Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Structure of the D-alanylgriseoluteic acid biosynthetic protein EhpF, an atypical member of the ANL superfamily of adenylating enzymes

**Asim K. Bera,[a] Vesna Atanasova,[a] Swarna Gamage,[b] Howard Robinson[c] and James F. Parsons[a]***

[a]Center for Advanced Research in Biotechnology, The University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850, USA, [b]Auckland Cancer Society Research Centre, School of Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand, and [c]Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA

Correspondence e-mail: parsonsj@umbi.umd.edu

Received 19 January 2010
Accepted 5 March 2010

**PDB References:** EhpF, native, 3hgu; SeMet derivative, 3hgv; phenazine-1,6-dicarboxylate complex, 3l2k.

The structure of EhpF, a 41 kDa protein that functions in the biosynthetic pathway leading to the broad-spectrum antimicrobial compound D-alanylgriseoluteic acid (AGA), is reported. A cluster of approximately 16 genes, including *ehpF*, located on a 200 kbp plasmid native to certain strains of *Pantoea agglomerans* encodes the proteins that are required for the conversion of chorismic acid to AGA. Phenazine-1,6-dicarboxylate has been identified as an intermediate in AGA biosynthesis and deletion of *ehpF* results in accumulation of this compound *in vivo*. The crystallographic data presented here reveal that EhpF is an atypical member of the acyl-CoA synthase or ANL superfamily of adenylating enzymes. These enzymes typically catalyze two-step reactions involving adenylation of a carboxylate substrate followed by transfer of the substrate from AMP to coenzyme A or another phosphopantetheine. EhpF is distinguished by the absence of the C-terminal domain that is characteristic of enzymes from this family and is involved in phosphopantetheine binding and in the second half of the canonical two-step reaction that is typically observed. Based on the structure of EhpF and a bioinformatic analysis, it is proposed that EhpF and EhpG convert phenazine-1,6-dicarboxylate to 6-formylphenazine-1-carboxylate *via* an adenylyl intermediate.

## 1. Introduction

*Pantoea agglomerans* is a Gram-negative plant and soil-dwelling bacteria with the potential to cause human infections (Cruz *et al.*, 2007). Historically, illnesses have been traced to either contaminated prepackaged pharmaceutical products or to plant-related traumatic injuries such as impalings (Maki *et al.*, 1976; Matsaniotis *et al.*, 1984; Ulloa-Gutierrez *et al.*, 2004). Additionally, cases of nosocomial infection have been reported that are frequently associated with central venous catheter usage (Uche, 2008). *P. agglomerans* is also of agricultural importance. It is an important biocontrol agent that is used to prevent fireblight disease in apples, pears and other economically important crops (Kearns & Mahanty, 1998; Kempf & Wolf, 1989; Wright & Beer, 2005). The ability of *P. agglomerans* to control the fireblight pathogen and to cause human disease is related in part to its ability to produce antimicrobial phenazine compounds such as D-alanylgriseoluteic acid (AGA; Giddens *et al.*, 2003; Stockwell *et al.*, 2002; Fig. 1). AGA has been shown to have broad-spectrum antimicrobial activity and may have significant potential in terms

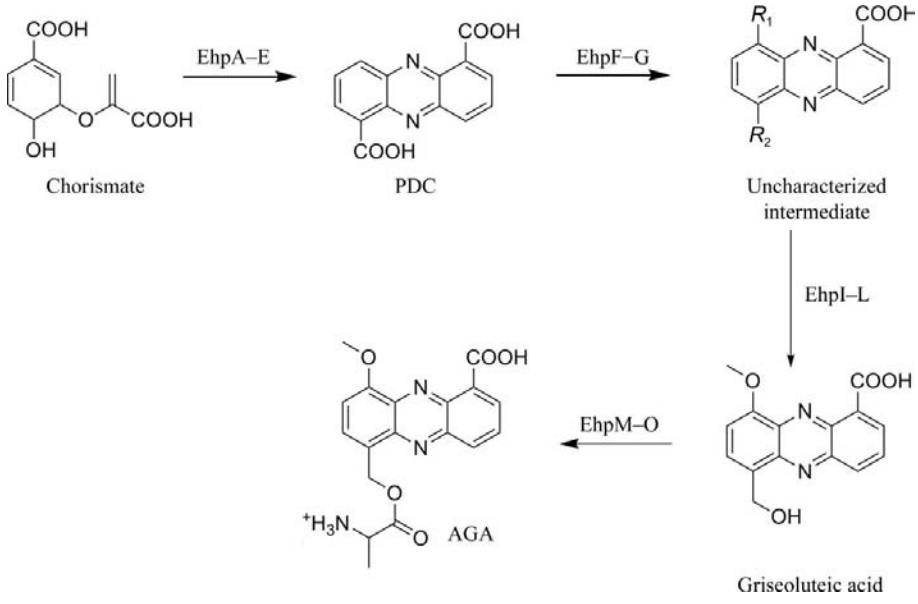


**Figure 1**
Modified version of the biosynthetic pathway leading to AGA first proposed by Giddens *et al.* (2002)[1].

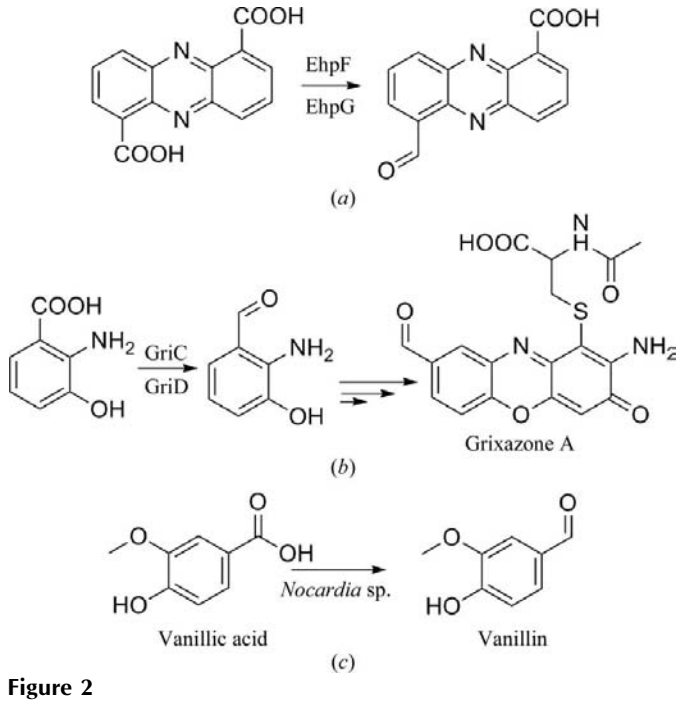


**Figure 2**
Proposed function of EhpF based on the crystallographic observations and its homology to GriC from the grixazone pathway and an arylcarboxylate reductase from *Nocardia*. (*a*) Proposed conversion of PDC to 6-formyl-phenazine-1-carboxylate catalyzed by EhpF and EhpG[1]. (*b*) The role of GriC and GriD in grixazone biosyntheis. (*c*) Vanillin production in *Nocardia* catalyzed by an arylcarboxylate reductase enzyme that features an N-terminal EhpF/GriC-like domain.

of the development of new antibiotics (Giddens & Bean, 2007). Paradoxically, based on the involvement of similar phenazine compounds such as pyocyanin (Bera *et al.*, 2009) in pathogenesis, AGA may play a role in *P. agglomerans* infections of humans. Additionally, the interesting biochemistry involved in AGA production may reveal key information about the evolution of secondary metabolic pathways and the roles of key enzymes in pathogenesis. Therefore, we have undertaken a structural and biochemical analysis of AGA biosynthesis.

The genes required for the conversion of chorismate to AGA in *P. agglomerans* (*ehpA–M*) are located on a 200 kbp plasmid native to certain isolates of the organism. Phenazine-1,6-dicarboxylate has been identified as an intermediate in AGA production; however, the other intermediates between PDC and AGA are unknown (Fig. 1). Giddens and coworkers proposed that EhpF and EhpG[1] convert PDC to an unknown downstream product (Giddens *et al.*, 2002). This uncharacterized intermediate was further proposed to be converted to griseoluteic acid (GA) by the combined actions of EhpI, EhpJ, EhpK and EhpL. EhpM, EhpN and EhpO were proposed to convert GA to AGA (Fig. 1).

Bioinformatic analysis revealed that EhpF and EhpG are homologues of GriC and GriD from *Streptomyces griseus* (sequence alignment provided as Supplementary Material[2]; Suzuki *et al.*, 2007). GriC and GriD together catalyze the formation of 2-amino-3-hydroxybenzaldehyde from 2-amino-3-hydroxybenzoate during one of the early steps in the biosynthesis of the antibiotic grixazone (Fig. 2). Additionally, both EhpF and GriC show similarity to the vanillin-specific carboxylic acid reductase from *Nocardia* (Fig. 2; Venkitasubramanian *et al.*, 2007). The N-terminal domain of this enzyme shows sequence similarity to both EhpF and GriC. A reasonable prediction based on the observations of Giddens *et al.* (2002) and the similarity of EhpF to GriC is that EhpF and EhpG catalyze the conversion of PDC to 6-formylphenazine-1-carboxylate (Fig. 2).

Here, we report the structure of EhpF alone and in complex with its predicted substrate PDC. We also provide a description of the similarities and differences observed between EhpF and members of the ANL superfamily of adenylating enzymes (Gulick, 2009). Among the structural homologues of EhpF identified by automated alignment algorithms are 2,3-dihydroxybenzoate-AMP ligase (DhbE; May *et al.*, 2002), D-alanine-poly(phosphoribitol) ligase (DltA; Osman *et al.*, 2009), benzoate-CoA ligase (Bains & Boulanger, 2007), firefly

[1] Giddens *et al.* (2002) annotated two genes, *ehpG* and *ehpH*, as separate open reading frames. Sequence analysis of EhpH reveals homology only to the C-terminal region of EhpG homologues, suggesting that EhpG and EhpH may be encoded by a single open reading frame and that the original annotation may be inaccurate, perhaps as a result of a sequencing error. 'EhpG' in this work refers to a combined 'EhpGH' enzyme.
[2] Supplementary material has been deposited in the IUCr electronic archive (Reference: HM5085). Services for accessing this material are described at the back of the journal.

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

Values in parentheses are for the highest resolution shell.

| | SeMet EhpF ($P3_121$) | Native EhpF | SeMet EhpF ($P2_12_12_1$) | EhpF–PDC complex |
|---|---|---|---|---|
| PDB code | | 3hgu | 3hgv | 3l2k |
| Space group | $P3_121$ | $P3_121$ | $P2_12_12_1$ | $I23$ |
| Wavelength (Å) | 0.979200 [Se peak] | 1.08090 | 0.979199 [Se peak] | 1.5418 |
| Unit-cell parameters (Å) | $a = b = 89.48$, $c = 188.19$ | $a = b = 89.85$, $c = 188.56$ | $a = 68.836$, $b = 110.143$, $c = 112.044$ | $a = b = c = 193.13$ |
| No. of measured intensities | 989742 | 1022966 | 558695 | 434418 |
| No. of unique reflections | 51819 | 65275 | 38714 | 29538 |
| Resolution of data (Å) | 30.00–2.10 | 30.00–1.95 | 30.0–2.30 | 48.28–2.80 |
| $R_{\rm merge}$ | 0.069 (0.572) | 0.075 (0.375) | 0.066 (0.506) | 0.100 (0.501) |
| Completeness (%) | 100 (100) | 100 (100) | 100 (100) | 99.9 (100) |
| Redundancy | 19.1 (19.2) | 15.7 (12.7) | 14.4 (14.3) | 14.71 (12.57) |
| Mean $I/\sigma(I)$ | 55.8 (5.6) | 29.9 (5.8) | 45.3 (5.2) | 12.0 (3.2) |
| Overall figure of merit | 0.66 (0.34) | | 0.64 (0.31) | |
| Refinement statistics | | | | |
| Resolution limits | | 30–1.95 | 28.34–2.30 | 26.78–2.80 |
| $R$ | | 0.193 (0.220) | 0.180 (0.213) | 0.212 (0.250) |
| $R_{\rm free}$ | | 0.240 (0.322) | 0.249 (0.331) | 0.288 (0.318) |
| No. of water molecules | | 394 | 355 | 16 |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | | 0.015 | 0.025 | 0.030 |
| Bond angles (°) | | 1.536 | 1.994 | 1.961 |
| Ramachandran plot | | | | |
| Preferred (%) | | 91.1 | 89.2 | 84.2 |
| Allowed (%) | | 8.9 | 10.6 | 15.1 |
| Outliers (%) | | 0.0 | 0.2 | 0.7 |
| Average $B$ (Å$^2$) | | | | |
| Main chain | | 42.7 | 39.33 | 36.14 |
| Side chain | | 44.3 | 40.62 | 35.46 |
| Water | | 45.2 | 42.10 | 50.87 |

luciferase (Nakatsu *et al.*, 2006) and 4-chlorobenzoate-CoA ligase (Reger *et al.*, 2008). Notably, EhpF lacks the C-terminal domain that is present in all of its known homologues. Consistent with the prediction that EhpF acts on PDC, we were able to obtain a structure of EhpF in complex with PDC. Purified EhpF, however, displays no activity, suggesting that another enzyme may be involved and required, possibly EhpG, the enzyme encoded immediately downstream of EhpF.

## 2. Experimental

### 2.1. Cloning and expression of EhpF

The *ehpF* gene from *P. agglomerans* was synthesized commercially and subcloned into the expression vector pET28a (EMD) for protein expression. The gene sequence was optimized for expression in *Escherichia coli* and was based on the published EhpF sequence (Giddens *et al.*, 2002). *E. coli* strain BL21 (DE3) was used to express EhpF. Cells transformed with the Pet28a-*ehpF* plasmid were grown in ZYP-5052 medium containing 100 µg ml$^{-1}$ kanamycin (Studier, 2005). Initial growth was at 310 K. When the culture density reached an optical density of ∼0.5 at 600 nm, the temperature was reduced to 293 K and the cells were harvested by centrifugation after ∼16 h. The cells were lysed by sonication and His-tagged EhpF was purified by $Co^{2+}$-ion affinity chromatography as directed by the resin manufacturer (Sigma). Human $\alpha$-thrombin (Haematologic Technologies) was used to remove the His tag from the purified protein. Thrombin was removed from the solution using benzamidine agarose and EhpF was further purified by passage over a second $Co^{2+}$ column. Pure EhpF was dialyzed against 25 m*M* Tris, 1 m*M* DTT pH 7.6, concentrated to ∼16 mg ml$^{-1}$ and stored at 193 K in 250 µl aliquots. The yield was typically ∼50 mg pure EhpF per litre of culture. Selenomethionine-labeled EhpF was produced using *E. coli* strain B834 (DE3) in minimal M9 medium as described previously (Doublié, 2007).

### 2.2. Crystallization

Initial crystallization screening of EhpF was conducted at ∼295 K using the sitting-drop vapor-diffusion method with Hampton Research SaltRx HT and Index HT screening kits. The protein solution was at 16 mg ml$^{-1}$ before crystallization. Equal volumes of protein solution and reservoir solution were mixed (typically 2 µl in total) to initiate the crystallization experiment. EhpF crystallizes in at least three distinct forms. Based on initial screening results, optimized hexagonal crystals were obtained using 1 *M* sodium/potassium tartrate and 0.1 *M* Tris–HCl pH 7.5 as the well solution. Diffraction-quality crystals were typically obtained after ∼2 d for both native and SeMet-labeled protein. A second crystal form was obtained using 5% Tacsimate pH 7.0, 0.1 *M* HEPES and 10% PEG 5000 monomethyl ether as the well solution. In this case, diffraction-quality crystals were obtained after 8–10 d for both native and SeMet-labeled protein. Crystals of the complex between EhpF and the proposed substrate PDC were grown using a well solution of 1.0 *M* monobasic sodium phosphate monohydrate and 1 *M* dibasic potassium phosphate pH 5.0. Cocrystallization experiments were initiated by mixing 16 mg ml$^{-1}$ EhpF and 10 m*M* PDC in a 3:1 ratio at room temperature immediately prior to mixing with the well solution as described above. Crystals grew in 2–3 d. PDC was synthesized as described previously (Flood *et al.*, 1972).

## 3. Data collection

Crystals were mounted in Hampton Research pre-mounted loops using 25% glycerol as a cryoprotectant and were then flash-cooled in liquid nitrogen. Single-wavelength anomalous dispersion (SAD) data sets were collected from SeMet-labeled protein crystals on the X29 beamline at the National

Synchrotron Light Source (NSLS; Brookhaven National Laboratory). A native data set was also collected on the NSLS X29 beamline. All data were indexed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Crystals grown in sodium/potassium tartrate belonged to space group $P3_121$. SeMet-labeled crystals diffracted to 2.1 Å resolution and native crystals diffracted to 1.9 Å resolution. The second crystal form, grown in Tacsimate, belonged to space group $P2_12_12_1$. SeMet-labeled crystals diffracted to 2.3 Å resolution. Native data were not collected for this crystal form. Both crystal forms contained two molecules per asymmetric unit. Diffraction data for the EhpF–PDC complex were collected in-house using a Rigaku Micro Max 007 rotating-anode generator and an R-AXIS IV$^{++}$ detector (Rigaku). The crystals were cooled to 105 K using an X-stream 2000 cryocooler (Rigaku). The crystals were cryoprotected by dipping the crystal in a drop made up of equal volumes of well solution and 50% PEG 400 before mounting them on the goniometer. Diffraction data were processed with *CrystalClear*/*d*TREK* (Pflugrath, 1999). These crystals belonged to space group *I*23 and diffracted to 2.8 Å resolution.

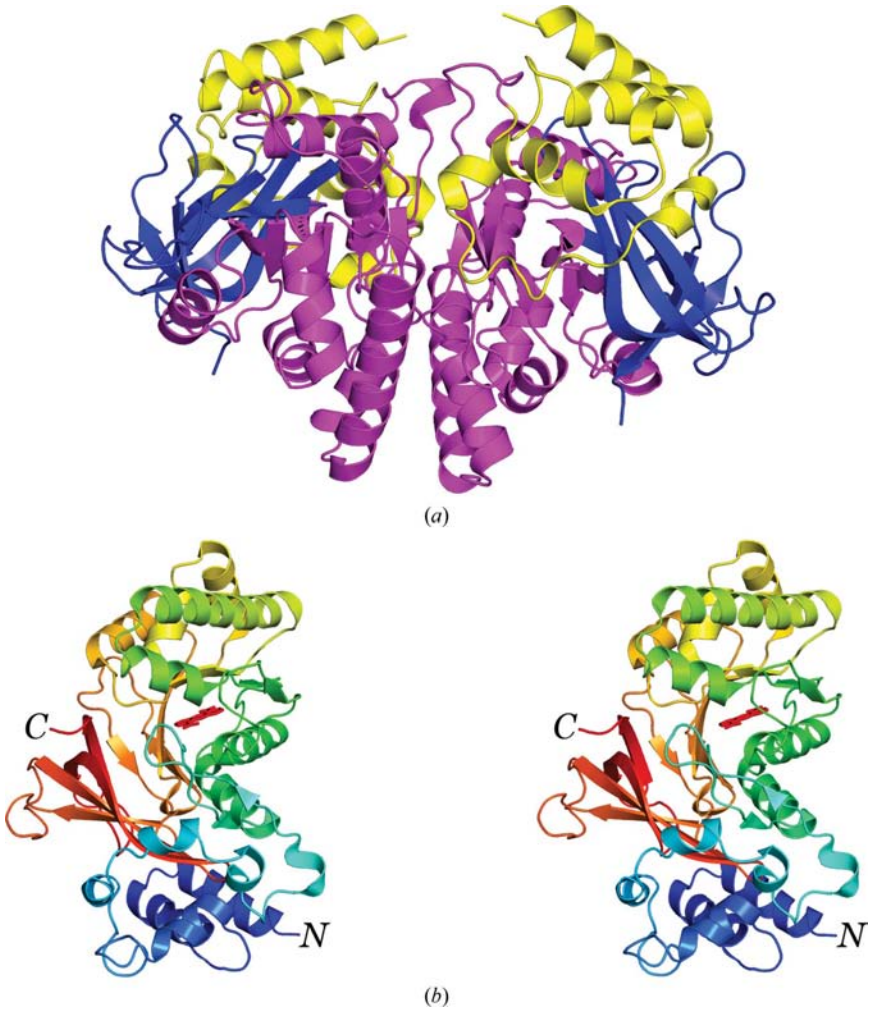


**Figure 3**
Structure of EhpF. (*a*) Ribbon diagram of the EhpF dimer. The subunits are colored by subdomain: residues 1–81 are shown in yellow, residues 82–267 are shown in magenta and residues 268–354 are shown in blue. (*b*) Ribbon diagram of the EhpF monomer colored in rainbow. The N- and C-termini are labeled. Bound PDC is shown in red.

### 3.1. Structural solution and refinement

Independent SAD phasing was carried out for both SeMet-labeled crystal forms using *HKL2MAP* (Pape & Schneider, 2004) with density modification involving solvent flattening. Phases from the 2.1 Å resolution data from the space group $P3_121$ SeMet-labeled crystals were used to solve the native unliganded EhpF structure. Automated model building was performed by *PHENIX* (Adams *et al.*, 2002). The EhpF–PDC complex structure was solved by molecular replacement using *Phaser* (McCoy *et al.*, 2007) with the native EhpF structure as the starting model. Statistics are shown in Table 1. The model was viewed and adjusted and parts were rebuilt with *Coot* (Emsley & Cowtan, 2004). *REFMAC*5 was used for refinement between viewing sessions (Murshudov *et al.*, 1997). The final statistics are shown in Table 1. The final models included residues 1–354 or 3–354 depending on the chain. The final 12 residues were disordered in all chains. In the apo EhpF structure no density was observed for residues 99–101 in both chains. In the apo SeMet-labeled structure residues 123–129 of both chains were disordered and were not included in the final model. Superimposing chains from the same structures results in calculated r.m.s. deviations of 0.4–0.5 Å. Superimposing the two apo EhpF structures on each other or on the PDC complex structure resulted in calculated r.m.s. deviations ranging from 0.4 to 0.7 Å.

## 4. Results and discussion

### 4.1. Overall structure

EhpF is a dimer in solution and in the crystallographic asymmetric unit (Fig. 3*a*). Each EhpF monomer folds into a single large mixed $\alpha/\beta$ domain (Fig. 3*b*). Residues 1–81 form an all-$\alpha$-helical subdomain. Residues 82–270 form a mixed $\alpha/\beta$ subdomain with a seven-stranded mixed $\beta$-sheet surrounded by five $\alpha$-helices, several of which contribute significantly to the dimer interface. The C-terminal portion of EhpF (residues 286–354) folds into a $\beta$-roll-type structure. The dimer interface of EhpF buries a total of approximately 3600 Å$^2$ of surface area. Subdomains 1 and 2 contribute to the interface, while the C-terminal $\beta$-roll motif does not make any intersubunit contacts (Fig. 3*a*). The long helix formed by residues 179–202

forms a significant portion of the dimer interface. The interface is a typical mixture of hydrophobic interactions and hydrogen bonds.

### 4.2. EhpF is an atypical member of the ANL superfamily

Overall, EhpF most closely resembles the enzymes of the proposed ANL superfamily (Gulick, 2009), which includes the acyl-CoA synthases, modular NRPS enzymes and firefly luciferase. Automated alignment algorithms illustrate that subdomains 2 and 3 of EhpF align well with the core subdomains of ANL-superfamily enzymes (Figs. 4 and 5). The N-terminal $\alpha$-helical subdomain of EhpF does not resemble the N-terminal region of any structurally characterized ANL-superfamily enzyme and interestingly EhpF entirely lacks the C-terminal domain that is seen in other adenylate-forming enzymes (Fig. 4). In 4-chlorobenzoate CoA ligase, for example, it has been shown that this C-terminal domain rotates significantly, adopting a second conformation that is required for the second half of the reaction: the transfer of 4-chlorobenzoate from AMP to CoA (Reger *et al.*, 2008). Should EhpF catalyze a similar reaction, the available evidence suggests that a second protein would have to be involved.

Figs. 4 and 5 present a comparison of EhpF and DhbE, a representative ANL-superfamily member. Annotated topological diagrams of EhpF and DhbE are available as Supplementary Material. DhbE adenylates 2,3-dihydroxybenzoate in the biosynthetic pathway leading to the cyclic siderophore bacillibactin (May *et al.*, 2002). EhpF and the larger N-terminal domain of DhbE are both comprised of three subdomains. However, only subdomains 2 and 3 are common; these are colored red in Fig. 4. The N-terminal subdomain of DhbE is topologically similar to its second subdomain. They are both $\alpha\beta\alpha$ three-layer sandwich motifs. EhpF, in contrast, has a unique all-helical N-terminal subdomain followed by a DhbE-like $\alpha\beta\alpha$ subdomain.

The first two subdomains of DhbE both contribute elements of two large domain-spanning $\beta$-sheets. One sheet is mostly made up of strands from subdomain 1 and is completed by the two strands from subdomain 2 (residues 184–190 and 200–204) that are connected by a P-loop nucleotide-binding motif. The second large $\beta$-sheet is composed of seven strands from subdomain 2 and two strands formed by residues 42–46 and 49–52 from the N-terminal subdomain. Since EhpF lacks this first $\alpha\beta\alpha$ structural unit, it has $\beta$-sheets that are smaller than those seen in DhbE. The first sheet only contains two strands (residues 92–96 and 104–108), while the larger second sheet contains seven strands rather than the nine seen in DhbE, since none are provided by the all-helical N-terminal subdomain of EhpF.

The $\beta$-roll-type motifs formed by EhpF and DhbE are quite similar, although a small helix is present in DhbE (residues 395–401) that is not observed in EhpF. Following the $\beta$-roll structure EhpF terminates, while DhbE continues for another ~125 residues forming the aforementioned C-terminal mixed $\alpha/\beta$ domain.

### 4.3. Nucleotide-binding site

EhpF exhibits many of the hallmarks of an ATP-dependent AMP ligase in both its primary and tertiary structures. Residues 96–104 of EhpF form a P-loop (phosphate-binding loop) structural motif that is common to many ATP-binding proteins

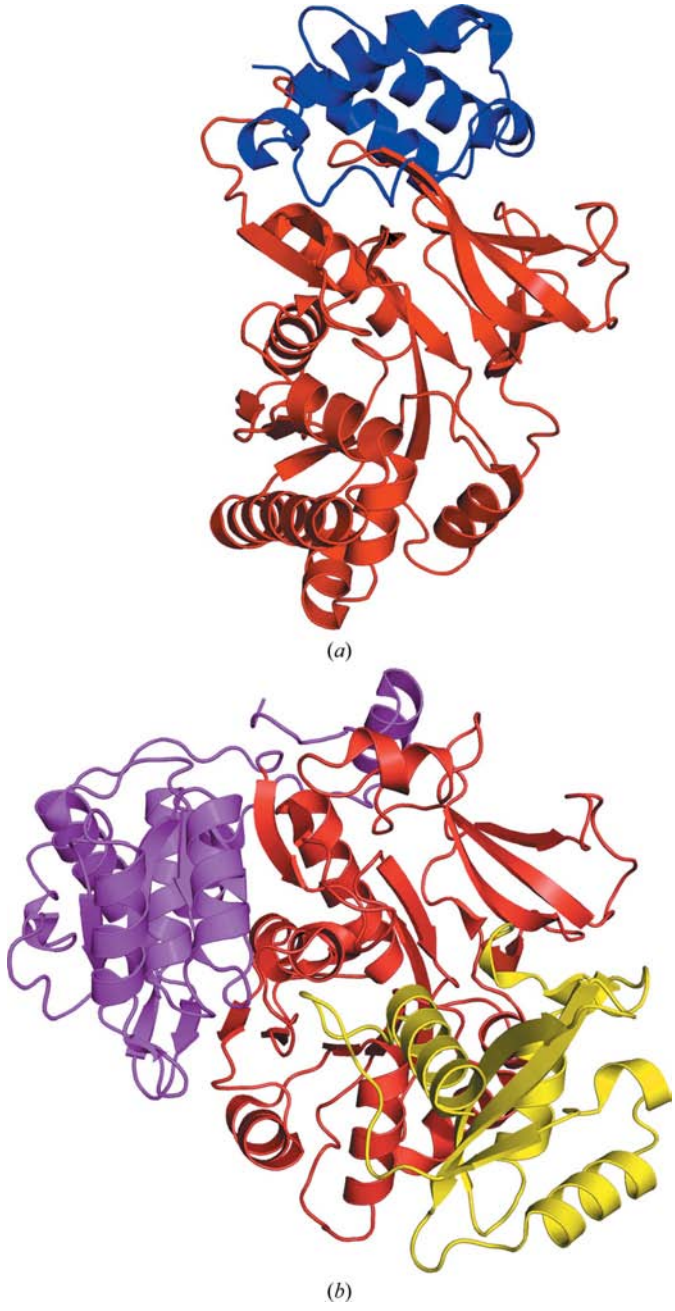


**Figure 4**
Comparison of the structure of EhpF with that of DhbE, a representative member of the ANL superfamily of adenylating enzymes (May *et al.*, 2002). Subdomains 2 and 3 of EhpF and subdomains 2 and 3 of domain I of DhbE (all shown in red) superimpose with an r.m.s.d. of ~3.3 Å for 229 $C^{\alpha}$ atoms. Structures are shown in identical orientations after superpositioning. (*a*) Ribbon diagram of the EhpF monomer. Subdomain 1 is colored blue, while subdomains 2 and 3 are colored red. (*b*) Ribbon diagram of DhbE (PDB code 1md9; May *et al.*, 2002). Subdomain 1 is shown in magenta, subdomains 2 and 3 are shown in red and the C-terminal domain (which is lacked by EhpF) is shown in yellow.

(May *et al.*, 2002). Examination of the amino-acid sequence of EhpF, including that of the P-loop ($^{96}$SGGTTGAPK$^{104}$), reveals that EhpF bears similarity to the adenylation domains of many nonribosomal peptide synthases, including DhbE and PheA (May *et al.*, 2002; Stevens *et al.*, 2006). Despite repeated attempts, we have not been successful in growing crystals of EhpF in complex with any nucleotide co-substrates. In addition to the P-loop residues, most of the expected elements of the putative ATP-binding site appear to be present in EhpF, however. Asp331, Ile350, Tyr262, Ile252 and the motif $^{235}$GGT$^{237}$ align with similar or identical residues involved in nucleotide recognition in other adenylating enzymes (Bains & Boulanger, 2007; May *et al.*, 2002; Reger *et al.*, 2008). Asp331 is expected to coordinate the 2′- and 3′-hydroxyl groups of the nucleotide sugar ring, a key interaction seen in ANL-superfamily enzymes.

At first glance, EhpF appears to lack two highly conserved residues, an arginine (Arg428 of DhbE) and a lysine (Lys519 of DhbE) that interact with the $\alpha$-phosphoryl O atoms of AMP in all adenylate-forming enzymes (Gulick, 2009). Based on sequence alignments alone, Arg346 of EhpF would be expected to be the conserved arginine. However, the loop connecting $\beta$12 and $\beta$13 of EhpF (residues 337–344) is longer than in other ANL-superfamily enzymes (Fig. 5). Because of this, Arg346 is out of register and is not near the active site. Structural alignments suggest that Pro352 or Leu353 occupies the analogous position in EhpF. The conserved lysine in question is contributed by the ANL-superfamily C-terminal domain that EhpF lacks. However, careful evaluation of the EhpF structure shows that an alternative scheme may have evolved to preserve these functional roles. Arg330 and Lys172 of EhpF are oriented in such a way that they could potentially fulfill the roles of Arg428 and Lys519 of DhbE (Fig. 6). Lys172 is part of the $\alpha$6 helix of EhpF, an element of the protein that does not superimpose with any region of DhbE (Fig. 5). Arg330 of EhpF (Gly412 of DhbE) appears to have the conformational freedom to potentially form an interaction with a bound nucleotide.

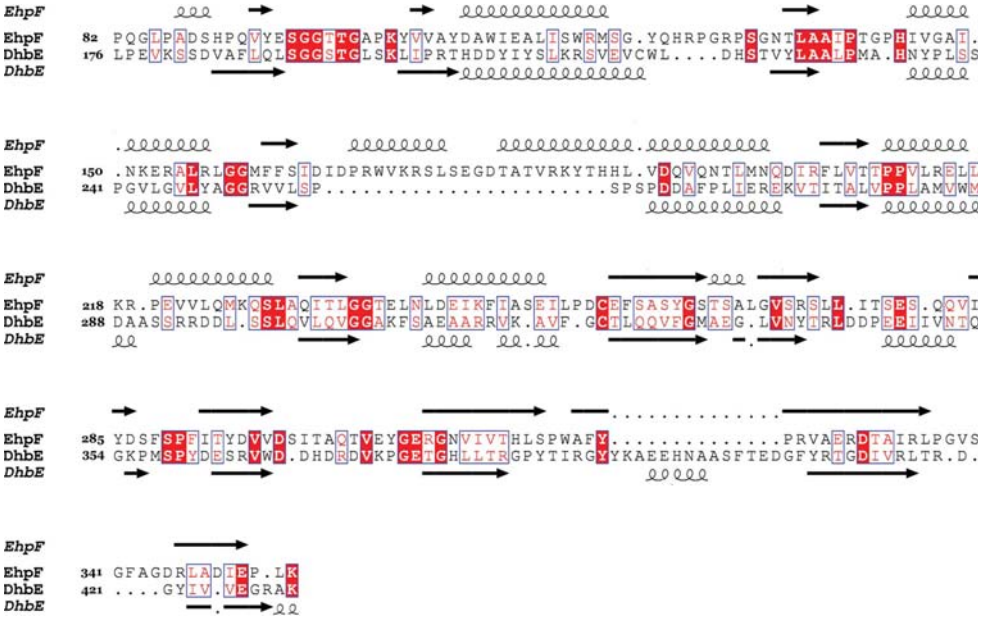


**Figure 5**
Structure-based sequence alignment of the homologous regions of EhpF and DhbE. The alignment was constructed using the *MATRAS* algorithm (Kawabata, 2003).

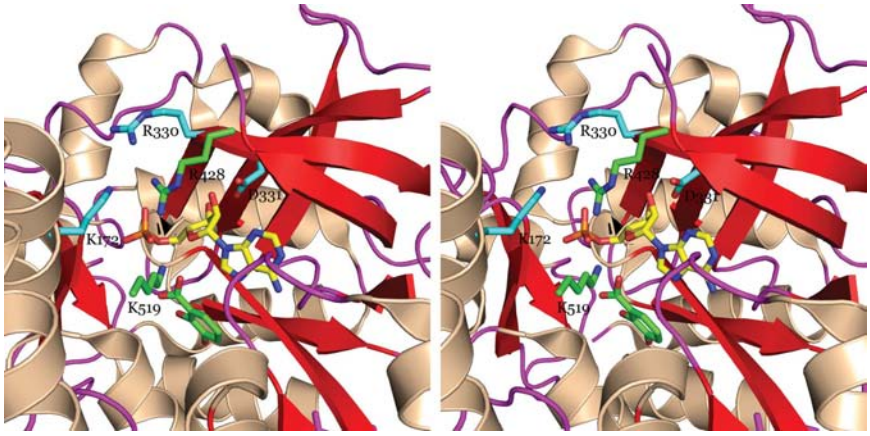


**Figure 6**
EhpF may use alternative residues to perform key functions. While most of the nucleotide-binding residues found in ANL-superfamily enzymes are conserved in EhpF, two key residues are missing at first glance. Stereoview of the potential nucleotide-binding site of EhpF, illustrating that Arg330 and Lys172 are in reasonable positions to perform roles analogous to the highly conserved residues Arg428 and Lys519 of DhBE. Following automated structural alignment, Arg428, Lys519 (both shown in green) and AMP (shown in yellow) from DhbE (PDB code 1md9) are displayed together with a ribbon diagram of EhpF. Asp331, Arg330 and Lys172 of EhpF are shown in blue.

### 4.4. Observed PDC-binding site

The structure of EhpF was solved in complex with its putative substrate PDC. Surprisingly, PDC is bound in a location distant from the canonical and predicted active site. The planar phenazine ring is bound between elements of the large $\beta$-sheet and the long loop connecting $\alpha$4 (residues 109–121) and $\beta$3. Phe205, Pro131, Leu157 and Leu275 interact with hydrophobic portions of the ring, while several arginine residues including Arg120, Arg127 and Arg153 make extensive polar contacts with the substrate carbox-

ylate groups and ring N atoms. Additionally, the phenazine ring is stacked against the side chain of Tyr124 (Fig. 7*a*). There are a number of possible explanations for the observed PDC-binding mode. Firstly, binding of PDC at the observed location could be physiologically irrelevant and a curiosity arising from the crystallization experiment. Secondly, the observed binding site could be an additional entryway to the active site. It has been shown in other enzymes that ATP is the second of the two substrates to bind and that its binding is triggered by structural rearrangements associated with binding of the first ligand (May *et al.*, 2002). It may be that PDC binding triggers a conformational change favoring nucleotide binding, which in turn triggers a second conformational change allowing PDC to slip deeper into the EhpF active site. In this scenario, the two substrates would approach the catalytic site from different sides of the enzyme. A third possibility is that the observed PDC-binding site is a regulatory binding site that blocks activity by inhibiting binding of ligand(s) to the catalytic site. It could be that the crystallization conditions favor PDC binding to the regulatory site. Lacking a functional assay, we are currently unable to discriminate between these possibilities. However, while in most cases it appears that substrates of many ANL-family enzymes access the catalytic site by simply binding prior to ATP, there is a precedent for the substrate to enter the active site through a path other than through the nucleotide-binding pocket. Long-chain acyl-CoA synthase homologues of EhpF are known to use a gating mechanism that allows 'one-way' movement of the nonpolar substrate through the protein, for example (Forneris & Mattevi, 2008; Hisanaga *et al.*, 2004). Indeed, given the hydrophobic nature and insolubility of PDC it would not be surprising if it were delivered to the EhpF active site by the enzyme upstream in the AGA-biosynthetic pathway, which is probably EhpE (Ahuja *et al.*, 2008).

### 4.5. Canonical substrate-binding site

A substrate-binding pocket analogous to the aryl acid-binding sites of DhbE, benzoate CoA ligase (Bains & Boulanger, 2007) and 4-chlorobenzoate CoA ligase is clearly evident in EhpF. It is defined by the side chains of the conserved stretch $^{262}$YGST$^{265}$ and the somewhat conserved stretch $^{144}$HIVGA$^{148}$. The side chains of Val270 and Ser271 also define this pocket, which appears to be large enough to easily accommodate PDC.

The observed PDC-binding pocket and the predicted binding pocket are 8–10 Å apart and are separated largely by the side chains of two arginine residues unique to EhpF: Arg120 and Arg153. If these residues were sufficiently flexible they could potentially act as gatekeepers, preventing or allowing access of PDC to the canonical active site. Fig. 8 illustrates the separation of the PDC-binding site and the catalytic site in EhpF by Arg120 and Arg153 as well as the potential consequences of these arginine residues acting as gatekeepers. Examination of the environment surrounding Arg120 and Arg153 suggests that their conformations are not rigidly constrained by steric concerns and that they may be able to adopt alternate conformations.

### 4.6. Role of EhpF in AGA biosynthesis

The AGA-biosynthetic machinery in *P. agglomerans* involves ∼13 proteins. The first five, EhpA–E, convert chorismic acid to PDC (Fig. 1). This portion of the pathway is similar to the phenazine-biosynthetic pathways seen in *Pseudomonas aeruginosa* and other organisms (Mavrodi *et al.*, 1998; Parsons *et al.*, 2004, 2007). The remaining proteins, EhpF–M, convert PDC to AGA. This portion of the pathway appears to be more specialized and less widely distributed compared with the phenazine-biosynthesis steps.

Functional clues as to the role of EhpF are gleaned by examining its structural or sequence homologues. Structurally, EhpF appears to belong to the ANL superfamily of adenylating enzymes, suggesting that an adenylyl intermediate may be involved in AGA biosynthesis. ANL-superfamily enzymes typically catalyze a two-step reaction transferring the activated substrate from AMP to CoA (Gulick, 2009). The absence of the C-terminal domain seen in other ANL-superfamily enzymes suggests that if EhpF catalyzes such a reaction

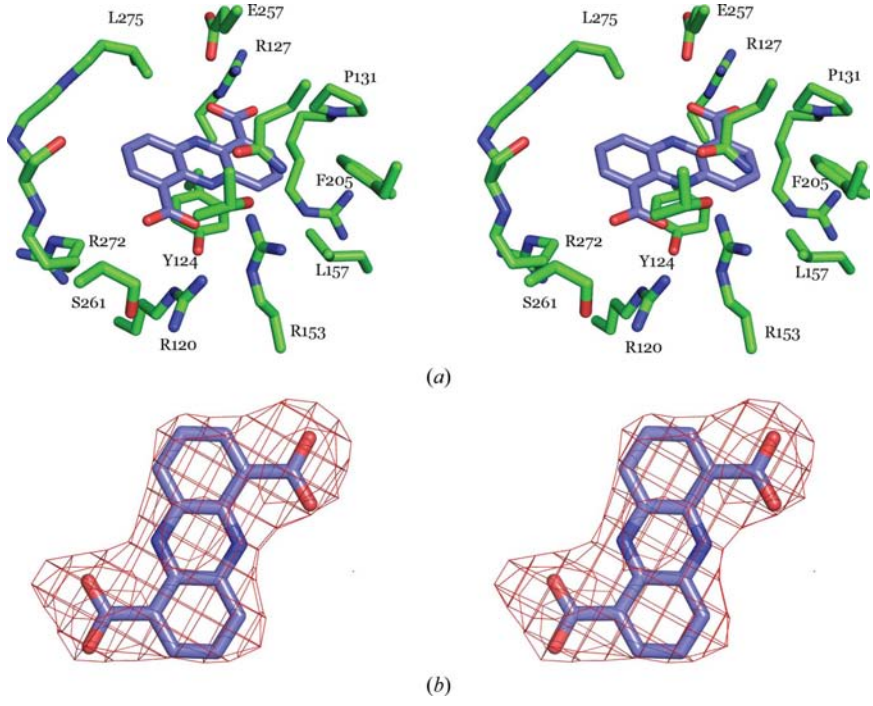


**Figure 7**
(*a*) Stereoview of the PDC-binding site. For clarity, Thr234 and Asn232 are shown but not labeled. (*b*) Stereo OMIT map and the final model of PDC. Positive difference density was calculated by omitting PDC from a round of refinement with *REFMAC*5. Density is contoured at 3$\sigma$.

it may not involve CoA and/or that a second enzyme that interacts with EhpF may be involved. Our attempts to directly evaluate the role of EhpG in an EhpFG complex have been complicated by uncertainties about the sequences of EhpG and EhpH. We have produced a number of *ehpGH* constructs, but none has yet yielded tractable protein.

Sequence homologues of EhpF provide further clues. As noted above, the GriC and GriD enzymes from *S. griseus* are believed to act as a carboxylate reductase in the grixazone-biosynthetic pathway (Suzuki *et al.*, 2007; Fig. 2). The structure of GriC is unknown; however, a comparison of the amino-acid sequences of GriC and EhpF reveals that they are 36% identical and 54% similar (see Supplementary Material). Like EhpF, GriC lacks the C-terminal domain that is characteristic of ANL-superfamily enzymes. In addition, the sequences of both GriD and EhpG identify them as belonging to the NAD(P)H-dependent aldehyde dehydrogenase superfamily. An additional similar reduction reaction is catalyzed by an enzyme from *Nocardia* that can convert vanillic acid to vanillin (Fig. 2). This reaction is catalyzed by an 1174-amino-acid enzyme that features an EhpF/GriC-like N-terminal adenylation domain and a C-terminal reductase domain. While no direct evidence is currently available to show that EhpF adenylates PDC, structural and bioinformatic evidence suggest that it is a strong possibility.

## 5. Conclusions

The structure of EhpF from *P. agglomerans* has been solved alone and in complex with its predicted substrate PDC. Surprisingly, PDC is bound in an unanticipated location approximately 8 Å from the expected active site. An intriguing possibility is that the observed binding pocket is a novel active-site entryway, essentially a back door. Such a scenario would require that two arginine residues, Arg120 and Arg153,

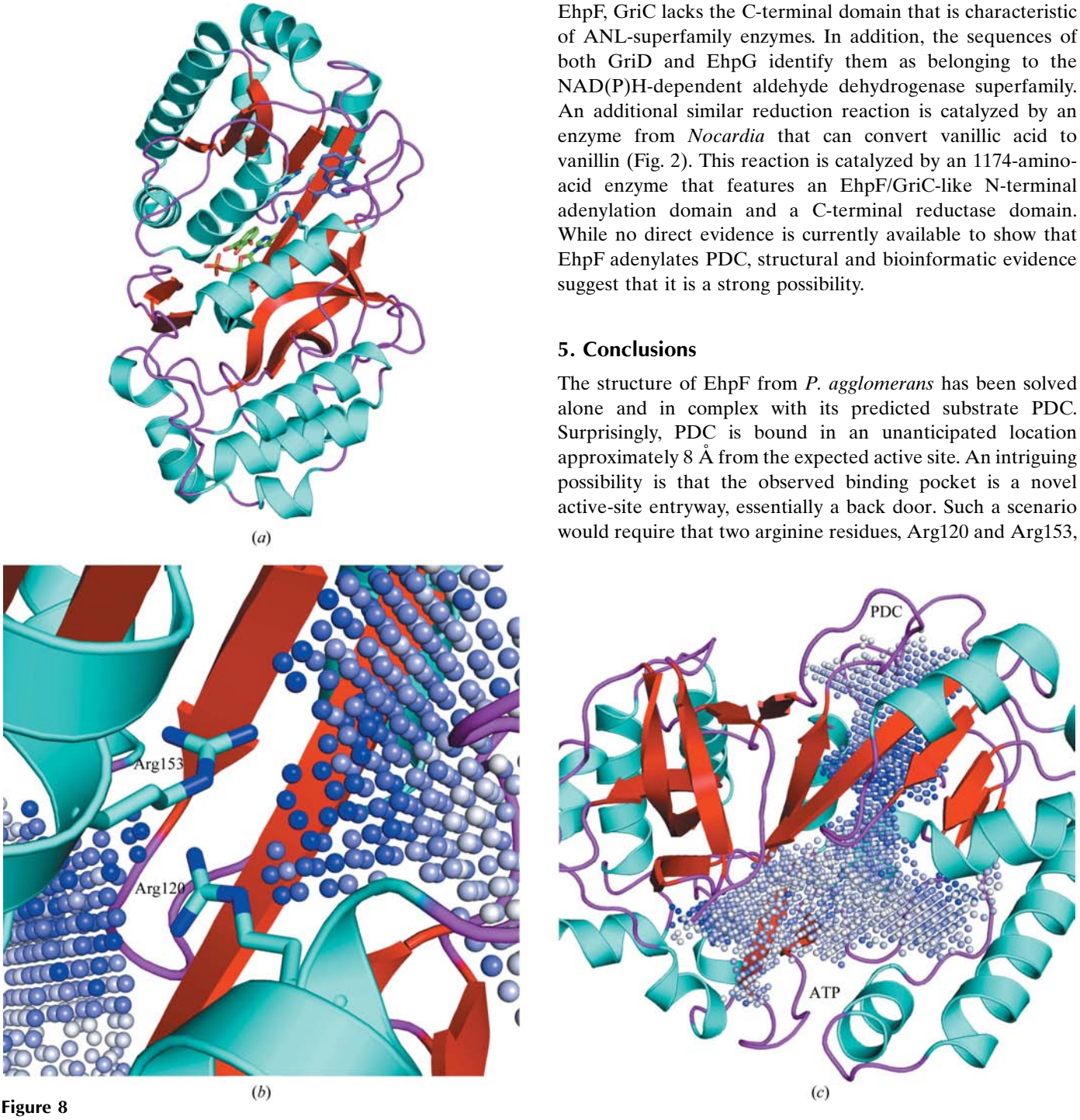


**Figure 8**
Arg120 and Arg153 separate the PDC-binding site from the ATP-binding/catalytic site of EhpF. (*a*) Cartoon illustrating the location of the observed PDC-binding site relative to the location of the canonical ANL-superfamily active site. The structures of PDC-bound EhpF and DhbE in complex with AMP and 2,3-dihydroxybenzoate (PDB code 1md9) were superimposed to generate the figure. (*b*) Cartoon representation illustrating the location of Arg120 and Arg153. Pockets were identified and illustrated using the *PyMOL* plugin *PocketPicker* (Weisel *et al.*, 2007). The orientation is the same as in (*a*). (*c*) Simulation of a gatekeeping function for Arg120 and Arg153. *PocketPicker* identifies a single large cavity running through the protein and connecting the PDC-binding and catalytic sites when Arg120 and Arg153 are truncated *in silico* to alanine residues.

act as gatekeepers and allow PDC to pass from its observed location through a channel occluded by these residues to the catalytic site (Fig. 8). Alternative explanations include PDC binding in a regulatory site or simply fortuitously in a location that has no biological significance.

EhpF does not display any activity *in vitro*, suggesting that another cellular factor is involved in catalysis. The observed amino-acid sequence homology between EhpF and GriC from *S. griseus* indicates that EhpF may be functionally similar to GriC. We predict that EhpF and EhpG comprise a two-component NAD(P)H-dependent arylcarboxylate reductase that converts PDC to 6-formylphenazine-1-carboxylate (Fig. 2). This prediction is consistent with the observed EhpF structure, the demonstrated functions of EhpF sequence homologues and the original data from Giddens and coworkers that showed PDC accumulation in *ehpF*-deficient strains of AGA-producing *P. agglomerans* (Giddens *et al.*, 2002).

This work was supported in part by a grant from the National Institute of Allergy and Infectious Diseases (AI067530) to JFP.

## References


Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Ahuja, E. G., Janning, P., Mentel, M., Graebsch, A., Breinbauer, R., Hiller, W., Costisella, B., Thomashow, L. S., Mavrodi, D. V. & Blankenfeldt, W. (2008). *J. Am. Chem. Soc.* **130**, 17053–17061.

Bains, J. & Boulanger, M. J. (2007). *J. Mol. Biol.* **373**, 965–977.

Bera, A. K., Atanasova, V., Robinson, H., Eisenstein, E., Coleman, J. P., Pesci, E. C. & Parsons, J. F. (2009). *Biochemistry*, **48**, 8644–8655.

Cruz, A. T., Cazacu, A. C. & Allen, C. H. (2007). *J. Clin. Microbiol.* **45**, 1989–1992.

Doublié, S. (2007). *Methods Mol. Biol.* **363**, 91–108.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Flood, M. E., Herbert, R. B. & Holliman, F. G. (1972). *J. Chem. Soc. Perkin Trans* **1**, 622–626.

Forneris, F. & Mattevi, A. (2008). *Science*, **321**, 213–216.

Giddens, S. R. & Bean, D. C. (2007). *Int. J. Antimicrob. Agents*, **29**, 93–97.

Giddens, S. R., Feng, Y. & Mahanty, H. K. (2002). *Mol. Microbiol.* **45**, 769–783.

Giddens, S. R., Houliston, G. J. & Mahanty, H. K. (2003). *Environ. Microbiol.* **5**, 1016–1021.

Gulick, A. M. (2009). *ACS Chem. Biol.* **4**, 811–827.

Hisanaga, Y., Ago, H., Nakagawa, N., Hamada, K., Ida, K., Yamamoto, M., Hori, T., Arii, Y., Sugahara, M., Kuramitsu, S., Yokoyama, S. & Miyano, M. (2004). *J. Biol. Chem.* **279**, 31717–31726.

Kawabata, T. (2003). *Nucleic Acids Res.* **31**, 3367–3369.

Kearns, L. P. & Mahanty, H. K. (1998). *Appl. Environ. Microbiol.* **64**, 1837–1844.

Kempf, H. F. & Wolf, G. (1989). *Phytopathology*, **79**, 990–994.

Maki, D. G., Rhame, F. S., Mackel, D. C. & Bennett, J. V. (1976). *Am. J. Med.* **60**, 471–485.

Matsaniotis, N. S., Syriopoulou, V. P., Theodoridou, M. C., Tzanetou, K. G. & Mostrou, G. I. (1984). *Infect. Control*, **5**, 471–477.

Mavrodi, D. V., Ksenzenko, V. N., Bonsall, R. F., Cook, R. J., Boronin, A. M. & Thomashow, L. S. (1998). *J. Bacteriol.* **180**, 2541–2548.

May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 12120–12125.

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.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K. & Kato, H. (2006). *Nature (London)*, **440**, 372–376.

Osman, K. T., Du, L., He, Y. & Luo, Y. (2009). *J. Mol. Biol.* **388**, 345–355.

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

Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.

Parsons, J. F., Greenhagen, B. T., Shi, K., Calabrese, K., Robinson, H. & Ladner, J. E. (2007). *Biochemistry*, **46**, 1821–1828.

Parsons, J. F., Song, F., Parsons, L., Calabrese, K., Eisenstein, E. & Ladner, J. E. (2004). *Biochemistry*, **43**, 12427–12435.

Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.

Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M. (2008). *Biochemistry*, **47**, 8016–8025.

Stevens, B. W., Lilien, R. H., Georgiev, I., Donald, B. R. & Anderson, A. C. (2006). *Biochemistry*, **45**, 15495–15504.

Stockwell, V. O., Johnson, K. B., Sugar, D. & Loper, J. E. (2002). *Phytopathology*, **92**, 1202–1209.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Suzuki, H., Ohnishi, Y. & Horinouchi, S. (2007). *J. Antibiot.* **60**, 380–387.

Uche, A. (2008). *South. Med. J.* **101**, 102–103.

Ulloa-Gutierrez, R., Moya, T. & Avila-Aguero, M. L. (2004). *Pediatr. Infect. Dis. J.* **23**, 690.

Venkitasubramanian, P., Daniels, L. & Rosazza, J. P. (2007). *J. Biol. Chem.* **282**, 478–485.

Weisel, M., Proschak, E. & Schneider, G. (2007). *Chem. Cent. J.* **1**, 7.

Wright, S. A. I. & Beer, S. V. (2005). *Proceedings of the First International Symposium on Biological Control of Bacterial Plant Diseases*, pp. 334–337.