Rv0216, a conserved hypothetical protein from *Mycobacterium tuberculosis* that is essential for bacterial survival during infection, has a double hotdog fold

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

Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, SE-751 24 Uppsala, Sweden

(RECEIVED March 4, 2005; FINAL REVISION April 22, 2005; ACCEPTED April 22, 2005)

Abstract

The *Mycobacterium tuberculosis* genome contains about 4000 genes, of which approximately a third code for proteins of unknown function or are classified as conserved hypothetical proteins. We have determined the three-dimensional structure of one of these, the *rv0216* gene product, which has been shown to be essential for *M. tuberculosis* growth in vivo. The structure exhibits the greatest similarity to bacterial and eukaryotic hydratases that catalyse the R-specific hydration of 2-enoyl coenzyme A. However, only part of the catalytic machinery is conserved in Rv0216 and it showed no activity for the substrate crotonyl-CoA. The structure of Rv0216 allows us to assign new functional annotations to a family of seven other *M. tuberculosis* proteins, a number of which are essential for bacterial survival during infection and growth.

Keywords: *Mycobacterium tuberculosis*; Rv0216; hotdog fold; protein structure; crystallography

Supplemental material: see www.proteinscience.org

Every year two million people die from tuberculosis (TB), a contagious disease caused by *Mycobacterium tuberculosis*. The slow growth of this bacillus and its tendency to hide in a latent stage makes it difficult to treat. This is exacerbated by the emergence of multidrug resistant strains, which makes the need for new, better drugs compelling (Stewart et al. 2003). A major step

forward in the battle against TB took place in 1998 when the complete sequence of the *M. tuberculosis* genome was published (Cole et al. 1998). This allowed the identification and functional annotation of ~4000 genes. In the initial analysis, 606 proteins of unknown function were identified, as well as 910 so-called conserved hypothetical proteins. A more recent annotation has reduced the number of proteins of unknown function to 272, but the number of conserved hypothetical proteins has risen to 1051 (Camus et al. 2002). Various genomic approaches have been used to evaluate the importance of particular gene products for the organism's survival. When using transposon site hybridization, for example, Sassetti and Rubin (2003) identified 194 genes that are essential for *M. tuberculosis* survival in a mouse model of TB. In this subset of genes, 55% code for proteins of unknown function. The availability of whole genome sequences from other mycobacteria

Reprint requests to: T. Alwyn Jones or Kristina Bäckbro, Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden; e-mail: alwyn@xray.bmc.uu.se or nina@xray.bmc.uu.se; fax: +46-18-536971.

Abbreviations: ACP, acyl carrier protein; CoA, coenzyme A; COG, cluster of orthologous groups; DHD, double hotdog; FAS, fatty acid synthase; PDB, Protein Data Bank; PMSF, phenylmethane sulfonyl fluoride; RMSD, root-mean-square distance; SeMet, seleno-methionine; SHD, single hotdog; TB, tuberculosis.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051442305>.

Table 1. Data collection statistics

	Peak	Inflection Point	Remote	Native
Beamline	ID29; ESRF	ID29; ESRF	ID29; ESRF	ID14 EH2; ESRF
Wavelength (Å)	0.97928	0.97935	0.97564	0.933
Resolution range (Å) ^a	89.2–1.9 (2.0–1.9)	89.9–2.6 (2.74–2.6)	89.9–2.5 (2.64–2.5)	90.2–1.9 (2.0–1.9)
No. of unique reflections	19,417	10,713	12,006	26,988
Redundancy	3.6	12.6	10.4	13.7
Completeness (%) ^a	84.4 (63.5)	100 (99.9)	100 (99.9)	100 (99.9)
Mean I/σI	8.7 (2.4)	6.4 (2.2)	6.5 (2.0)	7.0 (1.8)
R _{merge} (%) ^{a,b}	8.1 (43.2)	11.4 (39.2)	11.9 (43.2)	9.7 (41.7)

^a Values in parentheses refer to the outer resolution shell.

^b $R_{\text{merge}} = [\sum_{\text{hkl}} \sum_i |I - \langle I \rangle| / \sum_{\text{hkl}} \sum_i I] \cdot 100\%$.

provides additional information for the identification of essential gene products. The genome of *M. tuberculosis*, for example, is larger than the *Mycobacterium leprae* genome, which contains only 1605 protein-encoding genes. *M. leprae* seems to have undergone reductive evolution, retaining only the genes essential for mycobacterial pathogenesis (Cole et al. 2001). *M. tuberculosis* and *M. leprae* have 1433 genes in common and, after eliminating the genes that occur in other prokaryotes and in eukaryotes, there are only 219 genes that are unique to these two mycobacteria (Cole 2002). Most of these 219 are also present in *M. avium*, *M. marinum*, and *M. smegmatis*, indicating that they are mycobacteria specific (Marmiesse et al. 2004). Proteins that are essential for the pathogen, and at the same time specific to mycobacteria, are particularly interesting both for the understanding of mycobacterial biology and as potential drug targets. The gene product of *rv0216* (referred to here as Rv0216) satisfies both of these two criteria. It is essential for the survival of *M. tuberculosis* in vivo (Sasseti and Rubin 2003) and it is one of the 219 core genes conserved in several mycobacterial strains (Cole 2002; Marmiesse et al. 2004). At the Pasteur Institut TubercuList server (<http://genolist.pasteur.fr/TubercuList/>) it is classified as a conserved hypothetical protein of unknown function.

Three-dimensional structural information on Rv0216 might cast light on its biological function. Indeed, our crystallographic studies show that Rv0216 exhibits overall structural similarities to a group of enzymes that use thiol esters as substrates. The greatest similarity is to bacterial and eukaryotic R-specific enoyl hydratases, but Rv0216 is distinct from these enzymes in the composition of its putative catalytic residues. The new structure allows us to re-evaluate the structure–function relationship of a group of *M. tuberculosis* proteins that are likely to have structures similar to Rv0216. Half of these proteins have been shown to be essential for bacterial survival during infection (Sasseti and Rubin 2003) and/or growth (Sasseti et al. 2003).

Results

Overall structure

The crystal structure of His-tagged, full-length Rv0216 has been determined by multiple-wavelength anomalous dispersion (MAD) phasing of seleno-methionine (SeMet) substituted protein. The crystals contain a monomer of 337 residues in the asymmetric unit, and the initial model has been refined to a resolution of 1.9 Å resolution (Tables 1, 2).

Table 2. Phasing and refinement statistics

	Peak	Inflection	Remote	Native
Resolution (Å)	1.9	2.6	2.5	
No. of sites	3			
Mean figure of merit acentric	0.22			
Mean figure of merit centric	0.28			
Phasing power	0.49	0.79	0.28	
Anomalous R _{Cullis}	0.94	0.96	0.97	
Resolution (Å)				91.29–1.9
No. of unique reflections				25,592
No. of reflections in test set				1355
R _{work} (%)				18.6
R _{free} (%)				22.2
No. of atoms				
Protein				2482
Solvent				325
Ramachandran plot outliers ^a				0
Average atomic temperature factors				
Protein (Å ²)				18.2
Solvent (Å ²)				30.7
r. m. s. deviation from ideal geometry ^b				
Bond distance (Å)				0.012
Bond angle (°)				1.44

^a Calculated using a strict-boundary Ramachandran plot (Kleywegt and Jones 1996).

^b Ideal values from Engh and Huber (1991).

Figure 1. The structure of Rv0216, colored from red to blue, N- to C-terminal, shows a double hotdog fold. The monomeric protein has an N- and a C-terminal domain, each of which consists of a five-stranded β -sheet wrapping around a long α -helix. (A) Ribbon representation of the overall structure. (B) Topology diagram.

The somewhat elongated molecule is $\sim 35 \times 45 \times 55$ Å and is built up primarily of an extended anti-parallel β -sheet of 10 strands and two long α -helices ($\alpha 3$ and $\alpha 6$) (Fig. 1). Two similar structural repeats of ~ 140 residues are related by an approximate twofold rotation axis. Each repeat consists of a domain that is built up from a β -sheet of five strands (ordered 1-3-4-5-2) that wraps around a long α -helix. The β -sheets pack around the local dyad axis such that the second strand in each domain abuts to create the extended sheet. The helix packs onto the concave surface of the sheet and is located in the sequence between the first two β -strands of each domain. It is linked to the end of the first β -strand by an ~ 45 -residue stretch that folds to form a lid that interacts with the long helix of the other domain (Fig. 1). The lid of the first domain contains two short α -helices, while the lid of the second domain has only one. The last strand of the first domain is connected to the first strand of the second by a 40-residue loop that extends over the entire convex surface of the extended β -sheet. Each domain has the so-called hotdog topology first seen in *Escherichia coli* β -hydroxydecanoyl thiol ester dehydratase (Leesong et al. 1996). This enzyme, also called FabA, catalyses two reactions (dehydration and isomerization) on fatty acid thiol esters of acyl carrier protein (ACP) (Bloch 1971). It is a homodimer with a quaternary structure similar to Rv0216. The double hotdog (DHD) fold of Rv0216 has also been previously observed; the first report of it was in *E. coli* thioesterase II (Li et al. 2000).

Structural repeat

The single hotdog (SHD) domains of Rv0216 are remarkably similar to each other. The CA atoms of 121

pairs of residues can be aligned with a root-mean-square distance (RMSD) of 1.9 Å (Table 3). At the amino acid level this structural alignment shows 24 identical residues, corresponding to a sequence identity of $\sim 16\%$. The longest conserved stretch of amino acids is the pentapeptide G¹⁰⁶-DTLY/G²⁸⁶-DTLY starting at the short loop connecting the second and third β -strands of each domain. This is an interesting peptide stretch that crosses over the β -meander made up from the fourth and fifth strands, on the concave face of the sheet (Fig. 2A). The side-chain of D107 helps stabilize this loop by forming a hydrogen bond to the main-chain amide of 104. One other conserved side-chain, Q80, interacts via hydrogen bonds to the main chain of this pentapeptide. Two other conserved clusters occur along strands $\beta 4$ and $\beta 5$ (G¹³⁴-LR/G³⁰⁸-LR and D¹⁴⁴RxVLD/D³²⁴RxVLD), but only one set of conserved side-chain interactions is retained: the salt link across the strands between R136 and D149.

Similarity to other structures

Twelve structures have been deposited at the Protein Data Bank (PDB) (Berman et al. 2000) with the hotdog domain topology we observe in Rv0216. Ten of these are SHD homodimers, while two are DHD internal repeat monomers where the last strand of the first domain is connected to the first of the second domain by a long loop passing across the convex surface of the central β -sheet. Structure alignment statistics are given in Table 3 and the superimposed structures are shown in Figure 3. The best matching structures can be aligned so that ~ 250 CA atoms have RMSDs of < 2 Å while in the less similar structures the number of matching pairs is halved. The two most similar structures to Rv0216 are

Table 3. Structure alignment and DALI statistics

PDB code	Name	DALI score ^a	RMSD	No. of aligned C α atoms	Sequence identity (%)
2BI0	Rv0216 domain to domain DHD structures	13.2	1.9	121	16.0
1PN4	<i>Candida tropicalis</i> 2-enoyl-CoA hydratase 2	19.6	2.1	198	7.7
1C8U	<i>Escherichia coli</i> acyl-CoA thioesterase 2	13.1	2.0	169	5.0
	SHD homodimers to Rv0216 DHD				
1Q6W	<i>Archaeoglobus fulgidus</i> monoamine oxidase regulatory protein	12.0	1.9	253	16.3
1IQ6	<i>Aeromonas caviae</i> (R)-specific enoyl-CoA hydratase	9.9	1.9	233	12.8
1LO7	<i>Pseudomonas</i> 4-hydroxybenzoyl-CoA thioesterase	6.1	1.9	181	6.5
1IXL	<i>Pyrococcus horikoshii</i> hypothetical protein	6.5	1.9	172	11.3
1J1Y	<i>Thermus thermophilus</i> PaaI protein	6.4	2.1	180	9.2
1T82	<i>Shewanella oneidensis</i> putative thioesterase	4.8	2.1	179	3.9
1NNG	<i>Haemophilus influenzae</i> putative acyl-CoA thioester hydrolase	5.8	1.9	168	5.3
1NJK	<i>Escherichia coli</i> hypothetical protein Ybaw	5.5	2.0	163	5.9
1MKA	<i>Escherichia coli</i> β -hydroxydecanoyl thiol ester dehydrase	5.0	2.1	163	3.9
1Q4T	<i>Arthobacter</i> 4-hydroxybenzoyl-CoA thioesterase	5.7	2.3	119	5.9

(DHD) Double hotdog; (SHD) Single hotdog.

^a DALI and DaliLite servers (Holm and Sander 1994).

the (R)-specific enoyl-CoA hydratase from *Aeromonas caviae* (Hisano et al. 2003; PDB code 1IQ6) and the so-called monoamine oxidase regulatory protein from *Archaeoglobus fulgidus* (A.A. Fedorov, E.V. Fedorov, R. Thirumuruhan, and T.A. Jones, unpubl.; PDB code 1Q6W). In these three, the alignment of the β -sheet and central helix are very well conserved, as is the lid section of Rv0216 linking the first strand in each domain to the hotdog helix. The reverse turn between strands 3 and 4 is truncated

in both the 1IQ6 and 1Q6W structures. In the *Archaeoglobus fulgidus* SHD structure, the C terminus of the hotdog helix is extended by one turn. The structure-based sequence identity between Rv0216 and these two structures is 13% and 16%, respectively. In the least similar structure, the 4-hydroxybenzoyl-CoA thioesterase from *Arthobacter* sp. strain SU (Thoden et al. 2003; PDB code 1Q4T), the Rv0216-like lid is much reduced and replaced by an N-terminal extension from a dimer-related chain. In this

Figure 2. (A) A superposition of Rv0216 (side chains of N²³² and H²³⁷ colored dark orange-red) and three hotdog-folded enzyme complexes reveals a putative substrate-binding cavity formed between the two domains (ligands from the dehydrase/isomerase FabA 1MKA in light gray; the hydratase 2 1PN4 in gold and the thioesterase 1Q4T in dark gray). The conserved pentapeptide G²⁸⁶-DTLY from the second domain starts in the short light-blue loop between β 7- β 8 that crosses over the dark blue β 9- β 10 meander. (B) Putative active site of Rv0216 with the catalytic His and Asp residues from the superpositioned *E. coli* dehydrase/isomerase FabA, the *Arthrobacter* sp. thioesterase, and the *C. tropicalis* hydratase 2 together with its ligand, (3R)-hydroxydecanoyl-CoA (coloring as in A).

Figure 3. Structural alignment of Rv0216 and the 12 hotdog-folded proteins found in PDB. (A–C) The hotdog family of proteins shows a striking structural similarity although the sequence homology is very low. (A) A similar view to Figure 1A, showing that the first hotdog helix is less conserved than the rest of the protein (for clarity, the figure has been Z-clipped). (B) Both single and double hotdog-folded proteins share a very similar five plus five-stranded β -sheet despite a number of β -bulges. (C) The loop containing the pentapeptide G²⁸⁶-DTLY between β 2 and β 3 is well conserved in most of the hotdog structures.

comparison, the central 10 β -strands of Rv0216 align less well and the hotdog helix is displaced relative to the sheet. The 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3 (Thoden et al. 2002; PDB code 1LO7) has more similar overall structure to Rv0216 and the main hotdog helices are well aligned. The lid connection in this structure, however, is much shorter so that one edge of the hotdog helix is solvent accessible. Of the DHD structures, the eukaryotic hydratase 2 enzyme (Koski et al. 2004) shows the best overall structural similarity, including a close fit to the lid structures of Rv0216. Although the lid from the second domain is particularly well conserved, the first hotdog helix is highly distorted. The conserved structure-based sequence identity with Rv0216 is only 8%. The superimposed structures in Figure 3B show that the β -sheet is well conserved despite the numerous β -bulges, which produce local distortions. The hotdog helices move as rigid bodies to adapt to changes in sequence, although the first hotdog helix is the least well conserved in structure (Fig. 3A). The largest variations in the hotdog motif occur in the lid and the loops connecting secondary structures. Only the cross-over between β 2 and β 3 is a structurally well-conserved loop (Fig. 3C).

The hotdog-fold proteins with assigned functions are all enzymes. A comparison of them reveals that they appear to share a common active site: The SHD homodimers have two active sites whereas the DHD monomers have only one. The ligands are buried in a cavity formed by the dimer interface (Fig. 2A). They appear to enter between the central two strands of the β -sheet into a cleft formed by the N-terminal portion of the hotdog helix from one domain, and the C-terminal portion of the helix from the other domain. The lid loop, when present, folds on top of the helices to close

off the active site. In at least one structure, eukaryotic hydratase 2, the active site is a tunnel rather than an invagination (Koski et al. 2004).

Putative active site of RV0216

A number of sequences, in a wide range of actino- and proteobacteria, were found to be homologous to Rv0216 using an iterative approach with WU-BLAST2 (Lopez et al. 2003). These Rv0216-like proteins have only a few conserved regions (Fig. 4). Although separated in sequence, these conserved regions cluster in the same neighborhood of a deep pocket, sequestered from the bulk solvent. This pocket coincides with the active site of the enzymes given in Table 3. For three of these hotdog-fold enzymes, a histidine residue in the lid region (H237 in Rv0216) has been implicated in catalysis. In *E. coli* FabA (Leesong et al. 1996) this histidine (residue 70) is the target for a mechanism-based, suicide inhibitor. Together with a neighboring aspartyl side-chain from the twofold related hotdog helix (residue 84), these residues are proposed to catalyze both the dehydration and isomerization reactions performed by the enzyme. The other two enzymes are bacterial and eukaryotic hydratases (Hisano et al. 2003; Koski et al. 2004) that catalyze the R-specific hydration of 2-enoyl coenzyme A (CoA) intermediates to 3-hydroxyacyl-CoA molecules. It has been suggested that the catalytic residues are also a pair of histidine aspartyl residues. The bacterial hydratase is a SHD homodimer with two active sites, while the eukaryotic enzyme is a DHD monomer where only one active site has been conserved. The pair of catalytic residues is located on the same stretch of the lid region, separated by four intervening residues, and the lids in these enzymes have very similar

structures. In the eukaryotic hydratase, this local sequence was identified as a highly conserved segment termed the hydratase 2 motif (Qin et al. 2000), part of which (D-X-N-P-[LIV]-H) contains the proposed active site residues. Their lids differ in structure from the lid in FabA, but the catalytic histidine residues are structurally well conserved (Fig. 2B). The catalytic aspartyl residue of the dehydrase is spatially close to the aspartyl residues of the hydratases, despite being from another subunit. In their discussion of the hydration reaction, both Leesong et al. (1996) and Hisano et al. (2003) favor a mechanism where the aspartic acid residue activates a water molecule, which attacks the C3 atom of a 2-enoyl substrate while the histidine residue donates a proton to the C2 atom of the substrate. In contrast, Koski et al. (2004) favor a mechanism where the catalytic water molecule is first coordinated by the conserved aspartyl, asparagine, and histidine side-chains of the hydratase motif, which then attacks the substrate via a concerted transition state where both the proton and hydroxyl group are derived from the water (Bahnsen et al. 2002). In all these structures, the hydrogen bonds from the unpaired main-chain nitrogen atoms at the N terminus of the hotdog helix are likely to play a role in the interaction with the substrate. The FabA dehydrase is unique among these enzymes in having a second activity, namely, the isomerization of (E)-2-decenoyl-ACP to (Z)-3-decenoyl-ACP. Leesong et al. (1996) argue that this reaction is carried out by the same catalytic machinery.

In the structure of Rv0216, the region equivalent to the hydratase-2 motif is very similar to the structures of the R-specific hydratases. The sequence motif, however, is not conserved. Instead, Rv0216 contains an equivalent $N^{232}-X_4-H^{237}$ that is strictly conserved in all Rv0216-like proteins. The intervening residues show

some conservation. Position 234 has a preference for an alanine or glutamine residue, whereas 235 is a proline in many family members (Supplemental Material). There is no conserved aspartyl residue in the Rv0216-like family close to the hydratase-like active site. The side chains of N232 and H237, however, are positioned to coordinate a water molecule like in the hydratases, and the hotdog helix is positioned to form a stabilizing binding site for an enol/enolate-like intermediate.

In the Rv0216 family, the proposed active site is an enclosed cavity rather than a tunnel, with a volume and shape similar to the bacterial hydratase (Fig. 5A). The residue lining the end of the cavity is a highly conserved arginine residue (R152) that forms hydrogen bonds to the side chains of N91, S81, and T85. The asparagine at residue 91 is conserved in almost all family members. In one member both the residues equivalent to R152 and N91 are changed, and in another the asparagine is changed to a glutamic acid and therefore capable of forming a salt link with the residue equivalent to R152. The side chains that are equivalent to residue 81 are very highly conserved S/T residues. At residue 85, there is a preference for only a threonine. This cluster of residues around the arginine side chain results in satisfying four of the five possible hydrogen bonds via protein side chains and the fifth via a water molecule.

Several of Rv0216's structural neighbors use CoA or ACP-associated acyl-thioesters as substrates. The structure of *Arthrobacter* sp. strain SU 4-hydroxybenzoyl-CoA thioesterase has been described in complex with 4-hydroxyphenacyl-CoA (4HPC) (Thoden et al. 2003) and the *Candida tropicalis* 2-enoyl-CoA hydratase 2 was solved in both unliganded form and in the presence of (3R)-hydroxydecanoyl-CoA (3RHDC) (Koski et al. 2004). Superpositioning the latter enzyme-product com-

Figure 5. (A) The proposed substrate-binding pocket of Rv0216 (red) is similar in size to the (R)-hydratase, 1IQ6 (gray). However, the highly conserved residues (S81, T85, N91, and R152) in Rv0216-like proteins give the Rv0216 cleft a different lining compared with the (R)-hydratase and create a unique hydrogen-bonding network. (B) A conserved patch on the surface of Rv0216 (red) suggests a potential substrate binding site. The ligand is from FabA (Leesong et al. 1996) and is drawn to indicate the position of the putative active site.

plex onto Rv0216, the first half of the long decanoyl tail can be readily fitted into the deep crevice (Fig. 2B). However, since the binding pocket is closed by hotdog helix $\alpha 6$, in contrast to the eukaryotic enzyme, the four aliphatic carbons at the end of the acyl chain cannot be fitted without substantial conformational changes. The substrate, stretching up through the opening of the tunnel, is potentially interacting with a rim of well-conserved residues at the entrance to the proposed active site (including A90, L92, R119, K124, G131, L132, M155, L247, H279, and P282) (Fig. 5B).

Based on the above structural similarities to enoyl-CoA hydratases, the enoyl-CoA hydratase activity of Rv0216 was assayed using the minimal substrate, 2-butenoyl CoA (crotonyl-CoA), as substrate, and found to be inactive.

Related *M. tuberculosis* proteins

Rv0216 has been assigned to Cluster of Orthologous Groups (COG) 2030 (Tatusov et al. 1997) based on its sequence similarity but it is not assigned to any Pfam family (Bateman et al. 2004). However, some members of COG 2030 are also classified as members of Pfam 0175, a family whose members contain a domain of ~ 130 amino acids. This domain corresponds to the SHD domain of the bacterial hydratase of Hisano et al. (2003) and the C-terminal hotdog domain of Rv0216. Both COG 2030 and Pfam 0175 include seven other *M. tuberculosis* proteins. These proteins vary widely in size, from 142 to 3069 amino acids, and in Pfam annotation (Table 4). Five of these proteins contain a hydratase 2 motif as described by Qin et al. (2000). In light of our crystallographic study on Rv0216, we have re-evaluated this group of proteins. As part of our standard bioinformatics evaluation of *M. tuberculosis* targets, we have made use of fold-recognition evaluations. Before solving the structure of Rv0216, the 3D-PSSM server (<http://www.sbg.bio.ic.ac.uk/~3dpssm/>) of Sternberg and coworkers (Kelley et al. 2000) indicated that the structure might contain a hotdog-domain fold, most similar to PDB code 1IQ6. We have, therefore, also evaluated the sequences of the other seven *M. tuberculosis* proteins with the same fold recognition server (Table 4). Our analysis indicates that at least five of them (Rv2499c, Rv0130, Rv0636, Rv3389c, and Rv3538) are likely to be SHD dimers or DHDs that are related in structure to Rv0216. Two more (Rv0241c and Rv2524c) are likely to contain at least one SHD domain, but may be DHDs.

So far, we have been unable to identify another DHD Rv0216-like protein in the *M. tuberculosis* genome. As shown in the sequence alignments in Figure 4, however, Rv2499c appears to be an Rv0216-like SHD structure and hence likely to form a homodimer. This protein

matches the sequence profile of Rv0216-like proteins in both hotdog domains, in particular the conserved residues in the proposed active site. Two additional proteins, Rv0130 and Rv0636, are more similar to the SHD bacterial R-specific 2-enoyl-CoA hydratase (Fig. 4; Table 4). These similarities include the active site, the hydrophobic patch lining the entrance of the substrate-binding pocket (L35, F47, I51, F83, 1IQ6 numbering), and the $\beta 2$ - $\beta 3$ crossover loop (F83-D91). Although the hydratase 2 motif defined by Qin et al. (2000) is not present in Rv0130, only two residues differ (Fig. 4). In regions where the existing SHD crystal structures show the largest variation, namely, the connection between the hotdog helix and the $\beta 2$ -strand, there is poor sequence identity. In the final two β -strands, the two known structures show a preference for small, branched side chains pointing inward, which can be recognized in the *M. tuberculosis* sequences, but a correct alignment is difficult to achieve.

Two of the seven *M. tuberculosis* proteins, Rv3389c and Rv3538, are likely to be incorrectly classified as “possible” or “probable” dehydrogenases. This seems to have arisen because of sequence similarity to a multifunctional enzyme, where the second function is a dehydrogenase activity. We suggest that Rv3389c and Rv3538 are DHD enzymes, based on their structure-based sequence alignments (Fig. 4) and PSSM scores (Table 4). The second hotdog domains are very well conserved and clearly indicate R-specific enoyl hydratase-like active sites. The first hotdog domains show much lower overall sequence similarity but have local identity patterns that map to important areas of SHD structure. Residues L16–Y17 (1PN2 hydratase 2 numbering), on the first helix of the lid, close to the internal dyad axis, are positioned where there is a hydrophobic dependency as part of the dimer formation, while G22 terminates the helix at a Schellman turn (Schellman 1980). The P⁴²-TF sequence is the start of the hotdog helix with F44 close to the side-chain of Y17. The next sequence-conserved region, L⁷²-HGE, is located at the start of the $\beta 2$ in the vicinity of the active site, while the P⁸⁴-X-P peptide is at the important $\beta 2$ - $\beta 3$ crossover loop. The residues at the end of $\beta 3$ and the reverse turn contain another conserved region, P¹⁰²-KG, where the lysine residue interacts with one of the phosphates of the substrate in the eukaryotic DHD hydratase (Koski et al. 2004). Finally, R134 is the last residue in the final β -strand of the first hotdog domain, and is the beginning of a remarkably well-conserved feature in the other proteins (RGxGGFGG). This begins the linker to the second hotdog domain, and starts at a highly conserved proline residue, P162.

Two other proteins, Rv0241c and Rv2524c, show indications of DHD hydratase-like structures. Rv0241c is large enough to be a DHD, its C-terminal half shows

Table 4. *Hoidog-related proteins in M. tuberculosis*

Rv number active site residues	Pfam annotation	Size (aa)	Our classification ^a	Essential, in vivo (Sasseti and Rubin 2003)	Essential, in vitro (Sasseti et al. 2003)	Present in <i>M. leprae</i>	PSSM scores to SDH/DHD structures			
							Domain	1IQ6	1Q6W	IPN2
Rv0216 N232, H237	Not present	337	DHD, Rv0216-like	Yes	No	Yes	N ^b	2.7×10^{-3}	6.0×10^{-14}	2.9×10^{-3}
Rv2499c	Possible oxidase	185	SHD,	N/A ^d	No	Yes	C ^c	2.7×10^{-3}	3.2×10^{-7}	2.8×10^{-2}
N66, H71	Regulatory-related protein		Rv0216-like					2.7×10^{-3}	1.0×10^{-9}	5.1×10^{-3}
Rv0130 D40, H45	Hypothetical protein	151	SHD	No	No	No		2.7×10^{-3}	3.2×10^{-7}	2.9×10^{-3}
Rv0636 D36, H41	Hypothetical protein	142	SHD	Yes	Yes	Yes		2.7×10^{-3}	1.7×10^{-8}	2.9×10^{-3}
Rv3389c	Possible dehydrogenase	290	DHD	No	No	Yes	N	4.4×10^{-2}	8.0×10^{-3}	1.3×10^{-8}
D186, H191	Probable dehydrogenase	286	DHD	N/A	Yes	No	C	2.7×10^{-3}	4.5×10^{-9}	2.7×10^{-3}
Rv3538	Probable dehydrogenase	280	Probable DHD	N/A	No	Yes	N	8.5×10^{-3}	4.1×10^{-2}	4.6×10^{-6}
D189, H194	Hypothetical protein	3069	Possible DHD	N/A	Yes	Yes	C	3.3×10^{-9}	2.7×10^{-3}	2.7×10^{-3}
Rv0241c	Probable fatty acid synthetase							0.21	—	—
D190, H195							C	1.0×10^{-9}	2.7×10^{-3}	2.7×10^{-3}
Rv2524c							N	4.9	—	—
D1229, H1234							C	2.9×10^{-7}	2.7×10^{-3}	9.8×10^{-3}

^a Except for Rv0216 and Rv2499c, all entries have active sites that are compatible with hydratase/dehydratase activities.^b (N) N-terminal domain.^c (C) C-terminal domain.^d (N/A) Not applicable (not reported in Sasseti and Rubin 2003).

a good PSSM score to the bacterial R-specific enoyl hydratase, and its sequence contains a hydratase 2 motif. Its N-terminal region shows a poorer but still significant score to the same enzyme. Similarly, a 149-amino-acid segment (residues 1197–1345) in the much larger Rv2524c shows a high structural similarity to the bacterial R-specific enoyl hydratase and also contains a hydratase 2 motif. An equally long portion of the preceding sequence (residues 1038–1196) shows a much poorer match to the same hydratase. These results may indicate the presence of a DHD organization in both proteins but where the first hotdog domain is less similar to our search templates. Rv2524c is the only protein in Table 4 for which an enzymatic activity has been suggested. This is the type I fatty acid synthase (FAS-I) of *M. tuberculosis*, the multifunctional enzyme found in eukaryotes and advanced prokaryotes. The separate enzymatic activities have been identified and localized to functional domains. A dehydrase-related catalytic role for a histidine residue (H878) in the rat FAS-I system has been suggested (Joshi and Smith 1993). The identification of this residue was made possible by weak sequence similarity to the *E. coli* FabA enzyme before the crystal structure determination by Smith and coworkers (Leesong et al. 1996). In the sequence alignments of Joshi and Smith (1993) the aspartyl residue in the active site of FabA is not present. Our own alignments show the presence of an equivalent glutamyl residue in some sequences (including the mycoerotic acid synthase of *M. tuberculosis*, data not shown) but not in the rat and human FAS-I enzymes, for example. It should be remembered, however, that FabA is a homodimer and that the proposed catalytic residues are derived from different SHD domains. We have located a SHD in the human FAS-I sequence using PSSM, and after optimizing the sequence range to residues 820–968, we get significant scores (4.1×10^{-1} to 1MKA and 1.6×10^{-1} to 1J1Y) with H⁸⁷⁸ at the active site. We have been unable to locate a SHD domain before or after this region of the sequence, however.

Initially, the dehydrase activity of the *M. tuberculosis* FAS-I was suggested to reside in the region centered on H963, and was also based on sequence analysis (Fernandes and Kolattukudy 1996). Rv2524c was, however, later identified as containing the hydratase 2 motif by Qin et al. (2000). Since the β -hydroxyacyl dehydrase activity of this enzyme is the reverse of the reaction carried out by the R-specific enoyl hydratases, we can, with some confidence, assign D¹²²⁹ and H¹²³⁴ to be the catalytic residues for the dehydrase activity of Rv2524c. Interestingly, the type I enzyme is thought to be a homodimer (Smith et al. 2003), and although we suggest that the dehydrase activity resides in a DHD domain, an alternative hypothesis would be a dimer of

SHD domains. However, the FAS-I dimer can be dissociated by lowering the ionic strength and temperature while retaining dehydrase activity on model substrates (Kumar et al. 1970). Such activity is not feasible in a solitary SHD, but would be possible in a DHD structure. Although the catalytic machinery of all members of the FAS-I family does not seem to be conserved, it is possible that an underlying DHD structure may be a common feature. However, we do not rule out the possibility of homodimer formation, despite experimental evidence to the contrary. In the *M. tuberculosis* enzyme, the malonyl/palmitoyl transferase catalytic domain follows the dehydrase-like domain, separated by a short linker.

M. tuberculosis also contains the FAS-II system found in bacteria and plants where the enzymatic activities exist in separate polypeptides (Bloch 1977; Smith et al. 2003; Takayama et al. 2005). The gene encoding the dehydrase activity in this system has not yet been identified. In *E. coli*, on the other hand, two enzymes, FabA and FabZ, have this activity with overlapping chain length specificities. In their review, Takayama et al. (2005) identify a sequence similarity between FabZ and Rv0098 from *M. tuberculosis*. The latter protein has been identified by Sasseti and Rubin (2003) as being essential for *M. tuberculosis* survival during infection in an in vivo assay, is conserved in other mycobacteria (including *M. leprae*), and is, therefore, a good candidate for the “missing” β -hydroxyacyl-ACP dehydrase. Our PSSM analysis of the FabZ sequences from *E. coli* and *S. pneumoniae* indicates a structural relationship to the FabA of Leesong et al. (1996), with scores of 4.0×10^{-8} and 6.2×10^{-8} to 1MKA, but this relationship is not apparent in the Rv0098 sequence (no SHD or DHD hits in the top scores). Although this may merely indicate a divergence in the sequence so that the underlying structural similarity is no longer apparent, Table 4 suggests that a number of other gene products have the potential for a dehydrase active site. In particular, Rv0636 has been shown to be important in both the in vivo and in vitro studies (Sasseti and Rubin 2003; Sasseti et al. 2003) and is, therefore, an equally good candidate for the missing β -hydroxyacyl-ACP dehydrase in *M. tuberculosis*.

To elucidate the function of the *M. tuberculosis* proteins in Table 4 will require detailed biochemical, genetic, and structural analyses. As a start, we have expressed and purified Rv2499c, Rv0130, and Rv0636 (A.S. Covarrubias, T. Bergfors, T. Unge, and T.A. Jones, unpubl.) and investigated their hydratase activity. Only Rv0130 has shown activity in our assays using 2-butenoyl CoA (crotonyl-CoA) as a substrate. A protein concentration of 5 ng/mL gave a steady decrease in A₂₆₃ for >3 min. It must be emphasized that the known SHD and DHD hydratases/dehydrases show a variation in substrate specificity due to

differences in the structure of their active sites. The precise nature of the active sites of the proposed *M. tuberculosis* SHD/DHD's will require detailed crystallographic analyses and we hope that these structures will provide new insights into potential substrates.

Discussion

The World Health Organization estimates that two million people die each year from TB and that one-third of the earth's population is affected. New drugs are urgently required to combat this epidemic and, in particular, to reduce the treatment time. Structure-based drug design is most usually applied to known targets, preferably in well-understood, critical biochemical pathways. However, the availability of the complete sequence of the *M. tuberculosis* genome means that we are now able to analyze gene products of unknown function that are highly conserved in mycobacteria and have been identified as playing an important role for the survival of the organism. Sasseti et al. (2003) and Sasseti and Rubin (2003) have applied high-density mutagenesis to study the genetic requirements for mycobacterial survival during infection and growth, and we have used their results to guide target selection for some of our structural studies. Rv0216 is reported by Sasseti and Rubin (2003) as being one of the only 194 genes that are required for mycobacterial growth in vivo, therefore making it a potential drug target. The importance of this gene product is further strengthened by its conservation in the degenerate genome of *M. leprae*, the causative agent of leprosy. A hint of its biological function comes from its position in the *M. tuberculosis* genome, where it is adjacent to genes involved in fatty acid metabolism and cell wall processes. Rv0216, however, codes in the reverse direction.

Our structural analysis shows that the Rv0216 gene product has a double hotdog fold, characteristic of a class of enzymes that have thiol esters as substrates. Rv0216-like sequences show conservation around a potential active site region of this class of enzyme, but this is restricted to only one of the hotdog domains. Although Rv0216 shows strong evidence for a gene duplication event, we suggest that only one functional active site has been preserved. We believe this is localized to a variation of the hydratase 2 loop (Koski et al. 2004) in the C-terminal SHD. In Rv0216-like proteins, the proposed catalytic apparatus of the hydratases is modified so that although the histidyl residue is conserved, an asparaginyl residue replaces the catalytic aspartyl residue. Indeed, Rv0216 shows no activity on crotonyl CoA and its natural substrate remains to be identified. However, our new structure helps us evaluate seven *M. tuberculosis* gene products related to Rv0216. One of

these is likely to be a DHD containing the dehydrase activity of the FAS-I, and five others are likely to be SHD/DHD hydratases or dehydrases. We have expressed two of these, and one shows hydratase activity on crotonyl CoA substrate. We suggest that the other, Rv0636, is a good candidate for the "missing" β -hydroxyacyl-ACP dehydrase of FAS-II in *M. tuberculosis*. Interestingly, of the eight gene products that may contain SHD/DHD domains, four were found to be essential for survival in one of the high-density mutagenesis experiments of Rubin and coworkers (Sasseti and Rubin 2003; Sasseti et al. 2003), and six are present in *M. leprae*. This family may, therefore, have the potential for further structure-based drug design studies.

Materials and methods

Cloning, expression, and purification

The open reading frame coding for *rv0216* (nucleotides 258,913–259,926) was amplified by PCR from total DNA of *M. tuberculosis*, strain H37Rv, using the primer pair 5'-GATGGTATTGACCCCTGTCCGATAG (forward) and 5'-GCCCTTAGTAACCGAACCTAG (reverse). In a second PCR reaction, an N-terminal six-histidine tag was introduced using the forward primer 5'-ATGGCTCATCATCATCATCATCATGGTGCTAGCGGGTATGGGGGCATC. The high-fidelity polymerase Pfu Turbo (Stratagene) was used for the amplifications. After addition of a 3' A-overhang by incubation with Taq polymerase (Roche), 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* Rosetta (DE3) with a pLacI plasmid (Novagen) or in *E. coli* BL21 (DE3) (Invitrogen). The cells were cultured in LB media at 37°C. At OD₆₀₀ = 0.5–1.0 the temperature was lowered to 24°C and expression of the target gene induced with IPTG (50 mg/L) for 3 h. After harvesting by centrifugation, the cell pellet was resuspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol, 0.5% (v/v) Triton X-100 [pH 8.0]) with 1 mg/mL lysozyme, 1 mM PMSF, 0.01 mg/mL RNase A, and 0.02 mg/mL DNase I and lysed. The soluble fraction was incubated with 1 mL pre-equilibrated Ni-NTA agarose (Qiagen) slurry for 1 h at 4°C, then packed into a column. After washing with 20 column volumes of buffer (20 mM imidazole, 50 mM NaH₂PO₄, 300 mM NaCl, 10% (v/v) glycerol [pH 8.0]) the protein was eluted with four 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 (Amersham Biosciences) equilibrated in 20 mM Tris [pH 7.5], 150 mM NaCl, and 10% glycerol. The protein eluted in two fractions as a monomer and a dimer. These two fractions were concentrated separately to 4–5 mg/mL.

SeMet was introduced into the protein by metabolic inhibition (van Duyne et al. 1993). The SeMet protein was expressed in Rosetta (DE3) pLacI cells or 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 β -mercaptoethanol in all buffers to prevent SeMet oxidation. Mass spectrometry analyses of the native and SeMet proteins gave molecular weights of 36.69 and 36.80 kDa, respectively. The difference (0.11 kDa) indicates that 2.4, or 79%, of the three methionines have been substituted for SeMets.

Crystallization

All crystallization trials were carried out at 20°C by the vapor diffusion method (McPherson 1999). Initial crystallization conditions for the native protein were screened using Crystal Screen HT and a selection of additives from Additive Screens 1–3 (Hampton Research). Native and SeMet crystals of Rv0216 were obtained in droplets consisting of 1 μ L protein (4–5 mg/mL) in 20 mM Tris (pH 7.5), 150 mM NaCl, 10% glycerol, and 1 μ L reservoir (100 mM MES [pH 6.5], 1.3 M ammonium sulfate, 4% PEG400, 10 mM β -mercaptoethanol). The crystals grew to a size of 0.2 mm $\times$ 0.2 mm $\times$ 0.4 mm within 4 wk. Monomer and dimer fractions of the protein gave similar results in the crystallizations.

Data collection

Crystals of seleno-substituted Rv0216 were flash-frozen in liquid nitrogen using the mother liquor with 25% glycerol added as a cryo-protectant. Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) beamline ID29 from a single SeMet crystal at wavelengths above, below, and at the Se-absorption edge to resolutions of 2.7, 2.5, and 2.8 Å, respectively. An additional high-resolution 1.9 Å data set was collected close to the selenium peak. Indexing, integration, and scaling of the data were carried out using the HKL suite of computer programs (Otwinowski and Minor 1997). The SeMet crystal form was found to belong to space group P6₃22 with unit cell parameters $a = 77.9$ Å, $b = 77.9$ Å, and $c = 178.5$ Å. Assuming one monomer in the asymmetric unit gives a V_m of 2.2 Å³ Da⁻¹ (Matthews 1968). A complete native data set was later collected to a resolution of 1.9 Å. Data collection statistics are given in Table 1.

Structure determination and refinement

The structure was solved by MAD (Hendrickson 1991). The positions of two selenium atoms were found with the RSPS program (Knight 2000) using the Se-peak diffraction data set. A third selenium site was located by difference Fourier analysis after phasing with MLPHARE (Otwinowski 1991). Initial MAD phasing was carried out with SHARP (de la Fortelle and Bricogne 1997) and then extended to a resolution of 1.9 Å using SAD phasing. Electron density maps were subjected to solvent flattening and histogram matching using DM (Cowtan and Main 1998). The semi-interactive main-chain tracing and sequence decoration tools of O (Jones 2004) were used to build 330 of the 337 amino acids in the asymmetric unit. This model was improved by alternating cycles of refinement with REFMAC5 (Murshudov 1997) and interactive rebuilding with O (Jones et al. 1991). Water molecules were added with ARP/wARP (Perrakis 1997). One residue, Y94, in the putative active site shows relatively poor electron density and may interact with a sulfate group. However, since the density is not perfectly clear, it has been modeled as solvent. The final phasing and refinement statistics are given in Table 2. Co-

ordinates and structure factors have been deposited at the PDB with access code 2BI0.

Enoyl-CoA hydratase activity assay

Hydration of crotonyl-CoA (2-butenoyl-CoA) was used as an assay for enoyl-CoA hydratase activity (Moskowitz and Merrick 1969; Fukui et al. 1998). The reaction was carried out in a total volume of 1 mL. Enzyme (Rv0216, Rv0130, Rv0636, or Rv2499c) was added to 25 μ M of crotonyl-CoA (Sigma) in 50 mM Tris (pH 8.5), and the decrease in absorbance at 263 nm was measured at 25°C or 30°C.

Electronic supplemental material

Multiple alignment of the 45 different Rv0216-like sequences found, 1IQ6 and 1PN2 included for reference.

Acknowledgments

Our tuberculosis-related research is supported by grants from the Swedish Foundation for Strategic Research (SSF), the Swedish Natural Science Research Council, and European Commission programs SPINE (QLG2-CT-2002-00988) and X-TB (QLRT-2000-02018). Total DNA of *M. tuberculosis*, strain H37Rv, was a gift from Dr. Stewart Cole, Institut Pasteur.

References

- Bahnson, B.J., Anderson, V.E., and Petsko, G.A. 2002. Structural mechanism of enoyl-CoA hydratase: Three atoms from a single water are added in either an Elcb stepwise or concerted fashion. *Biochemistry* **41**: 2621–2629.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. 2004. The Pfam protein families database. *Nucleic Acids Res.* **32**: D138–D141.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.
- Bloch, K. 1971. β -Hydroxydecanoyl thioester dehydrase. In *The enzymes*, 3rd ed. (ed. P.D. Boyer), pp. 441–464. Academic Press, New York.
- . 1977. Control mechanisms for fatty acid synthesis in *Mycobacterium smegmatis*. *Adv. Enzymol. Relat. Areas Mol. Biol.* **45**: 1–84.
- Camus, J.C., Pryor, M.J., Medigue, C., and Cole, S.T. 2002. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**: 2967–2973.
- Cole, S.T. 2002. Comparative mycobacterial genomics as a tool for drug target and antigen discovery. *Eur. Respir. J. Suppl.* **36**: 78–86.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E., et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537–544.
- Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., et al. 2001. Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007–1011.
- 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.
- Engh, R. and Huber, R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**: 392–400.

- Fernandes, N.D. and Kolattukudy, P.E. 1996. Cloning, sequencing and characterization of a fatty acid synthase-encoding gene from *Mycobacterium tuberculosis* var. bovis BCG. *Gene* **170**: 95–99.
- 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.
- 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.
- Jones, T.A. 2004. Interactive electron-density map interpretation: From INTER to O. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2115–2125.
- 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.
- Joshi, A.K. and Smith, S. 1993. Construction, expression, and characterization of a mutated animal fatty acid synthase deficient in the dehydratase function. *J. Biol. Chem.* **268**: 22508–22513.
- Kelley, L.A., MacCallum, R.M., and Sternberg, M.J. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**: 499–520.
- Kleywegt, G.J. and Jones, T.A. 1996. ϕ/ψ -chology: Ramachandran revisited. *Structure* **4**: 1395–1400.
- Knight, S.D. 2000. RSPS version 4.0: A semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Crystallogr. D Biol. Crystallogr.* **56**: 42–47.
- Koski, M.K., Haapalainen, A.M., Hiltunen, J.K., and Glumoff, T. 2004. A two-domain structure of one subunit explains unique features of eukaryotic hydratase 2. *J. Biol. Chem.* **279**: 24666–24672.
- Kumar, S., Dorsey, J.A., Muesing, R.A., and Porter, J.W. 1970. Comparative studies of the pigeon liver fatty acid synthetase complex and its subunits. Kinetics of partial reactions and the number of binding sites for acetyl and malonyl groups. *J. Biol. Chem.* **245**: 4732–4744.
- 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.
- Lopez, R., Silventoinen, V., Robinson, S., Kibria, A., and Gish, W. 2003. WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.* **31**: 3795–3798.
- Marmiesse, M., Brodin, P., Buchrieser, C., Gutierrez, C., Simoes, N., Vincent, V., Glaser, P., Cole, S.T., and Brosch, R. 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**: 483–496.
- Matthews, B.W. 1968. Solvent content of protein crystals. *J. Mol. Biol.* **33**: 491–497.
- McPherson, A. 1999. *Crystallization of biological macromolecules*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 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 hydratase 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.
- 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.
- Qin, Y.M., Haapalainen, A.M., Kilpelainen, 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.
- 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.
- Schellman, C. 1980. The α_L conformation at the ends of helices. In *Protein folding* (ed. R. Jaenicke), pp. 53–61. Elsevier/North-Holland, New York.
- Smith, S., Witkowski, A., and Joshi, A.K. 2003. Structural and functional organization of the animal fatty acid synthase. *Prog. Lipid Res.* **42**: 289–317.
- Stewart, G.R., Robertson, B.D., and Young, D.B. 2003. Tuberculosis: A problem with persistence. *Nat. Rev. Microbiol.* **1**: 97–105.
- 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.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* **278**: 631–637.
- Thoden, J.B., Holden, H.M., Zhuang, Z., and Dunaway-Mariano, D. 2002. X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase. *J. Biol. Chem.* **277**: 27468–27476.
- Thoden, J.B., Zhuang, Z., Dunaway-Mariano, D., and Holden, H.M. 2003. The structure of 4-hydroxybenzoyl-CoA thioesterase from *arthrobacter* sp. strain SU. *J. Biol. Chem.* **278**: 43709–43716.
- 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.