Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of FabH and factors affecting the distribution of branched fatty acids in *Micrococcus luteus*

Jose H. Pereira,[a,b]‡ Ee-Been Goh,[a,b]‡ Jay D. Keasling,[a,b,c] Harry R. Beller[a,d] and Paul D. Adams[a,b,c]*

[a]Joint BioEnergy Institute, Emeryville, CA 94608, USA, [b]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [c]Department of Bioengineering, University of California Berkeley, Berkeley, CA 94720, USA, and [d]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

‡ These authors contributed equally to this work.

Correspondence e-mail: pdadams@lbl.gov

Received 28 April 2012
Accepted 22 June 2012

PDB Reference: MlFabH, 4ewp

Micrococcus luteus is a Gram-positive bacterium that produces iso- and anteiso-branched alkenes by the head-to-head condensation of fatty-acid thioesters [coenzyme A (CoA) or acyl carrier protein (ACP)]; this activity is of interest for the production of advanced biofuels. In an effort to better understand the control of the formation of branched fatty acids in *M. luteus*, the structure of FabH (MlFabH) was determined. FabH, or β-ketoacyl-ACP synthase III, catalyzes the initial step of fatty-acid biosynthesis: the condensation of malonyl-ACP with an acyl-CoA. Analysis of the MlFabH structure provides insights into its substrate selectivity with regard to length and branching of the acyl-CoA. The most structurally divergent region of FabH is the L9 loop region located at the dimer interface, which is involved in the formation of the acyl-binding channel and thus limits the substrate-channel size. The residue Phe336, which is positioned near the catalytic triad, appears to play a major role in branched-substrate selectivity. In addition to structural studies of MlFabH, transcriptional studies of *M. luteus* were also performed, focusing on the increase in the ratio of anteiso:iso-branched alkenes that was observed during the transition from early to late stationary phase. Gene-expression microarray analysis identified two genes involved in leucine and isoleucine metabolism that may explain this transition.

1. Introduction

Fatty-acid biosynthesis in bacteria is primarily carried out by the type II synthase system, in which FabH (β-ketoacyl-ACP synthase III) catalyzes the key initiation step in acyl-chain formation: Claisen condensation of an acyl-CoA with malonyl-ACP to form a β-ketoacyl-ACP. FabH plays an important role in mediating the fatty-acid composition of the cell by virtue of its acyl-CoA substrate specificity (Choi, Heath *et al.*, 2000; Li *et al.*, 2005). The nature of the acyl-CoA primer used by FabH determines whether the resulting fatty acid will be straight-chained (if acetyl-CoA is the primer, as is characteristic of most Gram-negative bacteria) or methyl-branched (as is characteristic of most Gram-positive bacteria). For methyl-branched fatty acids, the primer is an acyl-CoA derived from a branched amino acid rather than acetyl-CoA. For anteiso-branched fatty acids FabH uses 2-methylbutyryl-CoA (derived from isoleucine) as a primer, whereas for iso-branched fatty acids FabH uses isovaleryl-CoA or isobutyryl-CoA (derived from leucine or valine, respectively).

In light of recent interest in biofuels, fatty-acid biosynthesis has garnered increased attention because these highly reduced aliphatic molecules are favorable precursors for compounds that could substitute for petroleum-derived fuels. For example, enzymes and pathways have recently been reported that convert fatty acids to biofuel-relevant compounds such as

alkanes (Schirmer *et al.*, 2010), alkenes (Beller *et al.*, 2010; Mendez-Perez *et al.*, 2011; Rude *et al.*, 2011), fatty-acid ethyl esters (Steen *et al.*, 2010) and methyl ketones (Goh *et al.*, 2012). As the ability to design molecules with desirable fuel properties is critical for the biofuels industry, it will be important to develop an improved understanding of how bacteria modulate fatty-acid chain length, saturation and branching, all of which are key factors that mediate fuel properties. In this study, we have solved the crystal structure of FabH from the Gram-positive bacterium *Micrococcus luteus*. This is of relevance to biofuels because *M. luteus* naturally synthesizes long-chain iso- and anteiso-branched alkenes *via* the condensation of two activated fatty-acid molecules (primarily iso- and anteiso-branched C_{15} fatty acids). Analysis of the MlFabH structure provides insights into the selectivity of MlFabH for branched acyl-CoA substrates. Furthermore, the MlFabH structure also shows an elongated acyl-binding channel which promotes the activity of the enzyme with long-chain acyl-CoA substrates. In addition to the structural studies of MlFabH, we also performed transcriptional studies of *M. luteus*. Gene-expression microarray analysis identified two genes involved in leucine and isoleucine biosynthesis that are downregulated and may be the key to the increased ratio of anteiso- to iso-branched alkenes observed as *M. luteus* transitions from early to late stationary phase (Beller *et al.*, 2010).

2. Materials and methods

2.1. Cloning, expression and purification of MlFabH

PCR amplification of the Mlut_09310 ORF was carried out as described previously (Beller *et al.*, 2010). The amplified fragment was digested with *Nde*I and *Eco*RI and cloned into the His-tag expression vector pSKB3. This resulted in the vector pEG272, which was transformed into *Escherichia coli* strain BL21 (DE3). The strain was grown at 303 K in 2 l autoinduction medium for 20–24 h before being harvested for protein purification. Cell lysis and protein purification were carried out as described previously (Petty, 1996) with a few modifications. Briefly, the harvested cell pellet was resuspended in 50 m*M* Tris–HCl pH 8.0 with 10% glycerol, 500 m*M* NaCl, 30 m*M* imidazole and 5 m*M* dithiothreitol. The cells were lysed by sonication followed by three freeze–thaw cycles at 193 K in the presence of 1 mg ml^{-1} lysozyme and 0.1% Triton X-100. After lysis, the cell lysates were incubated with 10 µg ml^{-1} DNase I and 10 m*M* $MgCl_2$ at room temperature for 10 min before being subjected to centrifugation at 27 000*g* for 15 min. The clarified supernatant was incubated with 1 ml Ni–NTA resin at 277 K for 1 h in a gently rocking clean glass gravity column. The column containing the resin was washed with 50 m*M* Tris–HCl pH 7.9 with 10% glycerol, 500 m*M* NaCl, 30 m*M* imidazole and 5 m*M* dithiothreitol. The proteins were eluted with the same buffer but with 200 m*M* imidazole. The eluted proteins were concentrated and exchanged into 100 m*M* potassium phosphate buffer pH 7.0 with 25 m*M* NaCl using Amicon centrifugal devices (Millipore). Protein concentrations were determined with the Bio-Rad Protein Assay reagent and the purity of the protein was assessed on a 8–16% gradient SDS–PAGE gel.

2.2. Crystallization of MlFabH

MlFabH was concentrated and dialyzed against 20 m*M* Tris–HCl buffer pH 7.5 containing 25 m*M* NaCl and 5 m*M* dithiothreitol. The final concentration of MlFabH used in crystallization trials was 11 mg ml^{-1}. The MlFabH protein was screened using the sparse-matrix method (Jancarik & Kim, 1991) with a Phoenix Robot (Art Robbins Instruments, Sunnyvale, California, USA) and the following crystallization screens: Crystal Screen, SaltRx, PEG/Ion, Index and PEGRx (Hampton Research, Aliso Viejo, California, USA). Crystals of MlFabH were found in 0.1 *M* bis-Tris propane pH 9.0, 10% polyethylene glycol 200, 18% polyethylene glycol 8000. MlFabH crystals were obtained after 2 d by the sitting-drop vapor-diffusion method with drops consisting of a mixture of 0.5 µl protein solution and 0.5 µl reservoir solution.

2.3. X-ray data collection and structure determination

A crystal of MlFabH was placed in a reservoir solution containing 10%(*v*/*v*) glycerol and then flash-cooled in liquid nitrogen. X-ray data sets for MlFabH were collected on the Berkeley Center for Structural Biology beamlines 8.2.1 and 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory (LBNL). The diffraction data were recorded using an ADSC Q315r detector. The data sets were processed using the program *HKL*-2000 (Otwinowski & Minor, 1997).

The crystal structure of MlFabH was determined by the molecular-replacement method with the program *Phaser* (McCoy *et al.*, 2007) within the *PHENIX* suite (Adams *et al.*, 2010), using as a search model the monomer structure of β-ketosynthase (ZhuH) from *Streptomyces* sp. R1128 (PDB entry 1mzj; Pan *et al.*, 2002), which shows 46% sequence identity to the target. The atomic positions of six copies of MlFabH obtained from molecular replacement and the resulting electron-density maps were used to build the MlFabH structure and initiate crystallographic refinement and model rebuilding. Structure refinement was performed using the *phenix.refine* program (Afonine *et al.*, 2009). Translation–libration–screw (TLS) refinement was used, with each protein chain assigned to a separate TLS group. Manual rebuilding using *Coot* (Emsley & Cowtan, 2004) and the addition of water molecules allowed construction of the final model. 5% of the data were randomly selected for cross-validation. The final model of MlFabH had an *R* factor of 17.1% and an $R_{\rm free}$ of 21.2%. Root-mean-square deviation differences from ideal geometries for bond lengths, angles and dihedrals were calculated using *PHENIX* (Adams *et al.*, 2010). The overall stereochemical quality of the final models for MlFabH was assessed by the program *MolProbity* (Chen *et al.*, 2010).

2.4. RNA extraction and expression microarray analysis

For transcriptional studies, *M. luteus* cultures were grown in 15 ml tryptic soy broth in a 30 ml glass tube as described above

Table 1
Statistics for data collection and structure refinement of MlFabH.

Data collection	
Wavelength (Å)	0.9774
Resolution range (Å)	50–2.20 (2.24–2.20)
Crystal-to-detector distance (mm)	160
Φ collected/ΔΦ (°)	180/1.0
Exposure time (s)	3
Data-collection temperature (K)	100
Data statistics	
Space group	*P*1
Unit-cell parameters (Å, °)	a = 80.14, b = 85.02, c = 96.91, α = 69.82, β = 71.37, γ = 85.66
Total reflections	197918
Unique reflections	111230
Multiplicity	1.9 (1.7)
Data completeness (%)	96.5 (84.8)
$\langle I/\sigma(I)\rangle$	8.93 (1.27)
$R_{\rm merge}$† (%)	0.072 (0.407)
Structure refinement	
Resolution range (Å)	50–2.20
R factor‡ (%)	17.1
$R_{\rm free}$§ (%)	21.2
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.101
Protein residues	349
Water molecules	397
Average isotropic *B* factors (Å^2)	
Protein atoms	35.1
Solvent atoms	36.8
Ramachandran plot	
Favored region (%)	95.8
Outliers (%)	0.0

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\sum_{hkl}$ denotes the sum over all reflections and $\sum_i$ is the sum over all equivalent and symmetry-related reflections. ‡ *R* factor = $\sum_{hkl} ||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$. § $R_{\rm free}$ is calculated as for the *R* factor but for 5% of the data that were not included during crystallographic refinement.

and harvested at 12 and 24 h into 2 ml ethanol solution containing 5% phenol to stop further transcription and to preserve RNA integrity. Cell cultures were spun down and the pellets were immediately frozen in liquid nitrogen and stored at 193 K until RNA extraction. Extraction and purification of RNA were carried out with a Qiagen RNeasy Mini kit and by treatment on-column with RNase-free DNase I (Qiagen). The concentration and integrity of RNA were determined using a Thermo Scientific Nanodrop ND-1000 spectrophotometer and an Agilent 2100 BioAnalyzer, respectively.

To perform microarray analyses, 10 µg of total RNA primed with 5 µg of random hexamers (Roche, Germany) were reverse-transcribed using the SuperScript Indirect cDNA labeling kit (Invitrogen). Alexa Fluor 555 dyes (Invitrogen) were then incorporated into amino-allyl-dUTP-labeled cDNA and the fluorescently labeled cDNA was purified using a QiaQuick PCR purification kit (Qiagen) and dried under vacuum (Vacufuge Speed Vac, Eppendorf). Labeled cDNA was hybridized to a four-plex NimbleGen (Roche) *M. luteus* expression microarray chip at 323 K for 20–24 h. The microarray chips, which were custom-designed based on the first-draft genome sequence of *M. luteus* NCTC 2665 (NCBI accession No. NC_012803), include triplicates of ten different 60-mer probes for each of the 2374 ORFs in the genome. After hybridization, microarray chips were scanned with a GenePix 4000B scanner and data were extracted using *NimbleScan*

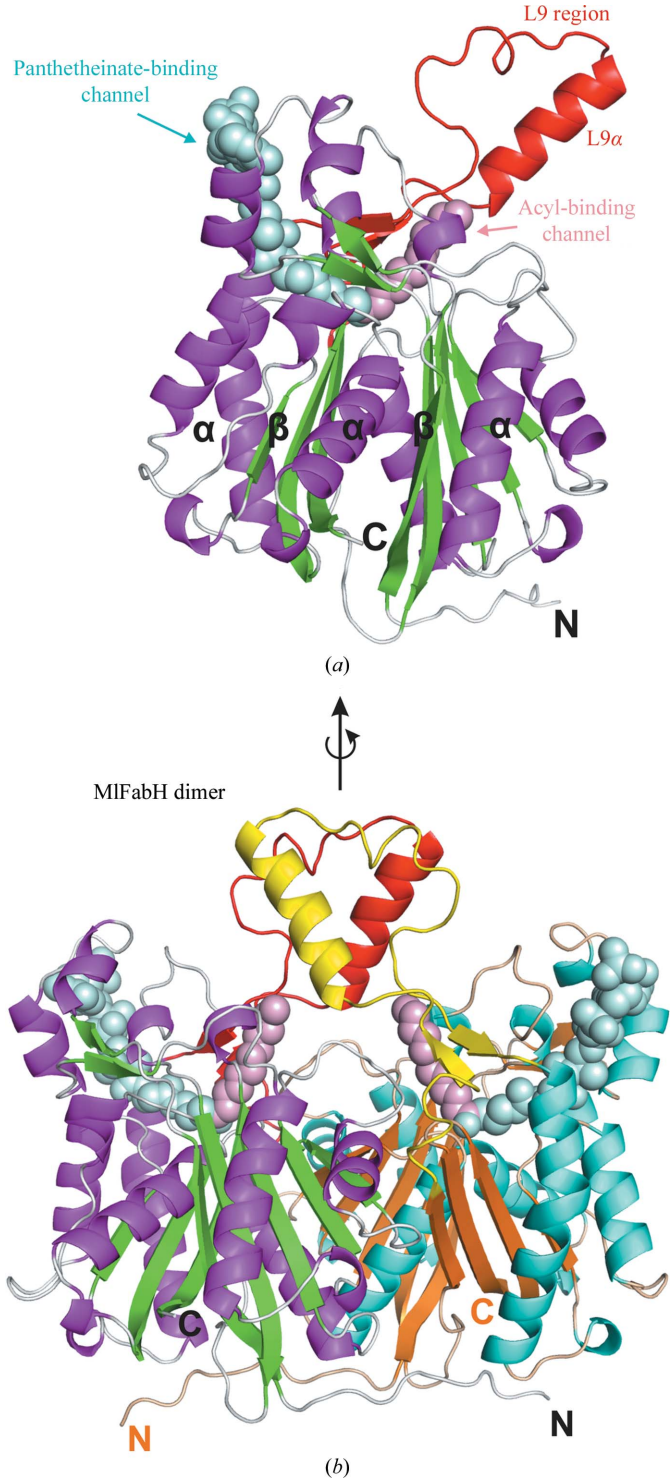

Figure 1
(*a*) The general MlFabH fold consists of a five-layered core based on three layers of α-helices interspersed with two layers of β-sheet (α-β-α-β-α). The substrate-binding site of MlFabH, which is represented as spheres, shows an L-shaped pocket composed of a solvent-accessible pantetheinate-binding channel (cyan) linked to a closed and elongated acyl-binding channel (pink). MlFabH possesses the largest L9 region among FabH structures deposited in the PDB and includes an antiparallel β-strand followed by a long α-helix (F9α; red). (*b*) The MlFabH structure has a dimer as the biological unit. The L9 region of both monomers (red and yellow) is located at the dimer interface and this region has been associated with the substrate specificity of the FabH enzyme class.

software. Array normalization was performed using the Robust Multiarray Average (RMA) technique (Irizarry *et al.*, 2003). The normalized expression values generated in RMA pair files were imported into *Excel* and statistical analyses were performed with the *Significance Analysis of Microarray* (*SAM*) add-on (Tusher *et al.*, 2001).

3. Results and discussion

3.1. General structure of MlFabH and the dimer interface

The MlFabH crystals diffracted to 2.2 Å resolution and belonged to the triclinic space group *P*1, with six copies of MlFabH (three dimers) per asymmetric unit. The statistics for X-ray data collection and structure refinement are summarized in Table 1. The FabH fold was first identified in the thiolase I structure, which catalyses the degradation of ketoacyl-CoA (Mathieu *et al.*, 1994, 1997). The MlFabH fold consists of a five-layered core based on three layers of α-helices interspersed with two layers of β-sheet (α-β-α-β-α; Fig. 1*a*). The MlFabH monomer shows structural internal pseudo-symmetry, despite there being no significant sequence identity between the two halves. The MlFabH structure, in keeping with all FabHs deposited in the Protein Data Bank (PDB), has a dimer as the biological unit (Fig. 1*b*). The substrate-binding site of MlFabH consists of an L-shaped pocket composed of a solvent-accessible pantetheinate-binding channel linked to a closed and elongated acyl-binding channel (Sachdeva *et al.*, 2008; Figs. 1*a* and 1*b*). The catalytic triad Cys123/His275/Asn306 is located at the bottom of the substrate-binding site connecting the two arms of the L.

3.2. The dimer interface of MlFabH and its role in supporting long acyl-CoA primer binding

The most obvious structural difference between MlFabH and all of the other FabHs deposited in the PDB (Davies *et al.*, 2000; Gajiwala *et al.*, 2009; Musayev *et al.*, 2005; Qiu *et al.*, 1999;

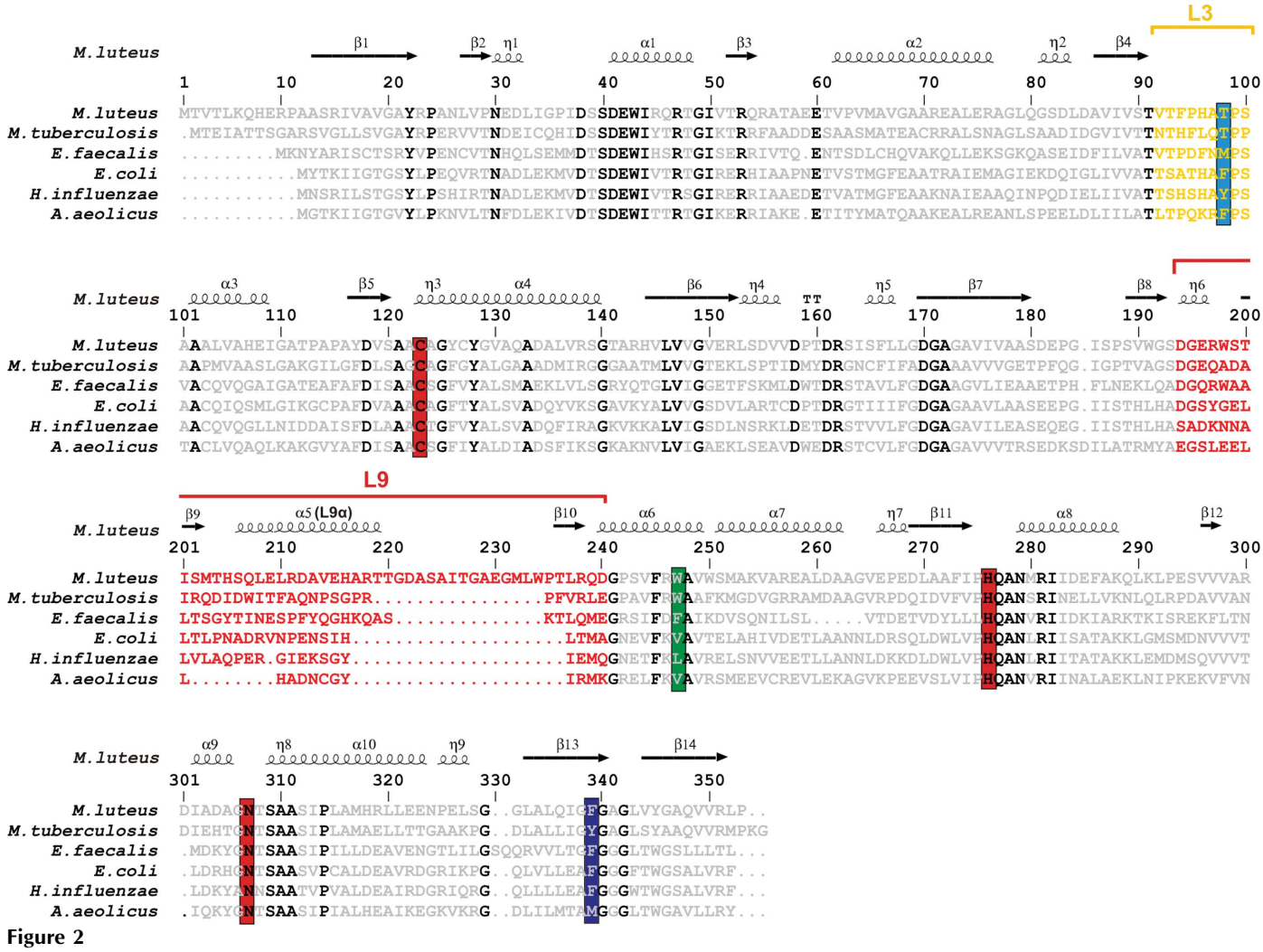

Figure 2
Sequence alignment of FabH from *M. luteus* (MlFabH), *M. tuberculosis* (MtFabH), *E. faecalis* (EfFabH), *E. coli* (EcFabH), *H. influenzae* (HiFabH) and *A. aeolicus* (AeFabH). The residues colored in yellow and red are the loop 3 (L3) and loop 9 (L9) regions, respectively. The catalytic residues Cys123, His275 and Asn306 are highlighted in red, the residue of FabH in the L3 region of the neighboring monomer of the dimer that is involved in acyl-binding channel formation is highlighted in cyan, the residue Phe336 (in MlFabH) that is important for branched-substrate specificity is highlighted in blue and residue Trp246 that directly impacts, through hydrophobic interactions, on the rotamer conformation of Phe336 is highlighted in green.

Sachdeva *et al.*, 2008) is the length of the L9 loop (the nomenclature is based on that of Davies *et al.*, 2000). MlFabH contains the longest L9 region (residues 193–239) of all of the known FabH structures and it contains an antiparallel β-strand followed by a long α-helix (F9α; Fig. 2). The L9 region is located at the MlFabH dimer interface and this loop has previously been associated with substrate specificity in ZhuH, an FabH homolog that is involved in polyketide synthesis (Pan *et al.*, 2002). Sequence analysis using *BLASTp* (Altschul *et al.*, 1997) against the previous FabH structures revealed that MlFabH has the highest similarity to FabH from *Mycobacterium tuberculosis* (MtFabH). A comparison between MlFabH and MtFabH in complex with lauroyl-coenzyme A (PDB entry 1u6s; Musayev *et al.*, 2005) shows that the L9 region interacts with the distal end of the acyl-binding channel (Fig. 1*a*). The distal end of the acyl-binding channel imposes an upper length limit on acyl-CoA substrates (Sachdeva *et al.*, 2008). Therefore, the high diversity in the L9 region supports a major role in determining the length of the channel for the acyl-binding site for this class of enzymes. Of particular interest is residue Trp233 in the L9 region of MlFabH, which may serve as a 'lid' to close the highly hydrophobic acyl-binding channel (Fig. 3*a*). A variant of MlFabH with the L9 region deleted (MlFabH-ΔL9) was created in order to understand its influence on the enzyme-binding affinity of different length acyl-CoA substrates. However, attempts to purify the MlFabH-ΔL9 variant indicated that it was expressed in inclusion bodies even after extensive optimization of the cell-growth conditions, suggesting that the L9 loop is important for correct folding of the protein. Furthermore, the MlFabH structure shows that the hydrophobic upper region of the acyl-binding site is stabilized by L9 residues from the neighboring monomer in the dimer. This suggests that the L9 region is not only important for determining length specificity for acyl-CoA substrates, but also for dimerization of MlFabH. In addition to the L9 region, there is a residue in MlFabH (Thr98) in loop L3 from the neighboring monomer that is involved in acyl-binding channel formation (Fig. 2). The FabH from *E. coli* (EcFabH) is most active with the acetyl-CoA substrate and is inactive with longer chain substrates. At the equivalent position to Thr98 in MlFabH, EcFabH has a large phenylalanine residue (Phe87), significantly reducing the size of the acyl-binding channel (Qiu *et al.*, 1999; Davies *et al.*, 2000; Gajiwala *et al.*, 2009). In contrast to EcFabH, MtFabH, like MlFabH, has a threonine

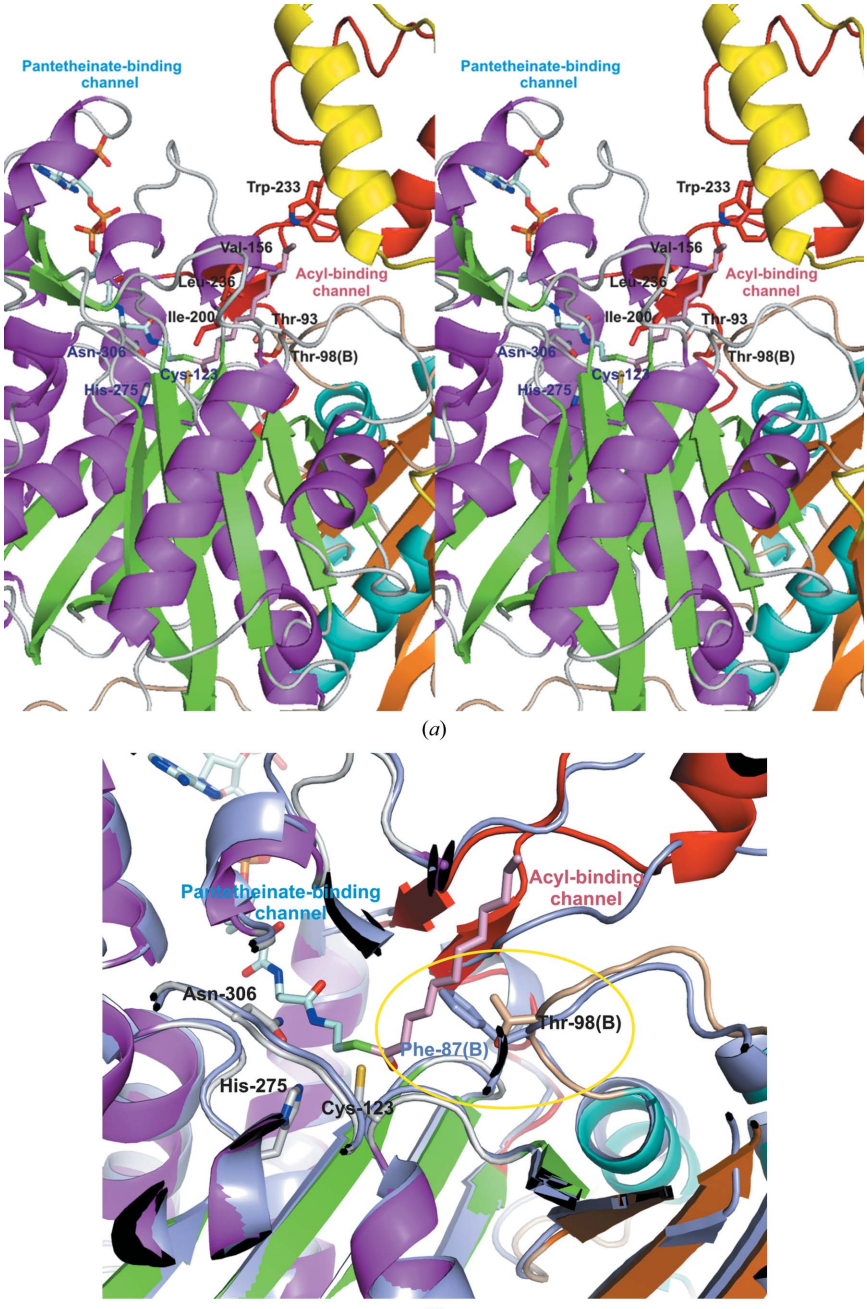

Figure 3
(*a*) Stereoview of the MlFabH active site in complex with lauroyl-coenzyme A. The similarities between the active sites of MlFabH and of MtFabH solved in complex with lauroyl-coenzyme A (Musayev *et al.*, 2005) permit inference of the residues that are involved in the selectivity for the acyl-CoA chain. Structural analysis of the model of the complex of MlFabH with lauroyl-coenzyme A shows that the L9 region interacts with the distal end of the acyl-binding channel of MlFabH. Therefore, the high diversity of the length and sequence of the L9 region that is observed in various FabHs plays a major role in determining the length of the acyl-CoA substrate for this class of enzymes. Of particular interest is residue Trp233 in the L9 region of MlFabH, which works as lid to close the highly hydrophobic end of the acyl-binding channel. (*b*) Structural comparison of the MlFabH–lauroyl-coenzyme A complex model and EcFabH (blue) shows that the Phe87 residue located in the L3 loop region in EcFabH constrains substrate specificity to acetyl-CoA, whereas the Thr98 residue at the same position in MlFabH permits the binding of a longer acyl-CoA substrate.

Table 2
Residues involved in branched-substrate selectivity in FabH.

Species	Gram classification	Key residue for branched-substrate selectivity	Rotomer conformation	Neighboring residue (position 1)	Neighboring residue (position 2)
M. luteus	+	Phe336	1	Trp246	Met251
E. faecalis	+	Phe312	1	Phe226	Val231
M. tuberculosis	n/a	Tyr304	1	Trp215	Met220
E. coli	–	Phe304	2	Val215	Leu220
H. influenzae	–	Phe303	2	Leu214	Leu219
A. aeolicus	–	Met296	2	Val207	Met212

residue (Thr87) in the L3 region. MtFabH also shows activity with long acyl-CoA substrates (Choi, Kremer *et al.*, 2000), supporting a key role for position 98 (MlFabH sequence numbering) in the formation of a long acyl-binding channel and allowing MlFabH to accept a long acyl-CoA substrate (Fig. 3*b*). This supports a previous structural comparison between MtFabH and EcFabH which drew a similar conclusion (Pan *et al.*, 2002): the large Phe87 side chain in EcFabH limits substrate specificity to acetyl-CoA, whereas the smaller residue at the same position in MtFabH permits the binding of a longer acyl-CoA substrate.

3.3. MlFabH substrate-binding site: specificity for branched acyl-CoA

The highly conserved triad of Cys/His/Asn residues in FabH is essential for catalytic activity (White *et al.*, 2005; Brown *et al.*, 2005). Site-directed mutagenesis studies have shown that the Cys residue is essential for condensation and transacylation activity, while the His and Asn residues are important for decarboxylation and condensation activity (Davies *et al.*, 2000). Although the exact mechanism by which FabH displays selectivity towards different primers is still under investigation, structural studies of a range of FabHs (from Gram-positive and Gram-negative bacteria) have provided valuable insights into the key factors that influence the substrate selectivity of FabH (Gajiwala *et al.*, 2009; Pan *et al.*, 2002).

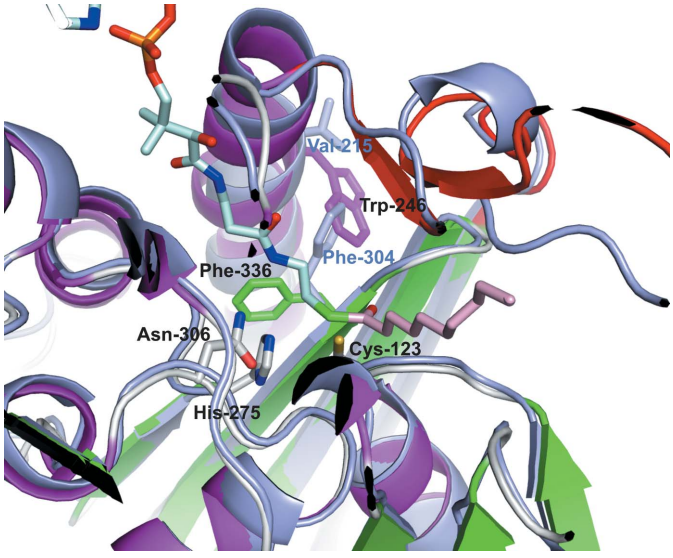

Figure 4
The positions of the phenylalanine residues equivalent to Phe336 in MlFabH found in Gram-negative bacteria occupy the binding site for the branched acyl-CoA group. Modeling analysis and the docking of a substrate (lauroyl-coenzyme A) into the active site of MlFabH shows that residue Phe336, which is located close to the catalytic triad, has a completely different orientation compared with residue Phe304 of the Gram-negative *E. coli* FabH. The position of Phe336 is influenced by the surrounding Trp246 residue *via* hydrophobic contacts. The residue equivalent to Trp246 in MlFabH in the Gram-negative EcFabH is Val215, which allows the conserved phenylalanine residue to assume a different rotamer orientation compared with MlFabH.

MlFabH uses acyl-CoA primers derived from branched amino acids (isoleucine, leucine and valine), as is characteristic of Gram-positive bacteria, rather than acetyl-CoA, which is used by most Gram-negative bacteria such as *E. coli*. MtFabH, which falls under neither classification, is capable of making iso- and anteiso-branched fatty acids, similar to MlFabH. The similarities between the active sites of MlFabH and MtFabH solved in complex with lauroyl-coenzyme A (Musayev *et al.*, 2005) permit inference of the residues that are involved in the selectivity for a branched acyl-CoA chain. Structural analysis of the active site of MlFabH shows that a conserved residue (Phe336) located close to the branched region of the substrate has a completely different orientation compared with the Gram-negative FabH structures from *E. coli* (EcFabH; Musayev *et al.*, 2005; Gajiwala *et al.*, 2009), *Haemophilus influenzae* (HiFabH; Gajiwala *et al.*, 2009) and *Aquifex aeolicus* (AaFabH; PDB entry 2ebd; RIKEN Structural Genomics/Proteomics Initiative, unpublished work).

The position of Phe336 is influenced by the neighboring residue Trp246 *via* a hydrophobic contact (Fig. 4). The equivalent residues to Trp246 in the Gram-negative EcFabH, HiFabH and AaFabH are Val215, Leu214 and Val207, respectively, which allow the conserved phenylalanine residues to assume different rotamer orientations compared with MlFabH (Fig. 4; Table 2). Gajiwala and coworkers explored the substrate specificity of FabH enzymes and identified residue Phe312 in FabH from the Gram-positive *Enterococcus faecalis* (EfFabH; equivalent to Phe336 of MlFabH) as a potential key determinant of substrate selectivity (Gajiwala *et al.*, 2009). However, Gajiwala and coworkers described the influence of two residues on the orientation of Phe312: Phe226 and Val231 (Table 2). Phe226 interacts *via* hydrophobic contacts with Phe312. Similar interactions are observed between Trp246 and Phe336 in MlFabH. The small side chain of Val231 allows Phe312 to assume the rotamer orientation found in the Gram-positive EfFabH. Comparison between MlFabH and EfFabH shows a relatively large side-chain residue (Met251) equivalent to Val231 of EfFabH, which does not influence the Phe336 rotomer position in MlFabH (Table 2). Furthermore, the FabH structure from the Gram-negative *A. aeolicus* (PDB entry 2ebd) also has a methionine (Met212) equivalent to Met251 of MlFabH, but the key residue for substrate specificity shows the same orientation as in the *E. coli* enzyme. We observe that *A. aeolicus* FabH has a

relatively small side-chain residue (Val207), similar to the residues of Gram-negative *E. coli* FabH (Val215) and *H. influenza* FabH (Leu214), compared with the Trp and Phe residues that are most commonly found in Gram-positive FabHs (Table 2). Therefore, the MlFabH structure demonstrates that the major influence on the rotamer orientation of the key conserved phenylalanine, which impacts the enzyme specificity for the branched acyl-CoA substrate, is the residue that is present at the equivalent position to Trp246 in MlFabH. Thus, hydrophobic residues such as Phe and Trp, which are characteristic of Gram-positive FabHs, at the position equivalent to 246 in MlFabH have a major influence on the conserved Phe (Phe336 in MlFabH), which is an important position for the branched-substrate specificity of FabH enzymes.

3.4. Gene-expression microarray analysis

We have previously observed that the ratio of anteiso-branched to iso-branched alkenes increases as *M. luteus* transitions from early to late stationary phase; these branched alkenes result from head-to-head condensation of iso- and anteiso-branched fatty acids (or their derivatives; Beller *et al.*, 2010). Gene-expression microarray analyses of *M. luteus* at 12 h (when anteiso-alkenes $\simeq$ iso-alkenes) and 24 h (when anteiso-alkenes >> iso-alkenes) identified two genes involved in leucine and isoleucine biosynthesis that may be related to the observed differences at early *versus* late stationary phase (Table 3). Mlut_10050, which encodes isoleucyl-tRNA synthetase, is downregulated at 24 h. This isoleucyl-tRNA synthetase is required for charging the tRNA with its cognate amino acid for protein synthesis. Mlut_12300, which encodes 2-isopropylmalate synthase, is also downregulated at 24 h. This enzyme converts 2-oxoisovalerate to 2-isopropylmalate, a key intermediate in leucine biosynthesis (Fig. 5). The down-regulation of both enzymes should serve to increase intracellular concentrations of isoleucine while decreasing those of leucine. Since isoleucine is the precursor of anteiso-branched primers and leucine is the precursor of odd iso-branched primers in fatty-acid biosynthesis, higher concentrations of isoleucine may account for the relative increase in anteiso-fatty-acid and anteiso-alkene production during late stationary

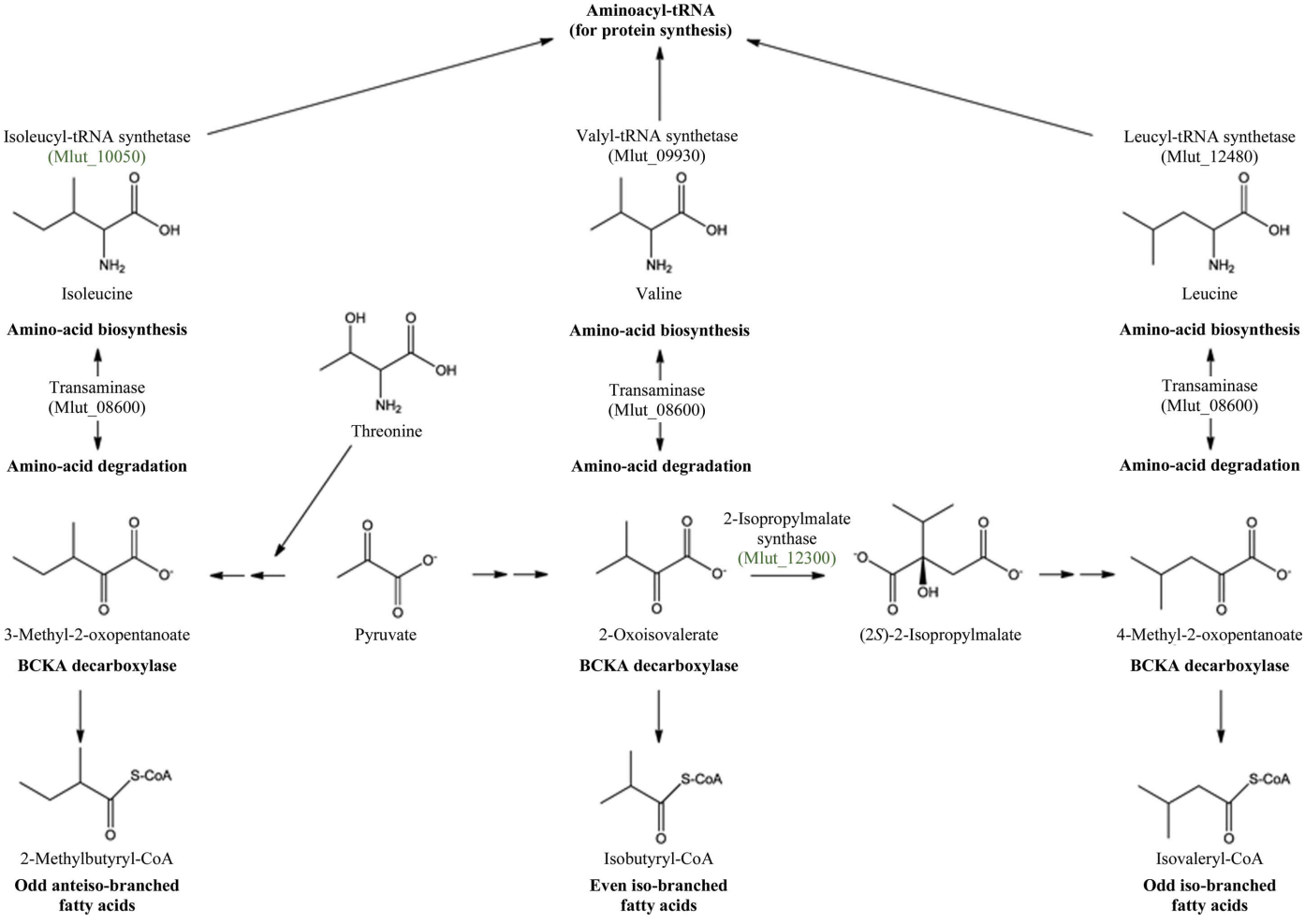

Figure 5
Methyl-branched acyl-CoA primers for FabH and their relationship to branched amino-acid precursors. Microarray analysis of *M. luteus* identified two genes, Mlut_10050 (annotated as isoleucyl-tRNA synthethase) and Mlut_12300 (annotated as 2-isopropylmalate synthase), as being significantly downregulated at 24 h compared with 12 h. The roles of these gene products in isoleucine, valine and leucine metabolism are highlighted in green.

Table 3
M. luteus genes associated with branched amino-acid biosynthesis/degradation and their relative transcriptional changes at 12 *versus* 24 h.

Genes with a less than twofold change were categorized as 'no change'.

Locus tag	Gene description	Change (fold)
Upregulated		
Mlut_06800	Pyruvate dehydrogenase (acetyl-transferring) E1 component, α subunit	2.4
Mlut_06810	Transketolase central region	3.1
No change		
Mlut_04360	Thiamine pyrophosphate protein TPP binding domain protein	1.0
Mlut_08500	Dihydroxy-acid dehydratase	0.7
Mlut_08550	Ketol-acid reductoisomerase	0.9
Mlut_08590	3-Isopropylmalate dehydrogenase	0.6
Mlut_08600	Branched-chain amino-acid aminotransferase	0.8
Mlut_08740	3-Isopropylmalate dehydratase, large subunit	0.9
Mlut_08750	3-Isopropylmalate dehydratase, small subunit	0.9
Mlut_09340	2-Oxo-acid dehydrogenase E1 subunit, homodimeric type	1.0
Mlut_09930	tRNA synthetase valyl/leucyl anticodon-binding	0.6
Mlut_12480	Leucyl-tRNA synthetase	1.0
Mlut_17800	Transketolase central region	1.0
Mlut_17810	Pyruvate dehydrogenase (acetyl-transferring)	1.0
Downregulated		
Mlut_08530	Acetolactate synthase, large subunit, biosynthetic type	2.5
Mlut_08540	Acetolactate synthase, small subunit	3.9
Mlut_10050	Isoleucyl-tRNA synthetase	3.7
Mlut_12300	2-Isopropylmalate synthase	2.2

phase. It is reasonable to assume that changes in alkene composition result from changes in the availability of fatty-acid primers. Previous studies have shown that the availability of primers *in vivo* can alter the fatty-acid composition of *Streptomyces glaucescens* (Han *et al.*, 1998). Other mechanisms that may affect the ratio of iso-alkenes to anteiso-alkenes at different growth stages include physiological responses [*e.g.* temperature changes may trigger responses that alter FabH substrate specificity (preference for anteiso *versus* iso primers; Singh *et al.*, 2009)].

4. Conclusions

In this study, we have addressed two different types of control of the biosynthesis of branched-chain fatty acids in *M. luteus*: (i) FabH structural characteristics that favor the binding of branched rather than straight-chain acyl-CoA primers and (ii) the differential expression of enzymes involved in isoleucine and leucine biosynthesis and metabolism that are likely to result in different primer concentrations for iso- and anteiso-branched fatty acids.

The amino-acid sequence of FabH is highly conserved among members of its family. However, structural comparisons of bacterial FabHs from Gram-negative *versus* Gram-positive bacteria have shown that subtle differences in the amino-acid sequences result in conformational differences that impact the three-dimensional shape of the substrate-binding pocket, thereby affecting substrate specificity. The crystal structure of MlFabH reveals a homodimer and a region, first identified in *E. coli* as the L9 loop (Davies *et al.*, 2000), which is essential for the dimer interactions and the formation of the upper acyl-binding channel (Pan *et al.*, 2002; Qiu *et al.*, 2005; Scarsdale *et al.*, 2001). In addition to the L9 loop region, residue Thr98 from the neighboring subunit, located in region L3, is also associated with the acyl-binding channel. A large side-chain residue at the equivalent position in *E. coli* FabH drastically reduces the substrate-binding pocket, allowing the enzyme to accept only small acyl-CoA substrates (*i.e.* acetyl-CoA). Additionally, the MlFabH structure supports the idea that the rotamer orientation of residue Phe336 is important for branched-substrate specificity and is directly affected by hydrophobic interactions with residue Trp246.

5. Data deposition

The atomic coordinates and structure factors of MlFabH have been deposited in the Protein Data Bank (PDB entry 4ewp). The microarray data (accession No. GSE38040) have been deposited in the GEO database hosted by NCBI.

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. We are grateful to the staff of the Berkeley Center for Structural Biology at the Advanced Light Source of Lawrence Berkeley National Laboratory. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Urzhumtsev, A. & Adams, P. D. (2009). *J. Appl. Cryst.* **42**, 607–615.
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Beller, H. R., Goh, E. B. & Keasling, J. D. (2010). *Appl. Environ. Microbiol.* **76**, 1212–1223.
Brown, A. K., Sridharan, S., Kremer, L., Lindenberg, S., Dover, L. G., Sacchettini, J. C. & Besra, G. S. (2005). *J. Biol. Chem.* **280**, 32539–32547.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Choi, K.-H., Heath, R. J. & Rock, C. O. (2000). *J. Bacteriol.* **182**, 365–370.
Choi, K.-H., Kremer, L., Besra, G. S. & Rock, C. O. (2000). *J. Biol. Chem.* **275**, 28201–28207.
Davies, C., Heath, R. J., White, S. W. & Rock, C. O. (2000). *Structure*, **8**, 185–195.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gajiwala, K. S., Margosiak, S., Lu, J., Cortez, J., Su, Y., Nie, Z. & Appelt, K. (2009). *FEBS Lett.* **583**, 2939–2946.

Goh, E. B., Baidoo, E. E., Keasling, J. D. & Beller, H. R. (2012). *Appl. Environ. Microbiol.* **78**, 70–80.
Han, L., Lobo, S. & Reynolds, K. A. (1998). *J. Bacteriol.* **180**, 4481–4486.
Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. & Speed, T. P. (2003). *Biostatistics*, **4**, 249–264.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Li, Y., Florova, G. & Reynolds, K. A. (2005). *J. Bacteriol.* **187**, 3795–3799.
Mathieu, M., Modis, Y., Zeelen, J. P., Engel, C. K., Abagyan, R. A., Ahlberg, A., Rasmussen, B., Lamzin, V. S., Kunau, W. H. & Wierenga, R. K. (1997). *J. Mol. Biol.* **273**, 714–728.
Mathieu, M., Zeelen, J. P., Pauptit, R. A., Erdmann, R., Kunau, W. H. & Wierenga, R. K. (1994). *Structure*, **2**, 797–808.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Mendez-Perez, D., Begemann, M. B. & Pfleger, B. F. (2011). *Appl. Environ. Microbiol.* **77**, 4264–4267.
Musayev, F., Sachdeva, S., Scarsdale, J. N., Reynolds, K. A. & Wright, H. T. (2005). *J. Mol. Biol.* **346**, 1313–1321.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pan, H., Tsai, S., Meadows, E. S., Miercke, L. J., Keatinge-Clay, A. T., O'Connell, J., Khosla, C. & Stroud, R. M. (2002). *Structure*, **10**, 1559–1568.
Petty, K. J. (1996). *Current Protocols in Molecular Biology*, edited by F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith & K. Struhl, p. 10.11.10–10.11.24. New York: John Wiley & Sons.
Qiu, X., Choudhry, A. E., Janson, C. A., Grooms, M., Daines, R. A., Lonsdale, J. T. & Khandekar, S. S. (2005). *Protein Sci.* **14**, 2087–2094.
Qiu, X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S., Lonsdale, J. & Abdel-Meguid, S. S. (1999). *J. Biol. Chem.* **274**, 36465–36471.
Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B. & Schirmer, A. (2011). *Appl. Environ. Microbiol.* **77**, 1718–1727.
Sachdeva, S., Musayev, F., Alhamadsheh, M. M., Scarsdale, J. N., Wright, H. T. & Reynolds, K. A. (2008). *Bioorg. Chem.* **36**, 85–90.
Scarsdale, J. N., Kazanina, G., He, X., Reynolds, K. A. & Wright, H. T. (2001). *J. Biol. Chem.* **276**, 20516–20522.
Schirmer, A., Rude, M. A., Li, X., Popova, E. & del Cardayre, S. B. (2010). *Science*, **329**, 559–562.
Singh, A. K., Zhang, Y.-M., Zhu, K., Subramanian, C., Li, Z., Jayaswal, R. K., Gatto, C., Rock, C. O. & Wilkinson, B. J. (2009). *FEMS Microbiol. Lett.* **301**, 188–192.
Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B. & Keasling, J. D. (2010). *Nature (London)*, **463**, 559–562.
Tusher, V. G., Tibshirani, R. & Chu, G. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
White, S. W., Zheng, J., Zhang, Y.-M. & Rock, C. O. (2005). *Annu. Rev. Biochem.* **74**, 791–831.