216. In the DHD hydratases, the lid section is only well conserved in the C-terminal SHD domain where it packs on top of the broken hotdog helix of the first domain to form the active site. In the dehydratase and the dehydratase-isomerase, the lid is much reduced and replaced by a short loop that lacks defined secondary structure.

The connections between  $\beta$ -strands in the hotdog foldfamily of structures show a great deal of variation (Castell et al. 2005). However, the distinctive  $\beta 2-\beta 3$  loop, packing under the  $\beta 4-\beta 5$  reverse turn, is structurally conserved in all family members. The loop between  $\beta 3$  and  $\beta 4$  is significantly longer in the N-terminal domain of Rv0216. The same loop is shorter in the C-terminal domain of Rv0216, while it is severely truncated in the C-terminal domains of 1PN2 and 1S9C.

#### Activity measurements

Based on the structural similarity to the enoyl hydratases, a spectroscopic assay (Moskowitz and Merrick 1969; Fukui et al. 1998) was used to investigate hydratase activity. Wild-type protein was found to hydrate crotonoyl-CoA (2-butenoyl-CoA) to hydroxybutyryl-CoA. The enzyme exhibits normal Michaelis-Menten kinetics for the hydratase reaction. For the wild-type enzyme, the  $K_{\rm M}$  value was 55  $\mu$ M,  $V_{\rm max}$  was  $1.6 \times 10^{-7}$  M/sec, and  $k_{\rm cat}$  was  $1.1 \times 10^2$  sec<sup>-1</sup>. The corresponding values for the mutant D40N were 40  $\mu$ M,  $4.4 \times 10^{-8}$  M/sec, and 0.057 sec<sup>-1</sup>, respectively. No kinetic values could be determined for the H45Q mutant. In the reverse reaction, with hydroxybutyryl-CoA as substrate, Rv0130 showed limited activity, but no kinetic values could be determined.

## Active site

The SHD hydratase, the dehydratase, and the dehydrataseisomerase all feature two active sites located on either side of the homodimer. The DHD hydratases, in turn, exhibit only one. Each of these catalytic sites is situated at equivalent positions in a deep cavity or a tunnel, formed by the dimer or domain interface. The residues responsible for the hydration of a trans-2-enoyl-CoA substrate in both bacterial and eukaryotic R-specific hydratases have been identified through site-directed mutagenesis (Hisano et al. 2003; Koski et al. 2004). In the dehydratase-isomerase, parts of the catalytic machinery have been targeted by a mechanism-based suicide inhibitor (Helmkamp and Bloch 1969; Leesong et al. 1996). The hydratase active site consists of a pair of histidyl/aspartyl residues located at a long loop in the lid region. In the dehydratase-isomerase FabA, a structurally equivalent histidine residue, situated in the small lid section, together with an aspartyl residue from the neighboring subunit, has been implicated in the dehydration and isomerization performed by the enzyme. FabZ, catalyzing only the dehydration reaction, features an almost identical active site, but with the aspartyl side chain exchanged for a glutamyl residue (Kimber et al. 2004; Kostrewa et al. 2005).

A superposition of the lid sections of Rv0130, 1IQ6, 106W, 1PN2, and 1S9C reveals a well conserved active site. In Rv0130, the side chains of H45 and D40 show a close fit to the His-Asp pair of the hydratases (Fig. 2B). The hydrogen-bonding network surrounding these residues also exhibits some conservation; the equivalent of H45 forms a short hydrogen bond to a backbone oxygen of an isoleucine (I60) in all hydratase structures. D40 of Rv0130 is positioned by two conserved hydrogen bonds, to a backbone amino group two residues ahead and to Q42 via a water molecule. In 1IQ6 the counterpart of D40 also forms a hydrogen bond to a serine residue (S62') and in 1Q6W to an aspartyl side chain (D80'). 1IQ6 and 1PN2 have an asparagine residue instead of Rv0130 Q42, whereas 1Q6W has a phenylalanine. Comparing Rv0130 with the dehydratases 1U1Z/1Z6B and the dehydratase-isomerase 1MKA, one finds that, in spite of the different location of the active site residues, the catalytically relevant nitrogen and oxygen atoms of the His-Asp and His-Glu pairs can be well superimposed.

Although the putative catalytic residues of Rv0130 and the enoyl-hydratases are very similar, the shapes of the cavities that contain them are different. The Aeromonas caviae 11Q6 features a deep but closed substrate-binding pocket. The depth of this pocket is restricted by the hotdog helix of the adjacent subunit, limiting substrates to C4-C6 enoyl-CoAs (Fukui et al. 1998; Tsuge et al. 2003a). 1Q6W, in contrast, has a long tunnel, going across the loop connecting the hotdog helix with strand β2 and on through the neighboring subunit. However, no functional information has been published about this protein. Due to the distortion of the N-terminal hotdog helix, the two DHD hydratases also exhibit a long curved tunnel passing the active site (Koski et al. 2004, 2005). However, compared with 106W, this cavity is solely created by the DHD-domain interface. The 25° kink in the central helix of Rv0130 results in a similar, long potential substrate-binding tunnel (Fig. 2C). The mouth of the tunnel is created by the splaying apart of the central  $\beta 2$ strands of the extended sheet. Two  $\beta$ -bulges, caused by P97 and by Y89', on either side of the mouth, together with F56 in the lid section and P99 on  $\beta$ 2, form a ring of residues that line the opening of the tunnel. Both the bulges and the two residues are conserved in all SHD and DHD hydratase structures. Other residues around the mouth, such as A98, K138, I87', and R148', are less well conserved. Further into the tunnel I60 and I44, lining one side of the potential catalytic histidine, have leucine/ isoleucine counterparts in all hydratase structures. On the opposite side, G63 is positioned at the N terminus of the hotdog helix. Any other residue at this position would seriously limit the space within the active site cavity. This residue is also conserved in most hotdog structures. H62, stacking against the potential catalytic H45, is retained in all hydratases but 1PN2. The smallest constriction of the tunnel is between G63, H45, I87', and N88', after which it expands to a large opening, created by the neighboring subunit. This cavity is formed as a result of the distorted hotdog helix and the long connection to strand  $\beta 2$ . In the structures lacking the hotdog proline, the length from the beginning of the helix to the beginning of  $\beta 2$  correlates well with the presence of a closed crevice, like in 1IQ6, or of an open-ended tunnel, like in the eukaryotic hydratases (Table 3).

In Rv0130, the low end of the tunnel is deeper and less hydrophobic than the DHD hydratases. A subset of the hydratase family seems to feature a similar proline residue in the hotdog helix as Rv0130. However, neither the sequences of the C-terminal part of the hotdog helix nor the small two-stranded  $\beta$ -sheet are well conserved within this subset. Consequently, the residues lining the

| Table 3.  | Compa    | rison oj | f the c | amino   | acid  | length i | between i      | the          |   |
|-----------|----------|----------|---------|---------|-------|----------|----------------|--------------|---|
| beginning | g of the | hotdog   | helix   | and the | he co | nserved  | $\beta$ -bulge | on $\beta$ ? | 2 |

| Protein | Hotdog<br>helix + loop | Number of residues | Cavity | Proline  |  |  |
|---------|------------------------|--------------------|--------|----------|--|--|
| Rv0130  | G63-N88                | 25                 | Т      | Yes, P71 |  |  |
| 1Q6W    | G68-F95                | 27                 | Т      |          |  |  |
| 1PN2    | P42-H73                | 31                 | Т      |          |  |  |
| 1S9C    | P58-H89                | 31                 | Т      |          |  |  |
| 2BI0    | P70-N91                | 21                 | С      |          |  |  |
| 1IQ6    | G54-Y76                | 22                 | С      |          |  |  |
| Rv0636  | G59-A79                | 20                 | с      |          |  |  |
| Rv3389c | P54-H80                | 26                 | t      |          |  |  |
| Rv3538  | P56-H87                | 31                 | t      |          |  |  |
| Rv2524c | P984-H1013             | 29                 | t      |          |  |  |
| Rv0241c | L61-H90                | 29                 | t      |          |  |  |
| PhaJ1   | G61-I82                | 21                 | с      |          |  |  |
| PhaJ2   | L53-H84                | 31                 | t      |          |  |  |
| PhaJ3   | D60-H94                | 34                 | t      |          |  |  |
| PhaJ4   | G62-N88                | 26                 | t      | Yes, P70 |  |  |

Type of substrate-binding cavity: tunnel (T) or crevice (C), as well as predicted tunnel (t) or crevice (c).

lower part of the tunnel (A86, V147, V127, Q74, Y149, and Y78) are only conserved in a few actino- and proteobacteria that are closely related to *M. tuberculosis*.

The C. tropicalis 1PN2 has been described in complex with its enzymatic product, (3R)-hydroxydecanoyl-CoA (3RHDC) (PDB code 1PN4; Koski et al. 2004). A superposition of Rv0130 and 1PN4 reveals that the pantetheine and acyl chain moieties of 3RHDC can be fitted easily into the putative substrate-binding cleft. An active site water molecule, coordinated by D40, Q42, and H45, occupies the same position as the (3R)-hydroxyl group of the superimposed ligand. This water is equivalent to W6 of 1PN2, which has been proposed to be the source of both the proton and the hydroxyl group added to the CoA-associated acyl thioester substrate of the enzyme (Koski et al. 2004). The ADP-phosphate, bound on the surface of the C. tropicalis enzyme, does not fit Rv0130 without conformational changes. However, the opening of the tunnel contains several side chains that might be involved in interactions with such a substrate. These include the conserved  $\beta$ -bulges (caused by P97 and Y89'), as well as four charged residues (K93, R95, K138, and R148') that might form salt-bridges to phosphate groups of a CoA moiety.

# Substrate specificity of related proteins

The most similar structure to Rv0130, the *R*-specific enoyl hydratase 1IQ6 (PhaJ) from *A. caviae*, has been implicated in the production of short-chain polyhydroxyalkanoates (PHAs) by converting enoyl-CoA intermediates from the fatty acid  $\beta$ -oxidation pathway to (*R*)-3hydroxyacyl-CoAs. These in turn are polymerized by PHA synthase (PhaC) (Fukui and Doi 1997; Fukui et al. 1998). PHAs are polymers of (R)-hydroxyalkanoic acids that many bacteria accumulate under periods of nutritional imbalance. The insoluble PHA polyesters can later serve as a carbon source during starvation (Madison and Huisman 1999; Sudesh et al. 2000). Tsuge et al. (2003b) recently cloned and characterized four R-hydratases from Pseudomonas aeruginosa involved in PHA biosynthesis, PhaJ1-PhaJ4. PhaJ1 was found to be active only on shortchain enoyl-CoAs (C4-C6), whereas PhaJ2, PhaJ3, and PhaJ4 were able to hydrate longer (C8-C12) enoyl-CoA chains. Comparing these enzymes with Rv0130, one finds that PhaJ1 seems to feature a SHD-type structure with a nondistorted hotdog helix followed by a short link to strand  $\beta$ 2, indicating a substrate-binding pocket instead of a tunnel. PhaJ2 and PhaJ3, on the other hand, are DHD-type enzymes with an extended connection between the N-terminal hotdog helix and B2, similar to the eukaryotic hydratase 2 structures (Fig. 3; Table 3). The last protein of the study, PhaJ4, seems to exhibit SHD topology with a proline insertion in the hotdog helix (P70). This might be indicative of a substrate-binding tunnel similar to Rv0130, which in turn would correlate well with the experimental substrate specificity.

# Related M. tuberculosis proteins

Castell et al. (2005) recently predicted the M. tuberculosis genome to contain at least five DHD- and three SHD-type proteins. Several of these are likely to be hydratase/ dehydratase-related enzymes, featuring minor variations of the so called hydratase 2 motif ([YF]-X(1,2)-[LVIG]-[STGC]-G-D-X-N-P-[LIV]-H-X(5)-[AS]) (Qin et al. 2000). In Rv0130, A37 and, in particular, Q42 and W43 deviate from this pattern. Rv0636, predicted by Castell et al. (2005) to be a SHD-folded protein, matches the hydratase 2 motif exactly. This protein is of particular interest since it has been shown to be essential for M. tuberculosis growth both in vitro (Sassetti et al. 2003) and in vivo (Sassetti and Rubin 2003). Rv0636 lacks the proline residue observed in the hotdog helix of Rv0130. The loop between  $\beta 2$  and the hotdog helix is also much reduced compared with both Rv0130 and 1Q6W. We therefore predict a putative Rv0636 homodimer to contain a substrate-binding pocket instead of an open tunnel (Table 3; Fig. 3). Although the active site of Rv0636 is well retained, the mouth of the cavity formed by  $\beta 2$  shows less conservation. The four potential M. tuberculosis DHD proteins that share the hydratase motif (Rv3389c, Rv3538, Rv2524c, and Rv0241c) are more similar to the eukaryotic hydratase 2 enzymes. In these proteins, the N-terminal hotdog helix as well as the following loop seem to be much longer compared with Rv0130 (Table 3).

|   |   | ww  |   |   | VAAA  |   | vv  | V  |   |  |   |  |  |   | I  |   |
|---|---|---|---|---|---|---|---|--|---|--|---|--|--|---|--|---|
| M.tuberculosis Rv0130<br>P.aeruginosa PhaJ4<br>A.caviae 11Q6<br>M.tuberculosis Rv0636<br>P.aeruginosa PhaJ1<br>A turkine 106W               | -MRTFES<br>MPFVI<br>IRVG<br>IRVG<br>MSQVQI                            | 10<br>SVADLAAA<br>PVAALKDY<br>AQS<br>GPYFDDLS<br>LREFSSVA<br>NIPYAELE<br>NPTYFFS1 | 20<br>AGEKVGQSD<br>VGKDLGHSE<br>SLEVGQKARL<br>SKGQVFDWAP<br>VGDQLPEKT<br>SVGQKAEYT- | 30<br>-WVTIT<br>-SKRFG<br>-GVTLS<br>-YP-LT<br>-SSIA<br>-PRTVT               | 40<br>-QEEVNLFADA<br>-QERVDQFAEC<br>-AAEVAAFAAL<br>-LGLAAAHQSJ<br>-RQDLVNYAGV<br>-ERDLQLFAAV<br>-FTDTWTFAYL | 5<br>TGDHQWI<br>TGDHQFI<br>SEDFNPL<br>VGNRLRL<br>VGNRLRL<br>SGDLNPI<br>SGDRNPV<br>TADEEPI | 0<br>HVDPE-<br>HLDPA-<br>ALDSD-<br>HWDDE-<br>HLDAA-<br>HTDVE- | 60<br>KAAK<br>FAAT<br>LCAA<br>IAKV<br>YAAT<br>FAKK | 70<br>GPF-GTTIA<br>TPF-GGTIA<br>TAF-ERPIV<br>/TGMPGPLA<br>/GL-DTAIA<br>TQF-KERIA<br>TTE-GKPTA | 80<br>HGFMTLALL<br>HGFLSLSLI<br>HGMLL-ASL<br>HPGLVCDVA<br>HGMLTMGIG<br>HGMLSGALI<br>OGMLVI STA | 90<br>PR-LQHQM-<br>PKLMEGLL-<br>FS-GLLGQ-<br>IG-QSTLA-<br>GG-YVTSW-<br>SAAIATVL-<br>LG-MVDOV- | 100<br>  | 110  | 120<br>QLPGKGS<br>TQRVK<br>VGDF<br>PGPG                                 | 130<br>YTVKGVKLA<br>VLPEGLKMAY   | 140<br>INYGLNKVREPA<br>VNYGLDTVREIQ<br>IYLGQ-SLSEKL<br>ANLFYRGLREHR<br>GAVTEYNVRETA<br>TIYLGQTLRETR<br>AFEGTKDVRELR |
| H.sapiens 159C<br>M.tuberculosis Rv3589c<br>M.tuberculosis Rv3538<br>P.aeruginosa PhaJ2<br>P.aeruginosa PhaJ3<br>C.tropicalis 1PN2          | MAA-TSO   | GFAG<br>MAIDF<br>MPIDLI<br>-MALDPE\<br>DLPAPPAI                                   | AIGQKLPF<br>PNSIGAVTEF<br>DVALGAQLPF<br>/LLNYPIPEV<br>.PGLFLRAAL                    | FSYAYT<br>MLFEWT<br>VEFSWT<br>-RQRYS<br>-RRGIR                              | -ELEAIMYALC<br>-DRDTLLYAIC<br>-STDVQLYQLC<br>-RRDSAFYALS<br>-GKALPERGLF<br>-DRDVILYNIA                      | VGASIKE<br>VGA-G<br>LGA-G<br>LGL-G<br>SQV-T   | )PKDLK-<br>-TgD<br>-SdPMNP<br>-GdPLD-<br>-VdPKH-<br>-TkQLK-   | FIYE<br>LAFTTE<br>RELSYL<br>ERQL<br>LERY<br>YVYE   | GSSDFSC<br>NSHGIDQQV<br>ADDTPQV<br>AYVDPCR<br>RQVCGFR<br>NDSDFQV                              | LPTFGVIIG<br>LPTFGVICC<br>LPTFGNVAA<br>DLQALPOMA<br>DDGLLPPTY<br>IPTFGHLIT                     | QKSP  | MMGGGLAEJ<br>PAFGA<br>TFHLTTPPI<br>LVLGHPGFW<br>HILAFPLQMAL<br>FNSGKSQNS                   | PGLS-INFAI<br>AAKVGTENPA<br>VQEPGIDIELSI<br>/LGNPATGVDALI<br>.LTDKREPEPLLC<br>GFAKLLRNENPM | (V<br>\L<br>\L<br>SL  |  | LHGEQYLELYK-<br>LHGSQGIRLHA-<br>LHASERVEVPA-<br>VHGEQRLEWR-R<br>VHLENRIDVL-R<br>LHGEHYLKVHSW                        |
| M.tuberculosis Rv2524c<br>M.tuberculosis Rv0241c  | VT(   | SPEFL<br>QPSGLKNI   | PAVANGTAT   | LTVDWH<br>VVPRTDQL  | -PERVADHT-C   | VTA-T<br>IDP-A  | FGE-P-<br>NVA-A-  | LAPSLT<br>YAAVTG                                   | NVPDALV<br>LRYGNQV  | GPCWPAVFA<br>PLTYPFA   |   | AIGSAVTDT<br>LTFPSVMSL   | GEPVVEGLL  | SL<br>SA  | <br>:  | /HLDHAARVVG-<br>IHTENHITQYR-  |
| M.tuberculosis Rv0130<br>P.aeruginosa PhaJ4<br>A.caviae 11Q6  | 1'<br>PVPV-G<br>PVRV-G<br>PVFV-G                                      | 50<br>SRVRAT<br>SRVRLC<br>DEVTAE  | 160<br>TSSLVGVE<br>GLTLLDVN<br>EVEVTALR   | 170<br>DLG<br>EKN   | 180<br>N-GTVQATV<br>P-GQWLIKA<br>E-DKPIATL  | 190<br>STTVE<br>RATLE<br>TTRIF  | EVE-GSA<br>EIE-GQE<br>TQ-G-G                                  | 200<br>KPACVAES<br>KPAYIAET<br>ALAVTGEA            | 210<br>IVRYV<br>LSLCF<br>WKLP   | 220  | 230   | 240  | 250<br>MRTFEST   | 260<br>/ADLAAAA-<br>/AALKDYV-<br>SAQSL-                                 | 270<br>GE-KVGQSD<br>G-KDLGHSE<br>EV-GQKARL                                     | 280<br>-WVTITQEEVNL<br>-WLTIDQERVDQ<br>-SKRFGAAEVAA   |
| M.tuberculosis Rv0216 (2BI0)<br>M.tuberculosis Rv0636<br>P.aeruginosa PhaJ1<br>A.fulgius 1Q6W   | FPAV-G<br>VVPV-PI<br>PVKL-G<br>PVFI-G                                 | DTLYTF<br>NDGKGAEL<br>DDLKVF<br>DTIAAS  | RTEVVGLRAN<br>VFNGRVK<br>ELEVLEKL<br>SAEVVEKQ                                       | ISPKPGR<br>SVD<br>PKN<br>DFD  | A-PTGLAGL<br>P-ESKSVTJ<br>R-VRMATRV<br>E-KSGVVTY  | RMTTI<br>ALTAT<br>FNQAGKO<br>KLEVK  | (DRTD<br>(TG-G<br>(VVDGEA)<br>(NQ-R-G)<br>(NQ-R-G)            | RLVLDFYR<br>-KKIFGRA<br>EIMAPEEK<br>ELVLTALY       | CAMLPAS<br>IASAKLA<br>LSVEL<br>SAL-IRK  | PDWKPGAVP  | GDDLSRIGA   | DAPAPAADPTA  | HWDGAVFRKRV<br>MALI<br>PPISIGMSQVQN<br>TPARN   | PGPHFDAG-<br>REFSSVKV-<br>IPYAELEV-<br>PIYFESIQ-                        | IAGAVLHST<br>GD-QLPEKT<br>G-QKAEYT-<br>IG-EKIEGL                               | -ADLVSGAPE<br>-YP-LTRQDLVN<br>SSIAERDLQL<br>-PRTVTETDIWT  |
| M.tuberculosis Rv3389c<br>M.tuberculosis Rv3538<br>P.aeruginosa PhaJ2<br>P.aeruginosa PhaJ3   | PLPA-AU<br>PLPA-AU<br>PLPA-EU<br>ALGG-LU                              | KLKCE<br>KLSVV<br>SARAV<br>EVIGF<br>SPFTVS  | /TEVADIQ<br>/TRFTDIW<br>/TRVTGLV<br>SVAVENLQ  | DKG<br>DKG<br>DKG<br>PHD  | EGKNAIVVL<br>KAAVICS<br>ADKGALLYS<br>KGATFSI  | RGRGC<br>ETTAT<br>EKVLS<br>VTRLE  | DP-ESG<br>TP-DG<br>DA-LSG<br>DQ-LG-                           | SLVAETLT<br>LLLWTQKR<br>EVLAVARS<br>-LLWVGDS       | TLVLRGQ<br>SIYARGE<br>TTFLR<br>KVLCR  |  |   |  | GGFGGKRG<br>GGFGGKRG<br>GDGGFGGSRQVI<br>-G-VKVPGEIPPI                                      | GERPAAPEF<br>PSGSDVAP-<br>PETPHRLP-<br>CAEQEPLPL                        | PD-RHPDAR<br>ER-APDLQV<br>ER-TPDLRL<br>EP-VDNWKA                               | -IDMPTREDQAL<br>-AMPILPQQALL<br>-DLPTRPEQALY<br>-PADIGRRY   |
| C.tropicalis 1PN2<br>M.tuberculosis Rv2524c<br>M.tuberculosis Rv0241c   | QLPTVP  | GEIKII<br>AQLTVI<br>DAVGVF  | FEPIATI<br>ATAANAT<br>WRAENLR   | PKG<br>DTDMG<br>EHR   |   | IGSKSV<br>VSVVV<br>VTNVS  | /dn-ksg<br>/tg-adg<br>SVGN                                    | ELIYSNEA<br>AVIATLEE<br>DVAWHQVT                   | TYFIRNC<br>RFAILGR<br>TFLHQQR   |  |   | Q/<br>TGSAE  | ADNKVYADRPAF/<br>LADPARAGGAV:<br>TSLSGEPKPPI   | ATNQELAP-<br>SANATDTPR<br>PQKKPKLP-                                     | KR-APDYQV<br>RRRRDVTIT<br>PP-AAVLRI  | -DVPVSEDLAAL<br>-APVDMRPFAV-<br>-T-PAKIRRYAA  |
|   | M   | N   | N   |   | AAAAA   | N   |   |  |   |  |   |  | a final  |   |  |   |
| M.tuberculosis Rv0130<br>P.aeruginosa PhaJ4<br>A.caviae 1IQ6<br>M.tuberculosis Rv0216 (2BI0)<br>M.tuberculosis Rv0636<br>P.aeruginosa PhaJ1 | 290<br>FADATGI<br>FAECTGI<br>FAALSEI<br>LARLTLI<br>YAGVSGI<br>FAAVSGI | HQWIHV<br>DHQFIHV<br>DHQFIHV<br>DFNPLHL<br>NIAATHH<br>DLNPIHW<br>DRNPVHL          | 00<br>PERAAA-G-<br>PEKAAKT<br>PAFAAT-T-<br>WR<br>DEIAKV-V-<br>JAAYAATT              | 310<br>PFGTTIAH<br>PFGGTIAH<br>AFERPIVH<br>VSGRRLVY<br>GLDTAIAH<br>QFKERIAH | 320<br>GFMTLALLPRL<br>GFLSLSLIPKL<br>MLL-ASLFSC<br>GHTIGLALAQ<br>MLTMGIGGGY<br>MLSGALISAA                   | 330<br>QHQ<br>M-EGL<br>LLG<br>ATR<br>VTS<br>I-ATV   | 34<br>LVL<br>QQL<br>L<br>WVG                                  | 0<br>VKGVI<br>PEGLI<br>PG<br>LP<br>LP              | 350<br>KLAINYGLN<br>KMAVNYGLD<br>KGSIYLGQ-<br>NLATVLDWE<br>-DPGAVTEY<br>SPGTIYLGQ             | 360<br>KVRFPAPVP<br>TVRFIQPVR<br>SLSFKLPVF<br>SCDHTAPVH<br>NVRFTAVVP<br>TLRFTRPVK              | 370<br>VGSRVRA<br>VGSRVRL<br>EGDTLYS<br>VPNDGKGAE<br>LGDDLKV                                  | 380<br>TSSLVGVEDLO<br>GLTLLDVNEKM<br>EVEVTALR<br>ELHIESAQAHA<br>LVFNGRVKSVD<br>ELEVLEKLPKM | 390<br>IN-GTVQATV<br>IPGQWLIK<br>E-DKPIATL<br>DGGVLGLRSLV<br>P-ESKSVTI<br>IRVRMATR         | 400<br>-STTVEV<br>-ARATLE<br>-TTRIFT<br>(AVSDSA<br>-ALTATT<br>-VFNQAGKQ | 410<br>EG-SAKPAC<br>IE-GQEKPA<br>QG-GA-LAV<br>SE-PDRQVLI<br>GGKKI<br>VVDGEAEIM | 420<br>/AESIVRYV<br>/IAETLSLCFV-<br>/GEAVVKLP<br>DWRFSALQF<br>FGRAIASAKLA-<br>APEEKLSVELAE                          |
| A.fulgius 1Q6W  | FAYLTA  | OFFPLHTD  | VEFAKK-T-   | IFGKPIAQ  | QMLVLSIALQM   | VDQ   | VIL   | SNYDVSSV   | IAFFGIK   | DVRFLRPVF  | IGDTIAA   | SAEVVEKQDFD  | E-KSGVVTY  | KLEVKN  | QRGELVL  | FALYSALIRKTP  |
| H.sapiens 1S9C<br>M.tuberculosis Rv3389c<br>M.tuberculosis Rv3538<br>P.aeruginosa PhaJ2<br>P.aeruginosa PhaJ3                               | YRLS-GI<br>IYRLSGI<br>YRLC-GI<br>YRLN-GI<br>ARAA-GI                   | WNPLHIC<br>DRNPLHSC<br>DRNPLHSC<br>DDNPLHAE<br>DYNPIHLS                           | PNFASL-A-<br>PWFATQLA-<br>PEFAAA-A-<br>PAAALR-A-<br>APSAKL-F-                       | GFDKPILH<br>GFPKPILH<br>GFPRPILH<br>GFPRPILH<br>GFPRAIAH                    | GLCTFGFSARR<br>GLCTYGVAGRA<br>GLCTYGMTCKA<br>GLCTLGVAFHA<br>GLWNKARSLAA                                     | VLQ<br>LVA<br>IVD<br>VLRGL<br>LGER-   | QFA<br>ELG<br>ALL<br>ADYRAG<br>LPA                            | DNDVS<br>GCVAA<br>DSDAT                            | RFKAV<br>NITSI<br>AVAGY<br>QLGHL<br>SGYRV   | KARFAKPVY<br>AARFTKPVF<br>GARFAGVAY<br>QVRFSAPVF<br>EVRFQKPVL                                  | PGQTLQT<br>PGETLST<br>PGETLTV<br>PGETLRT<br>LPASLTL   | EMWK<br>VIWR<br>NVWK<br>EMWS<br>LASA   | EG-NRI<br>TEPGRA<br>DG-RRL<br>DGS<br>AAADG   | HFQTKV<br>VFRTEV<br>VASVVA<br>FRTRV<br>QFSL                             | QE-TGDIVIS<br>AG-SDGAEAF<br>PT-RDN/<br>VE-RDVVVLI<br>RG-KDDLPHM                | SNAYVDLAPTSG<br>XVVLDDGAVEYV<br>XVV-LSGVELVP<br>DNGRVGPPPERP<br>MAGHWSRLQG  |
| C.tropicalis 1PN2<br>M.tuberculosis Rv2524c<br>M.tuberculosis Rv0241c   | YRLS-CI<br>VS-CI<br>VG-CI   | ORNPLHIC<br>DHNPIHTC<br>DHNPIHTN  | PNFAKG-A-<br>RAAALL-A-<br>IPIAAKL-F-  | KFPKPILH<br>GLESPIVH<br>GFPTVIAH  | CMCTYGLSAKA<br>CMWLSAAA<br>CMFTAAA  | LID<br>QHAVT<br>VLA   | KFG<br>A-T  | DGQARPPA<br>EARFP                                  | MFNEI<br>RLVGW<br>DAVRY   | KARFTGIVF<br>TARFLOMVR<br>SVRFAKPVL  | PGETLRV<br>PGDEVDF<br>LPATAGL   | LAWK<br>RVERVGI<br>YVAEG   | ESDDTI<br>DQGAEI   | -VFQTHV<br>-VDVAAR<br>-TLRNMA   | VD-RGTIAIN<br>VG-SDLVMS/<br>KG-YPHLTAT   | NAAIKLVGDKA<br>ASARLAAPKTVY<br>FVRGL  |

Figure 3. Structure-based sequence alignment. The alignment is grouped according to the closest related structures. The SHD-type proteins are represented twice to enable comparison with the N- and C-terminal domains of the DHD structures.

#### Discussion

As expected, Rv0130 forms a single hotdog fold (Castell et al. 2005) with a large, highly curved  $\beta$ -sheet that wraps around a long, central  $\alpha$ -helix. Like many SHD proteins, two Rv0130 molecules pack side-by-side to form a homodimer with an extended sheet. Rv0130 features a highly conserved R-specific hydratase motif (Qin et al. 2000), buried deeply between the two subunits of the dimer. Indeed, our biochemical studies show that Rv0130 is able to hydrate a *trans*-2-enoyl-CoA moiety with a  $k_{cat}$  of  $1.1 \times 10^2 \text{ sec}^{-1}$  and a  $K_{\rm M}$  of 55  $\mu$ M. The two potential active site residues, D40 and H45, were confirmed to be important for this catalytic activity by site-directed mutagenesis. The active site also shares some similarity with the P. aeruginosa and the P. falciparum dehydratases, as well as the E. coli dehydratase-isomerase, indicating the catalytic relationship between the hydratase motif and the dehydratase/dehydratase-isomerase active sites.

In contrast to most bacteria, M. tuberculosis uses two types of systems for fatty acid synthesis, FAS I and FAS II. FAS I is a large multifunctional enzyme, commonly found in eukaryotes, containing all the necessary functions to perform de novo fatty acid synthesis. The FAS II system, on the other hand, contains several dissociated enzymes, each catalyzing a separate reaction (for recent review, see Takayama et al. 2005). Both FAS I and FAS II systems include a dehydration step, which in FAS II is catalyzed by a  $\beta$ -hydroxyacyl-ACP dehydratase, FabZ, which still remains to be identified in *M. tuberculosis*. In FAS I we have suggested that the dehydratase activity of M. tuberculosis FAS I is localized in a DHD (Castell et al. 2005) with a well conserved hydratase motif at the active

site (Oin et al. 2000). In a recent major advance in our understanding of this system, Ban and colleagues have described the mammalian and fungal FAS I structures at 4.5 and 5.0 Å resolution, respectively (Jenni et al. 2006; Maier et al. 2006). In both systems, they localize the dehydratase domain to a DHD of "about twice the expected size" (Maier et al. 2006), confirming our hypothesis. It is becoming clear, therefore, that the hydratase motif is able to shift the reaction equilibrium toward either hydration or dehydration. Based on this catalytic ambiguity and the data describing *M. tuberculosis* growth in vitro and in vivo (Sassetti and Rubin 2003; Sassetti et al. 2003), Castell et al. (2005) suggested Rv0636 to be the missing FabZ of M. tuberculosis. Interestingly, the classical dehydratase/dehydratase-isomerases also exhibit some catalytic ambiguity; both FabZ and FabA possess substantial hydratase activity in vitro (Heath and Rock 1995, 1996; Sharma et al. 2003). In vivo, the FAS II enoyl-ACP reductase FabI pulls the equilibrium toward dehydration by reducing the enoyl-ACP chain produced by FabZ/FabA to acyl-ACP (Heath and Rock 1995; Sharma et al. 2003). Thus only a low dehydratase activity might be required from a putative hydratase-motif-containing enzyme to act in fatty acid elongation FAS II.

Compared with the previously known structures, the packing of the C-terminal portion of the Rv0130 hotdog helix is altered by a number of changes, including a proline residue situated in the middle of the helix. This modification opens up the dimer interface and creates a long tunnel housing the active site. In the eukaryotic hydratase 2 enzymes, a similar structural feature enables the processing of long chain CoA-thioesters (Koski et al. 2004, 2005). Koski and colleagues argue that in their structure(s) the long loop between the hotdog helix and  $\beta$ 2 is highly flexible, adapting to accommodate a long fatty acid substrate. In Rv0130, the end of the tunnel seems to be more rigid, defined by the C-terminal half of the hotdog helix and by the small  $\beta$ -sheet. It is likely that Rv0130 also is able to act on long fatty acid enoyl chains, but with different substrate specificity. In the known 3D structures, the length of the hotdog helix and connecting loop to  $\beta 2$  correlates with the presence of an active site cavity or tunnel. Of the four R-hydratases from P. aeruginosa, PhaJ1 is predicted to feature a substratebinding pocket, whereas PhaJ2, PhaJ3, and PhaJ4 may feature substrate-binding tunnels. These pocket/tunnel predictions correlate well with the reported substrate specificities. Applying the same reasoning to the five other hydratase/dehydratase proteins from M. tuberculosis that Castell et al. (2005) predicted to exhibit SHD/DHD-type structures, it is suggested that a Rv0636 homodimer features a substrate-binding pocket, while the Rv3389c, Rv3538, Rv2524c, and Rv0241 DHDs have substratebinding tunnels. However, since FAS II FabZ has a

preference to act on long-chain fatty acid derivatives (Takayama et al. 2005), it will most likely require a substrate-binding tunnel, thus making Rv0636 a less likely candidate. On the other hand, the sequence variation of the C-terminal portion of the hotdog helix and the following loop is quite extensive in the large family of *R*-specific enoyl hydratases, making sequence to structure predictions difficult. One example of how acyl chain length specificity can depend on small sequence changes was reported by Tsuge et al. (2003a), who was able to alter the substrate specificity of the *A. caviae* hydratase 1IQ6 from C4–C6 to C4–C10 by a single point mutation. Thus, the precise relation between *R*-hydratase sequence and substrate specificity still requires detailed crystallographic analysis.

# Materials and methods

## Cloning, expression, and purification

The open reading frame coding for rv0130 was amplified by PCR from cosmid MTCI5 (Philipp et al. 1996) of M. tuberculosis, strain H37Rv, using the primers 5'-TACTGCCGACCAGA GGAGAG (forward) and 5'-ATTCTGACCGCGAGTTGC (reverse). In a second PCR reaction an N-terminal six-histidine tag was introduced using the forward primer 5'-ATGGCTCATCAT CATCATCATCATGGTCGCACCTTCGAGTCGGTCG. The high fidelity polymerase PfuUltra (Stratagene) was used for the amplifications. After addition of a 3' A-overhang by incubation with Taq polymerase (New England Biolabs), the DNA fragment was ligated to the pCR T7 TOPO vector (Invitrogen). Cloning was performed in E. coli TOP10 (Invitrogen), and the correctness of the isolated gene was verified by DNA sequence analysis. Expression was performed in E. coli BL21(DE3) (Invitrogen). The cells were cultured in LB media at 37°C. At  $OD_{600} = 0.5-1.0$  the temperature was lowered to 24°C and expression of the target gene induced with 0.1 mM IPTG overnight. After harvesting by centrifugation the cell pellet was resuspended in lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, 10% [v/v] glycerol, 0.5% [v/v] Triton X-100 at pH 8.0) with 1 mg/mL lysozyme, 0.01 mg/mL RNase A, and 0.02 mg/mL DNase I and lysed using a One Shot Cell disruptor (Constant Systems Ltd.). The soluble fraction was incubated with 0.7 mL pre-equilibrated Ni-NTA agarose (Qiagen) slurry at 4°C for 1 h then packed into a column. After washing with 20 column volumes of buffer (20 mM imidazole, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10% [v/v] glycerol at pH 8.0), the protein was eluted with five column volumes of 250 mM imidazole in the same buffer. Fractions containing the protein were pooled and further purified on a HiLoad 16/60 Superdex 75 prep grade column (GE Healthcare) equilibrated in 20 mM Tris (pH 7.5), 150 mM NaCl, 10% glycerol.

Selenomethionine (SeMet) was introduced into the protein by metabolic inhibition (van Duyne et al. 1993). The SeMet protein was expressed in BL21(DE3) cells (Invitrogen) grown in minimal media supplemented with SeMet, lysine, threonine, phenylalanine, leucine, isoleucine, and valine. Expression was done at room temperature overnight. The purification protocol for the SeMet protein was the same as for the native protein but with 10 mM  $\beta$ -mercaptoethanol in all buffers to prevent selenomethionine oxidation.

#### **Mutagenesis**

Mutagenesis was performed with the QuikChange site-directed mutagenesis kit (Stratagene). Temperature cycling was done according to the manual with the *rv0130*-containing pCR T7 plasmid as template. The two mutagenic primers D40N (5'-CCGACGCAAC GGGTAACCACCAGTGGATCCA-3') and antisense D40N (5'-TGGATCCACTGGTGGTTACCCGTTGCGTCGG-3') were used for mutation of Asp40 to Asn and the primer pair H45Q (5'-GATCACCAGTGGATCCAAGTCGACCTGGATCCACTGGTGATCCACTGGTGATCCACTGGTGATCCACTGGTGATCCACTGGTGATCCACTGGTGA TC-3') for mutation of His45 to Gln. After digestion of parental DNA template with DpnI, the DNA was transformed into Top10 supercompetent cells. Mutations were verified with sequence analysis.

## Enzymatic activity

The hydratase activity was measured spectrophotometrically in an assay that monitors the decrease in absorbance at 263 nm when crotonoyl CoA is hydrated to 3-hydroxybutyryl-CoA (Moskowitz and Merrick 1969; Fukui et al. 1998). The reaction was started by addition of enzyme (0.15 pmol of the wild-type enzyme) to varying concentrations (5-100 µM) of crotonoyl-CoA in 50 mM Tris (pH 8.0) at 30°C. The wild-type enzyme was freshly diluted from a 0.15 mM solution before each measurement. For the D40N mutant 0.77 nmol of enzyme, from a 0.31 mM solution, was added to each experiment and for the H45O mutant up to 3.4 pmol were tested on 25 and 100 µM substrate. In order to get consistent results, the wild-type enzyme had to be diluted just before the reaction was started. The D40N mutant was not diluted but even here a correlation was seen with the freshness of the enzyme and a new aliquot of protein was taken from the storage at  $-80^{\circ}$ C the same day as the measurement was done. The reactions were carried out in a total volume of 1 mL in a quartz cuvette with a 1-cm light path and readings were taken every second. The  $\epsilon_{263}$  of the enoyl-thioester bond is taken to be 6.7  $\times$  $10^3 \text{ M}^{-1} \text{ cm}^{-1}$  (Stern et al. 1956). The kinetic constants were determined by measuring the reaction rate during the first 2 min of the reaction. Kinetic constants were calculated from Hanes-Woolf plots.

### Crystallization

Purified SeMet protein was crystallized by the vapor diffusion method (McPherson 1982). The best crystals were grown at room temperature by mixing 0.5  $\mu$ L of 2 mg/mL protein solution, still in its purification buffer, with 0.5  $\mu$ L of 0.1 M HEPES (pH 7.5), 2% PEG 400, and 2 M ammonium sulfate. Crystals in sizes ~0.07 × 0.05 × 0.05 mm<sup>3</sup> could be harvested within 3–4 d.

#### X-ray data collection and processing

Before flash freezing in liquid nitrogen, Rv0130 crystals were transferred to a second drop containing mother liquor with 20% PEG400 added as a cryo-protectant. X-ray diffraction data were collected at ESRF beamline ID14:3 from a single SeMet crystal above the Se-absorption edge to a resolution of 1.7 Å. The data were indexed, integrated, and scaled using the HKL suite (Otwinowski and Minor 1997) and were found to belong to spacegroup P2<sub>1</sub>2<sub>1</sub>2 with unit cell parameters a = 58.7 Å, b = 108.7 Å, and c = 50.4 Å, assuming two molecules in the

Chem. 2

asymmetric unit give a Matthews coefficient of 2.5  $Å^3/Da$  (Matthews 1968), which corresponds to a solvent content of 50.6%. Data collection statistics are given in Table 1.

Four selenium atoms were located using SHELXD (Schneider and Sheldrick 2002) and verified using MLPHARE (Otwinowski 1991). Phase estimates were calculated by SHARP (de la Fortelle and Bricogne 1997). The NCS operators could be identified using O (Jones et al. 1991) and refined with Imp (Jones 1992; Kleywegt et al. 2001). Subsequent twofold averaging, solvent flattening, and histogram matching in DM (Cowtan and Main 1998) improved maps considerably. The ARP/wARP suite (Perrakis 1997) was used for automatic main- and side-chain tracing. Crystallographic refinement was done in REFMAC5 (Murshudov 1997) against the SeMet data to a resolution of 1.8 Å. After each cycle of refinement, the  $\sigma_A$ -weighted  $|2F_{obs}| - |F_{calc}|$  and  $|F_{obs}| - |F_{calc}|$ maps (Read 1986) were used for further model rebuilding in O (Jones et al. 1991). Water molecules were added using ARP/ wARP (Perrakis 1997) and O, in peaks (>2 $\sigma$ ) of the  $|2F_{obs}|$  -|F<sub>calc</sub>| electron density map. Final refinement statistics are given in Table 1. Structural alignments were made using the Lsq tools of O, and structure-based sequence alignments were made using the INDONESIA package (D. Madsen, P. Johansson, and G.J. Kleywegt, unpubl.). Molecular figures were made in O and rendered in Molray (Harris and Jones 2001). Accessible surface area calculations were performed by PDBsum (Laskowski et al. 2005). Coordinates and structure factors of Rv0130 have been deposited at the Protein Data Bank with codes 2C2I and R2C2ISF, respectively.

#### Acknowledgments

Financial support was received from Swedish Foundation for Strategic Research (SSF), the Swedish Natural Science Research Council, and European Commission programs SPINE (QLG2-CT-2002–00988) and NM4TB (LSHP-CT-2005-018923). Total DNA of *M. tuberculosis*, strain H37Rv, was a gift from Dr. Stewart Cole (Institut Pasteur, Paris, France).

#### References

- Benning, M.M., Wesenberg, G., Liu, R., Taylor, K.L., Dunaway-Mariano, D., and Holden, H.M. 1998. The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3. *J. Biol. Chem.* 273: 33572–33579.
- Camus, J.C., Pryor, M.J., Médigue, C., and Cole, S.T. 2002. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiol.* 148: 2967–2973.
- Castell, A., Johansson, P., Unge, T., Jones, T.A., and Bäckbro, K. 2005. Rv0216, a conserved hypothetical protein from *Mycobacterium tuberculosis* that is essential for bacterial survival during infection, has a double hotdog-fold. *Protein Sci.* 14: 1850–1862.
- Chothia, C. and Janin, J. 1982. Orthogonal packing of β-pleated sheets in proteins. *Biochemistry* 21: 3955–3965.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393: 537–544.
- Cowtan, K. and Main, P. 1998. Miscellaneous algorithms for density modification. Acta Crystallogr. D Biol. Crystallogr. 54: 487–493.
- de la Fortelle, E. and Bricogne, G. 1997. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* 276: 472– 494.
- Dye, C., Scheele, S., Dolin, P., Pathania, V., and Raviglione, M.C. 1999. Consensus statement. Global burden of tuberculosis: Estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 282: 677–686.

- Engh, R. and Huber, R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr. A 47: 392–400.
- Fukui, T. and Doi, Y. 1997. Cloning and analysis of the poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) biosynthesis genes of Aeromonas caviae. J. Bacteriol. 179: 4821–4830.
- Fukui, T., Shiomi, N., and Doi, Y. 1998. Expression and characterization of (*R*)specific enoyl coenzyme A hydratase involved in polyhydroxyalkanoate biosynthesis by *Aeromonas caviae*, J. Bacteriol, 180: 667–673.
- Harris, M. and Jones, T.A. 2001. Molray—A web interface between O and the POV-Ray ray tracer. Acta Crystallogr. D Biol. Crystallogr. 57: 1201– 1203.
- Heath, R.J. and Rock, C.O. 1995. Enoyl-acyl carrier protein reductase (*fab1*) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. J. Biol. Chem. 270: 26538–26542.
- . 1996. Roles of the FabA and FabZ β-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. J. Biol. Chem. 271: 27795–27801.
- Helmkamp Jr., G.M. and Bloch, K. 1969. β-hydroxydecanoyl thioester dehydrase. Studies on molecular structure and active site. J. Biol. Chem. 244: 6014–6022.
- Hendrickson, W.A. 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254: 51–58.
- Hisano, T., Tsuge, T., Fukui, T., Iwata, T., Miki, K., and Doi, Y. 2003. Crystal structure of the (*R*)-specific enoyl-CoA hydratase from *Aeromonas caviae* involved in polyhydroxyalkanoate biosynthesis. J. Biol. Chem. 278: 617– 624.
- Holm, L. and Sander, C. 1994. Searching protein structure databases has come of age. *Proteins* 19: 165–173.
- Imaeda, T., Kanetsuna, F., and Galindo, B. 1968. Ultrastructure of cell walls of genus *Mycobacterium. J. Ultrastruct. Res.* 25: 46–63.
- Jarlier, V. and Nikaido, H. 1990. Permeability barrier to hydrophilic solutes in Mycobacterium chelonei. J. Bacteriol. 172: 1418–1423.
- Jenni, S., Leibundgut, M., Maier, T., and Ban, N. 2006. Architecture of a fungal fatty acid synthase at 5 Å resolution. *Science* 311: 1263–1267.
- Jones, T.A. 1992. A, yaap, asap, @#\*? A set of averaging programs. In Molecular replacement (eds. E.J. Dodson et al.), pp. 91–105. SERC Daresbury Laboratory, Warrington, UK.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47: 110–119.
- Kimber, M.S., Martin, F., Lu, Y., Houston, S., Vedadi, M., Dharamsi, A., Fiebig, K.M., Schmid, M., and Rock, C.O. 2004. The structure of (3R)hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas* aeruginosa. J. Biol. Chem. 279: 52593–52602.
- Kleywegt, G.J. and Jones, T.A. 1996. Phi/psi-chology: Ramachandran revisited. Structure 4: 1395–1400.
- Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M., and Jones, T.A. 2001. Around O. In *International tables for crystallography* (eds. M.G. Rossmann and E. Arnold), pp. 353–356, 366–367. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Koski, M.K., Haapalainen, A.M., Hiltunen, J.K., and Glumoff, T. 2004. A twodomain structure of one subunit explains unique features of eukaryotic hydratase 2. J. Biol. Chem. 279: 24666–24672.
- 2005. Crystal structure of 2-enoyl-CoA hydratase 2 from human peroxisomal multifunctional enzyme type 2. J. Mol. Biol. 345: 1157–1169.
- Kostrewa, D., Winkler, F.K., Folkers, G., Scapozza, L., and Perozzo, R. 2005. The crystal structure of PfFabZ, the unique β-hydroxyacyl-ACP dehydratase involved in fatty acid biosynthesis of *Plasmodium falciparum*. *Protein Sci.* 14: 1570–1580.
- Laskowski, R.A., Chistyakov, V.V., and Thornton, J.M. 2005. PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* 33: D266–D268.
- Leesong, M., Henderson, B.S., Gillig, J.R., Schwab, J.M., and Smith, J.L. 1996. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: Two catalytic activities in one active site. *Structure* 4: 253–264.
- Li, J., Derewenda, U., Dauter, Z., Smith, S., and Derewenda, Z.S. 2000. Crystal structure of the *Escherichia coli* thioesterase II, a homolog of the human Nef binding enzyme. *Nat. Struct. Biol.* **7:** 555–559.

- Madison, L.L. and Huisman, G.W. 1999. Metabolic engineering of poly(3hydroxyalkanoates): From DNA to plastic. *Microbiol. Mol. Biol. Rev.* 63: 21–53.
- Maier, T., Jenni, S., and Ban, N. 2006. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* 311: 1258–1262.
- Matthews, B.W. 1968. Solvent content of protein crystals. J. Mol. Biol. 33: 491–497.
- McPherson, A. 1982. *Preparation and analysis of protein crystals*. John Wiley & Sons, New York.
- Moskowitz, G.J. and Merrick, J.M. 1969. Metabolism of poly-β-hydroxybutyrate. II. Enzymatic synthesis of D-(–)-β hydroxybutyryl coenzyme A by an enoyl hydrase from *Rhodospirillum rubrum*. *Biochemistry* 8: 2748–2755.
- Murshudov, G.N. 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53: 240–255.
- O'Regan, A. and Joyce-Brady, M. 2001. Latent tuberculosis may persist for over 40 years. *BMJ* 323: 635.
- Otwinowski, Z. 1991. Maximum likelihood refinement of heavy atom parameters. In CCP4 study weekend. Isomorphous replacement and anomalous scattering (eds. W. Wolf et al.), pp. 80–86. Daresbury Laboratory, Warrington, UK.
- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276: 307–326.
- Perrakis, A. 1997. wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr. D Biol. Crystallogr.* 53: 448–455.
- Philipp, W.J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., Bergh, S., Bloom, B.R., Jacobs Jr., W.R., and Cole, S.T. 1996. An integrated map of the genome of the tubercle bacillus, Mycobacterium tuberculosis H37Rv and comparison with Mycobacterium leprae. Proc. Natl. Acad. Sci. 93: 3132–3137.
- Qin, Y.M., Haapalainen, A.M., Kilpeläinen, S.H., Marttila, M.S., Koski, M.K., Glumoff, T., Novikov, D.K., and Hiltunen, J.K. 2000. Human peroxisomal multifunctional enzyme type 2. Site-directed mutagenesis studies show the importance of two protic residues for 2-enoyl-CoA hydratase 2 activity. *J. Biol. Chem.* 275: 4965–4972.
- Read, R.J. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A42: 140–149.
- Sassetti, C.M. and Rubin, E.J. 2003. Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci.* 100: 12989–12994.
- Sassetti, C.M., Boyd, D.H., and Rubin, E.J. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 48: 77–84.
- Schneider, T.R. and Sheldrick, G.M. 2002. Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. 58: 1772–1779.
- Sharma, S.K., Kapoor, M., Ramya, T.N., Kumar, S., Kumar, G., Modak, R., Sharma, S., Surolia, N., and Surolia, A. 2003. Identification, characterization, and inhibition of *Plasmodium falciparum* β-hydroxyacyl-acyl carrier protein dehydratase (FabZ). *J. Biol. Chem.* 278: 45661–45671.
- Stern, J.R., del Campillo, A., and Raw, I. 1956. Enzymes of fatty acid metabolism. I. General introduction; crystalline crotonase. J. Biol. Chem. 218: 971–983.
- Sudesh, K., Fukui, T., Iwata, T., and Doi, Y. 2000. Factors affecting the freezefracture morphology of in vivo polyhydroxyalkanoate granules. *Can. J. Microbiol.* 46: 304–311.
- Takayama, K., Wang, C., and Besra, G.S. 2005. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.* 18: 81–101.
- Tsuge, T., Hisano, T., Taguchi, S., and Doi, Y. 2003a. Alteration of chain length substrate specificity of *Aeromonas caviae R*-enantiomer-specific enoylcoenzyme A hydratase through site-directed mutagenesis. *Appl. Environ. Microbiol.* 69: 4830–4836.
- Tsuge, T., Taguchi, K., Seiichi, T., and Doi, Y. 2003b. Molecular characterization and properties of (*R*)-specific enoyl-CoA hydratases from *Pseudomonas aeruginosa*: metabolic tools for synthesis of polyhydroxyalkanoates via fatty acid β-oxidation. *Int. J. Biol. Macromol.* **31**: 195–205.
- van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. 229: 105–124.
- Venkatachalam, C.M. 1968. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* 6: 1425–1436.