

**Qilong Mao,^a William L. Duax^{a,b}
and Timothy C. Umland^{a,b*}**

^aHauptman–Woodward Medical Research
Institute, 700 Ellicott Street, Buffalo, NY 14203,
USA, and ^bDepartment of Structural Biology,
School of Medicine and Biomedical Sciences,
University at Buffalo, Buffalo, NY, USA

Correspondence e-mail:
umland@hwi.buffalo.edu

Received 11 October 2006
Accepted 2 January 2007

Crystallization and X-ray diffraction analysis of the β -ketoacyl-acyl carrier protein reductase FabG from *Aquifex aeolicus* VF5

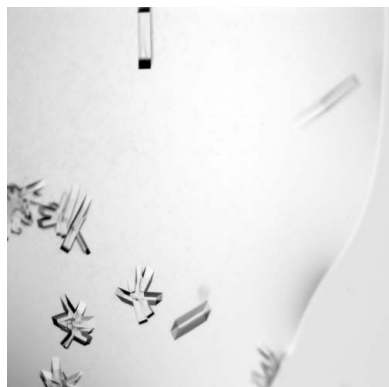
The gene product of *fabG* from *Aquifex aeolicus* has been heterologously expressed in *Escherichia coli*. Purification of the protein took place using anion-exchange and size-exclusion chromatography and the protein was then crystallized. Diffraction data were collected to a maximum resolution of 1.8 Å and the initial phases were determined by molecular replacement. The *A. aeolicus* FabG protein is a putative β -ketoacyl-acyl carrier protein reductase. Structure–function studies of this protein are being performed as part of a larger project investigating naturally occurring deviations from highly conserved residues within the short-chain oxidoreductase (SCOR) family.

1. Introduction

The short-chain oxidoreductase (SCOR) enzyme family includes at least 300 biochemically characterized enzymes found in bacteria, plants, insects and mammals (Oppermann *et al.*, 2003). This ancient family possesses members that catalyze a diversity of reactions (*i.e.* oxidation, reduction, epimerization and synthase reactions) using a wide range of substrates (*e.g.* steroids, sugars, prostaglandins, alcohols, acids, aromatics, xenobiotics *etc.*). SCOR proteins are known to participate in steroid metabolism, fatty-acid synthesis and clearance of toxins. Several are targets for drug development for the treatment of a variety of diseases (*e.g.* steroid-dependent cancers and bacterial infections). Over 13 500 putative members of the SCOR family have been identified through sequence similarity, but the vast majority have unknown substrates and functions and a significant number may have been incorrectly annotated. This large and diverse protein family does not contain a single amino-acid residue that is completely conserved in all SCOR members. Approximately 70% of the SCOR family belong to the TGxxxGIG subfamily, named after their signature sequence within the β 1 α 2 turn of the embedded Rossmann fold (unpublished data).

The diversity of the SCOR family makes it ideal for developing methods for identifying specific sets of residues (fingerprints) useful for predicting the biological properties (*e.g.* cofactor-specificity, substrate-specificity, oligomeric state *etc.*) of the large number of experimentally uncharacterized proteins that have been putatively assigned to the family (Duax *et al.*, 2003, 2005). This fingerprinting methodology employs knowledge of both the sequence and the experimentally determined three-dimensional structure. To further refine the methodology and to test the predictions, additional SCOR crystal structures are required (i) to provide a representative crystal structure to fingerprint families thus far assigned by sequence alone, (ii) to investigate the effect of engineered mutations or natural variations within a fingerprint and (iii) to identify the structural and functional roles of specific residues. As part of this rational proteomics project, the gene product of *fabG* (locus *aq_1716*; SWISS-PROT/TrEMBL accession No. 067610) from the thermophile *Aquifex aeolicus* VF5 was chosen for structural studies.

A. aeolicus FabG (Aae FabG) contains 248 amino acids, with a calculated weight of 26.9 kDa per monomer. It has been annotated based upon sequence as being a β -ketoacyl-acyl carrier protein reductase (β -k-ACPR; also referred to as 3-oxoacyl acyl-carrier-protein reductase; EC 1.1.1.100) and participates in the fatty-acid



biosynthesis pathway as a component of bacterial fatty-acid synthase II (FAS II; Rock & Jackowski, 2002). The NADP(H)-dependent FabG reduces the acyl carrier protein-linked β -ketoacyl group to a β -hydroxyacyl group during the elongation of the fatty acid's acyl chain. However, there are no published reports of the biochemical and biophysical characterization of Aae FabG. The crystal structures of several FAS II β -k-ACPRs have been determined (Fisher *et al.*, 2000; Price *et al.*, 2001, 2004; Wickramasinghe *et al.*, 2006; Yang *et al.*, 2002, 2003). These homologues have 42–51% amino-acid sequence identity to Aae FabG. Significantly, Aae FabG exhibits deviations at 10% (four out of 40) of the highly conserved SCOR TGxxxGIG fingerprint residues, with three of these variations resulting in a change in chemical properties (Asn65, Asp84, Thr159 and Val180 in Aae FabG; the fingerprint residues are Asp, Gly, Ala and Ile, respectively). In the previously determined crystal structures of FabG homologues, only the *M. tuberculosis* enzyme (Yang *et al.*, 2002, 2003) exhibits any of these fingerprint deviations: a variation corresponding to the Ile to Val180 variation observed in Aae FabG. These data indicate that the *A. aeolicus* enzyme may possess unique properties, perhaps as a consequence of the stability requirements of the thermophile. Thus, Aae FabG is an attractive structural target for use in further refinement of the SCOR fingerprint methodology.

2. Experimental methods

2.1. Cloning

The *fabG* gene from *A. aeolicus* was amplified by polymerase chain reaction using genomic DNA as a template. The restriction-enzyme sites *NdeI* and *XhoI* were designed into the forward primer and reverse primer, respectively. The forward primer sequence was 5'-GCGCATATGGAAATAAACTCCAGGG-3' and the reverse primer sequence was 5'-GCGCTCGAGTTAAACATTCTCCGT-TTAC-3', with the restriction sites denoted in bold. The reverse primer also encodes the complement to a stop codon.

The amplified *fabG* fragment containing *NdeI* and *XhoI* restriction sites was ligated into the pGEM-T Easy vector (Promega) using T4 DNA ligase (Roche Applied Sciences) to create the T-*fabG* plasmid. DNA sequencing confirmed the fidelity of the amplified gene. The *fabG* fragment was excised from the T-*fabG* vector using *NdeI* and *XhoI* restriction enzymes and ligated into the pETDuet-1 expression vector (Novagen), previously digested with *NdeI* and *XhoI*, yielding pETDuet-*fabG*. This protocol produced a recombinant protein that has an identical amino-acid sequence to that of the wild-type protein, without any additional residues arising from the parental vector.

2.2. Expression and purification

The final pETDuet-*fabG* construct harboring *A. aeolicus fabG* was transformed into *Escherichia coli* Rosetta (DE3) strain (Novagen) for protein expression. Cells from an overnight 10 ml pre-culture were used to inoculate 1 l LB medium containing ampicillin (100 $\mu\text{g ml}^{-1}$) and chloramphenicol (35 $\mu\text{g ml}^{-1}$) at 310 K and 220 rev min^{-1} . Protein expression was induced with 1 mM isopropyl β -D-thiogalactopyranoside (IPTG) when the culture reached an OD_{600} of approximately 0.4–0.6 and the culture was incubated for 4–6 h prior to harvesting. The harvested cells were stored at 253 K.

The cell pellets were thawed and resuspended in 100 ml buffer A (50 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 mM EDTA, 1 mM DTT) and 1 ml protease-inhibitor cocktail (Sigma). Cells were lysed by sonication. Whole cell lysate was then centrifuged at 35 000g for 60 min at 277 K. The supernatant was loaded onto a 5 ml HiTrap Q Sepharose Fast Flow anion-exchange column (Amersham Bio-

sciences) equilibrated with buffer A. After washing the column with 20 column volumes (CV) of buffer A, the protein was eluted with a linear gradient (10–300 mM NaCl over 15 CV). The peak containing Aae FabG was pooled and diluted tenfold with buffer A to reduce the NaCl concentration. The sample was then loaded onto a Mono Q 10/10 anion-exchange column (Amersham Biosciences) equilibrated with buffer A. After washing with 10 CV buffer A, Aae FabG was eluted using a linear gradient of NaCl (10–300 mM over 10 CV). The peak containing Aae FabG was pooled and concentrated to 2 ml. The final purification step was size-exclusion chromatography (HiLoad 16/60 Superdex 200 column, Amersham Biosciences) using buffer B (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 1 mM EDTA, 1 mM DTT). The peak containing Aae FabG was collected and concentrated to $\sim 10 \text{ mg ml}^{-1}$ in buffer B for crystallization. SDS-PAGE was used to judge the purity of protein samples.

2.3. Crystallization

Initial crystallization screening of Aae FabG at 10 mg ml^{-1} in buffer B used a robotic 1536-condition microbatch-under-oil screen at 296 K (Luft *et al.*, 2001). Crystallization leads were identified in over 50 of the conditions. One initial condition [100 mM MES pH 6.0, 100 mM NH_4Cl , 40% (v/v) PEG 400, with droplets containing 200 nl protein solution and 200 nl screen cocktail], which produced single crystals, was optimized using hanging-drop vapor diffusion to obtain crystals suitable for diffraction studies. The optimal conditions were 7 mg ml^{-1} protein and a reservoir solution consisting of 100 mM MES pH 6.0, 120 mM NH_4Cl , 36% (v/v) PEG 400, with the crystallization drop containing 2 μl protein solution and 2 μl reservoir solution. The maximum dimensions of the crystals were $0.20 \times 0.10 \times 0.05 \text{ mm}$.

2.4. Diffraction data collection and processing

No additional cryoprotectant was required for preparation of the crystals for X-ray diffraction studies under cryoconditions as the PEG 400 in the mother liquor served this purpose. Crystals were mounted directly from the crystallization droplet in nylon loops (Hampton Research) and flash-cooled by plunging into liquid nitrogen. Diffraction data were collected at SSRL beamlines 9-1 and 11-1 *via* remote control using the Stanford Automated Mounting System (Cohen *et al.*, 2002) and *Blu-Ice* control software (McPhillips *et al.*, 2002). Both beamlines were equipped with an ADSC Quantum-315 CCD detector. Data were processed using the *HKL-2000* suite

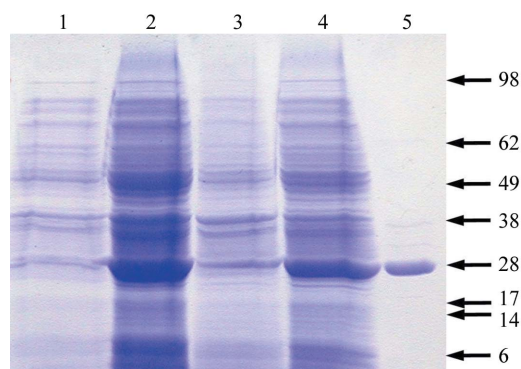


Figure 1 Coomassie-stained SDS-PAGE analysis of Aae FabG purification. The contents of the lanes are as follows: lane 1, pre-induction whole cell lysate; lane 2: post-induction whole cell lysate; lane 3: insoluble fraction of IPTG-induced cell lysate; lane 4: soluble fraction of IPTG-induced cell lysate; lane 5: final purified Aae FabG sample. Molecular weights (kDa) are marked on the right.

Table 1
Diffraction data statistics.

Values in parentheses are for the highest resolution shell.

	Crystal 1	Crystal 2
Beamline	SSRL 11-1	SSRL 9-1
Wavelength (Å)	0.9793	0.9795
Temperature (K)	100	100
Space group	$P2_12_12$	$P2_12_12$
Resolution limits (Å)	50.0–2.50 (2.59–2.50)	40.0–1.80 (1.86–1.80)
Unit-cell parameters (Å)	$a = 105.2$, $b = 64.5$, $c = 73.7$	$a = 105.7$, $b = 63.6$, $c = 73.6$
Mosaicity (°)	0.95	0.37
Total No. of reflections	58002	231843
Unique reflections	17476	46052
Redundancy	3.3 (2.8)	5.0 (3.4)
$\langle I/\sigma(I) \rangle$	11.6 (5.1)	14.1 (5.0)
Completeness (%)	96.1 (79.1)	98.3 (89.6)
R_{merge} (%)	6.6 (25.5)	5.6 (30.6)
Overall B factor from Wilson plot (Å ²)	33.7	18.0

(Otwinowski & Minor, 1997), yielding scaled intensities. Molecular replacement made use of the automated web service *CaspR* (Claude *et al.*, 2004).

3. Results and discussion

Aae FabG expressed well in *E. coli* Rosetta (DE3) cells and was completely soluble. The heterologously expressed protein possessed an amino-acid sequence identical to that of the wild-type protein, lacking any affinity tag. Thus, a low-resolution followed by a high-resolution anion-exchange approach was utilized to purify the protein from the lysate supernatant, with a final gel-filtration step. The final yield of pure protein was approximately 15 mg per litre of culture and only a single band corresponding to the expected weight of Aae FabG was observed on SDS-PAGE (Fig. 1).

Crystallization leads were obtained from a 1536-condition robotic microbatch-under-oil screen and were subsequently optimized by vapor diffusion. Initial attempts to collect diffraction data yielded modest-quality data to a maximum of 2.5 Å resolution (Table 1; crystal 1). However, the mosaicity was larger than desired (~1°). These crystals grew in <24 h and it was likely that the fast growth

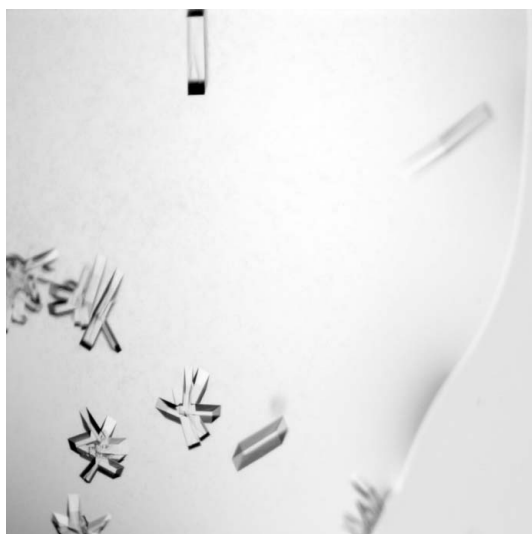


Figure 2
Crystals of Aae FabG grown *via* vapor diffusion in a hanging drop.

resulted in a poorly ordered crystal lattice. Crystallization conditions were altered in order to slow nucleation and growth, leading to high-quality crystals (Fig. 2) that diffracted to 1.8 Å resolution (Table 1; crystal 2). The diffraction data collected from all crystals studied were consistent with the orthorhombic crystal system, with unit-cell parameters $a = 105.7$, $b = 63.6$, $c = 73.6$ Å and systematic absences consistent with space group $P2_12_12$. The Matthews coefficient (Matthews, 1968) would be 4.6, 2.3 and 1.5 Å³ Da⁻¹ for one, two and three subunits of Aae FabG within the asymmetric unit, respectively. This result suggests the presence of two subunits per asymmetric unit.

Molecular-replacement (MR) techniques were applied to the 2.5 Å resolution data set (crystal 1). The availability of several homologous β -k-ACPR structures in the Protein Data Bank suggested MR as the method of choice for structure solution. The *CaspR* webserver for automated MR using homology modeling was used (Claude *et al.*, 2004), with a search for two subunits in the asymmetric unit requested. The FabG homologues from *P. falciparum* (PDB code 2c07, chain A) and *E. coli* (PDB code 1q7b; chain A) were input as initial search models and for use in the automated homology-modeling routine. Both have ~49% sequence identity to Aae FabG. Multiple sequence alignment identified three segments for truncation from the search models, corresponding to residues 1–6, 52–56 and 192–197 of the Aae FabG sequence. The best MR solution determined was with the truncated 1q7b monomer search model placed into the asymmetric unit as a dimer. This solution produced a c factor (the correlation coefficient calculated using $|F_{\text{obs}}|$ and $|F_{\text{calc}}|$) of 55.8% and a conventional R factor of 47.1% using data in the resolution range 15.0–3.0 Å. Automated *CNS* refinement of this solution gave a final R of 36.7% and an R_{free} of 44.3%. This refinement used the sequence of the truncated *E. coli* protein rather than the Aae FabG sequence. Initial analysis of the crystal structure reveals that Aae FabG is present as a tetramer comprised of a dimer within the asymmetric unit related to a second dimer by crystallographic twofold rotational symmetry. The tetrameric quaternary structure was as expected based upon the FabG-homologue crystal structures previously deposited in the PDB.

4. Conclusions

Aae FabG has been cloned and successfully expressed in *E. coli*. Good-quality diffraction data have been collected to 1.8 Å resolution and phasing has been performed by MR methods. Structure refinement is under way, as well as efforts to crystallize the enzyme–NADP cofactor complex. Previously, this protein has only been characterized by sequence analysis, which indicated that it is likely to participate in fatty-acid biosynthesis. We are interested in the structural/functional aspects of this protein as part of a larger project aimed at identifying ensembles of co-varying residues, or fingerprints, that define SCOR subfamilies and are predictive of substrate and cofactor preference. Furthermore, this project aims to determine in detail the structural/functional role of these fingerprint residues.

We thank Dr Robert Huber (Universitt Regensburg, Germany) for the kind gift of *A. aeolicus* genomic DNA and the Hauptman–Woodward Institute's High Throughput Crystallization Laboratory directed by Dr George DeTitta and Mr Joseph Luft for performing initial crystallization screening. This project was partially funded by NIH grant DK026546 (WLD). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US

Department of Energy, Office of Basic Energy Sciences, with support from DoE and NIH.

References

- Claude, J. B., Suhre, K., Notredame, C., Claverie, J. M. & Abergel, C. (2004). *Nucleic Acids Res.* **32**, W606–W609.
- Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. (2002). *J. Appl. Cryst.* **35**, 720–726.
- Duax, W. L., Pletnev, V., Addlagatta, A., Bruenn, J. & Weeks, C. M. (2003). *Proteins*, **53**, 931–943.
- Duax, W. L., Thomas, J., Pletnev, V., Addlagatta, A., Huether, R., Habegger, L. & Weeks, C. M. (2005). *Ann. NY Acad. Sci.* **1061**, 135–148.
- Fisher, M., Kroon, J. T., Martindale, W., Stuitje, A. R., Slabas, A. R. & Rafferty, J. B. (2000). *Structure*, **8**, 339–347.
- Luft, J. R., Wolfley, J., Jurisica, I., Glasgow, J., Fortier, S. & DeTitta, G. T. (2001). *J. Cryst. Growth*, **232**, 591–595.
- McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Oppermann, U., Filling, C., Hult, M., Shafqat, N., Wu, X., Lindh, M., Shafqat, J., Nordling, E., Kallberg, Y., Persson, B. & Jornvall, H. (2003). *Chem. Biol. Interact.* **143–144**, 247–253.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Price, A. C., Zhang, Y. M., Rock, C. O. & White, S. W. (2001). *Biochemistry*, **40**, 12772–12781.
- Price, A. C., Zhang, Y. M., Rock, C. O. & White, S. W. (2004). *Structure*, **12**, 417–428.
- Rock, C. O. & Jackowski, S. (2002). *Biochem. Biophys. Res. Commun.* **292**, 1155–1166.
- Wickramasinghe, S. R., Inglis, K. A., Urch, J. E., Muller, S., van Aalten, D. M. & Fairlamb, A. H. (2006). *Biochem. J.* **393**, 447–457.
- Yang, J. K., Park, M. S., Waldo, G. S. & Suh, S. W. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 455–460.
- Yang, J. K., Yoon, H. J., Ahn, H. J., Lee, B. I., Cho, S. H., Waldo, G. S., Park, M. S. & Suh, S. W. (2002). *Acta Cryst. D* **58**, 303–305.