The Crystal Structure of the *Escherichia coli* YfdW Gene Product Reveals a New Fold of Two Interlaced Rings Identifying a Wide Family of CoA Transferases*  

Arnaud Gruez, Véronique Roig-Zamboni, Christel Valencia, Valérie Campanacci, and Christian Cambillau†  

From the Architecture et Fonction des Macromolécules Biologiques, UMR 6098, CNRS and Universités Aix-Marseille I and II, 31 chemin J. Aiguier, F-13402 Marseille, Cedex 20, France

Because of its toxicity, oxalate accumulation from amino acid catabolism leads to acute disorders in mammals. Gut microflora are therefore pivotal in maintaining a safe intestinal oxalate balance through oxalate degradation. Oxalate catabolism was first identified in *Oxalobacter formigenes*, a specialized, strictly anaerobic bacterium. Oxalate degradation was found to be performed successively by two enzymes, a formyl-CoA transferase (frc) and an oxalate decarboxylase (oxc). These two genes are present in several bacterial genomes including that of *Escherichia coli*. The frc ortholog in *E. coli* is yfdW, with which it shares 61% sequence identity. We have expressed the YfdW open reading frame product and solved its crystal structure in the apo-form and in complex with acetyl-CoA and with a mixture of acetyl-CoA and oxalate. YfdW exhibits a novel and spectacular fold in which two monomers assemble as interlaced rings, defining the CoA binding site at their interface. From the structure of the complex with acetyl-CoA and oxalate, we propose a putative formyl/oxalate transfer mechanism involving the conserved catalytic residue Asp169. The similarity of yfdW with bacterial orthologs (~60% identity) and paralogs (~20–30% identity) suggests that this new fold and parts of the CoA transfer mechanism are likely to be the hallmarks of a wide family of CoA transferases. It has been proposed that the gut microorganisms, especially *O. formigenes*, maintain a transepithelial gradient of oxalate from the blood to the lumen of the intestines and consequently have an important symbiotic relationship with their hosts (6). The anaerobic mechanism of oxalate degradation involves, successively, an oxalate/formate antitransporter located in the cell membrane (7, 8), a formyl-CoA transferase (*frc*) and an oxalyl-CoA decarboxylase (*oxc*) (11). In this pathway the oxalate has first to be coupled to coenzyme A, a reaction catalyzed by formyl-CoA transferase; and a second step involves the oxalyl-CoA decarboxylase, yielding CO₂ and formyl-CoA. We have undertaken a structural genomics program aimed at solving the structures of *Escherichia coli* proteins of unknown function widespread among several Gram⁻ and Gram⁺ bacteria (12–14). Two of our targets, the yfdW and yfdU open reading frames, are closely related to the two *O. formigenes* enzymes. The *O. formigenes* formyl-CoA transferase (*frc*) is the closest protein neighbor of YfdW that has a known function, with both proteins sharing 61% identity (Fig. 1). yfdW also shares 50–60% sequence identity with orthologs from various species, all of which probably share the same function. These comparisons indicate that besides specialized organisms such as *O. formigenes*, several bacteria, including *E. coli*, may participate in oxalate detoxification. Other bacteria have *frc*-like genes sharing lower sequence identity with yfdW and coding for different CoA transferases: (R)-benzylsuccinate-CoA transferase (23% identity), (R)-phenyllactate-CoA transferase of *Clostridium difficile* (25%), (R)-carnitine-CoA transferase (24%), a putative cholate-CoA transferase (27%), and 2-methylacyl-CoA racemase (25%). However, no three-dimensional structures are available for any of them. Here we report on the YfdW structure in the apo-form in complex with a surrogate ligand, acetyl-CoA (AcCoA), and with both acetyl-CoA and the substrate, oxalate. The YfdW open reading frame product is a dimer displaying an amazing new fold of two interlaced rings, probably conserved among the orthologs and paralogs of the CoA transferase family. We have identified the formyl-CoA binding site as well as an oxalate resting site, and we propose a formate/oxalate transfer mechanism involving the catalytic nucleophile, Asp169. We believe that these features are hallmarks of a wide CoA transferase family.

**MATERIALS AND METHODS**

**Expression and Purification**—The subcloning and expression of the yfdW gene have already been described elsewhere (13, 14). In brief, the yfdW gene was amplified by PCR from *E. coli* K12 genomic DNA and

---

*This study was supported by a grant from the French Ministry of Industry (Agence Séquençage des Gènes) and was collaborated with the Information Génétique et Structural laboratory and the Aventis Company. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Fax: 33-491-164-536; E-mail: cambillau@afmb.cnrs-mrs.fr.

‡ To whom correspondence should be addressed. Fax: 33-491-164-536; E-mail: cambillau@afmb.cnrs-mrs.fr.
was subcloned into the pDest17 vector by recombination using the Gateway technology (Invitrogen). Expression was carried out using the Tuner(DE3)pLysS E. coli strain. The protein was purified on a nickel column followed by gel filtration on a Superdex 200 column in Hepes, 5 mM, NaCl, 150 mM, pH 7.5. The selenomethionine-labeled protein was expressed by blocking the methionine biosynthesis pathway (15) and was purified as described above.

Crystalization—Crystals of the SeMet-labeled apo-protein were grown by mixing 1 l of a protein solution at 11 mg/ml in 5 mM sodium-Hepes and 150 mM sodium chloride, with 500 nM of precipitant solution consisting of 0.9 mM ammonium phosphate and 0.1 mM sodium-Hepes, pH 7.5. Crystals grew within 5–15 days as rod-shaped prisms in the space group P212121 (Table I). Crystals contain 53% solvent (V_m = 2.7 Å^3/Da) assuming 2 monomers/asymmetric unit.

Crystals of the complex with AcCoA or with a mixture of AcCoA and oxalate were obtained by co-crystallization with 29% polyethylene glycol 6000 as precipitant in 0.1 M cacodylate, pH 6.5. Crystals belong to the space group P6_5 and contain 72% solvent (V_m = of 4.4 Å^3/Da) with a dimer in the asymmetric unit.

X-ray Structure Determination and Refinement—Selenomethionine-labeled crystals were soaked in a solution containing the reservoir buffer and 27–30% (v/v) glycerol and were subsequently flash-frozen. An x-ray data set of the SeMet apo-protein was collected at 100 K on beamline BM14 (ESRF, Grenoble) at 1.8 Å resolution. The binary and ternary complexes were collected at 2.0 and 2.2 Å resolution at ID14–EH4 and ID-29 (ESRF, Grenoble), respectively. Oscillation images were integrated, scaled, and merged using DENZO (16) and the CCP4 program suite (17) (Table I).

The apo-enzyme structure was solved at 1.8 Å by the single wavelength anomalous diffraction (SAD) method using the program SOLVE (18), which located 18 of the 22 selenium atom sites within the asymmetric unit and provided us with the initial phases. Solvent flattening was performed by RESOLVE (19) The good quality of the RESOLVE map made it possible to build 80% of the model automatically with wARP (20). This step was followed by manually building the missing parts of the protein with Turbo-Frodo (21) and by refinement with Refmac5 (22) and Arp-wARP (20). The TLS/restraint structure refinement of Refmac5 (22) converged to R and R_free of 16.5 and 19.3%, respectively, between 20.0 and 1.8 Å resolution (Table I).

The complexes with acetyl-CoA and the mixture of acetyl-CoA and oxalate were solved by molecular replacement with AMoRe (23) using the apo-protein dimer as a search model. The models were subjected to TLS/restraint refinements with Refmac5 (22) in the ranges of 20.0 to 2.0 and 20.0 to 2.2 Å resolution, followed by manual rebuilding with Turbo-Frodo (21) and water addition with Arp-wARP (20) (Table I). Protein geometry was assessed with PROCHECK (24). The figures were generated with VMD (25), GRASP (26), and Turbo-Frodo (21). The atomic coordinates and structure factors have been deposited with the Protein Data Bank at RCSB (accession codes 1PT7, 1PT9, and 1PT8 for the apo-enzyme and the binary and the ternary complexes, respectively).

RESULTS AND DISCUSSION

Completeness and Quality of the Models—The crystal structure of the product of the yfdW gene was determined at 1.8, 2.0, and 2.2 Å resolution for the apo-enzyme and the binary and ternary complex, respectively. In each of the three cases, a dimer is contained in an asymmetric unit. The final apo-protein model consists of 830 residues (415 residues/monomer of the 416 in sequence), 8 phosphate ions, 2 glycerol molecules, and 591 water molecules. The model of the binary complex contains 830 residues, 2 AcCoA molecules, and 708 water molecules, and the ternary complex contains, in addition, 2 oxalate molecules and counts 453 water molecules.

A total of 92.4, 92.9, and 91.2% of the residues are located in the most favorable region of the PROCHECK (24) Ramachandran plot of the apo-protein and the binary and ternary complex, respectively (Table I). The apo-protein has no residues in the disallowed region, whereas in the complex structures Leu^363 is located in the disallowed region. Its good electron density map indicates that this high energy conformation is not artifactual, however. The unfavorable interactions of the Leu^363 side chain are likely to be outweighed by a favorable hydrogen bond between the carbonyl group of Pro^362 and the nitrogen atom of Arg^154 involved in a γ-turn.

Overall Structure of the Monomers—Each monomer can be divided into four distinct structural domains featuring an ellipsoidal ring structure (Fig. 2A). The large hole observed at the center of the molecular surface is filled by the second monomer (see next section) (Fig. 2, B and C). A first large Rossman fold
domain is located at the N terminus (Fig. 2A, bottom, orange). Progressing anticlockwise (Fig. 2A), the small domain 2 (yellow) bridges domain 1 with domain 3 (blue) and is followed by the split C-terminal domains 4a and 4b (green). The N- and C-terminal regions are close together and form a globular superstructure closing the ring.

Domain 1 (residues 1–139 and 166–200) is formed by 6 parallel β-strands (β1–β6) flanked by 6 helices on either sides (Fig. 2A). Domain 2 (residues 140–165 and 201–248) contains four helices and coil/turn regions. Domain 3 encompasses residues 250–339. Strands β7–β9 form an antiparallel β-sheet covered on its outwards face by helices α11 and α14. The C-terminal domain 4 can be split into two subdomains. Subdomain 4a contains a two-stranded antiparallel β-sheet (β10 and β11) and helices α15 and α16 (residues 339–383). Strand β11 interacts with helix α3 of the N-terminal region. Subdomain 4b comprises a coil-and-turn region with two C-terminal α-helices, α17 and α18 (residues 384–416), which close the ring and are tightly packed against domain 1. The arrangement of domain 1 against domain 4b results in a globular substructure.

The interactions between the N- and C-terminal regions participating in the ring closure involve 16 hydrogen bonds and two salt bridges. Residues participating in the intramolecular interactions are located between Arg46 and Glu81 and between Table I

| Data collection and refinement statistics | Apo-protein | Acetyl-CoA | Acetyl-CoA/Oxalate |
|------------------------------------------|-------------|-----------|-------------------|
| Space group                              | P2₁,2₁,2₁   | P6₂       | P6₂               |
| Unit cell (Å)                            | 60.9, 118.6, 136.3 | 146.9, 146.9, 129.5 | 146.8, 146.8, 129.9 |
| Wavelength (Å)                           | 0.98000     | 0.9330    | 0.979257          |
| Resolution (Å)                           | 1.80 (1.89–1.8)⁺ | 2.0 (2.11–2.0)⁺ | 2.2 (2.32–2.2)⁺  |
| Completeness (%)                         | 96.7 (98.7)⁺ | 99.0 (97.5)⁺ | 96.7 (98.3)⁺     |
| Rₚ (%)                                   | 4.8 (26.8)⁺ | 7.8 (26.5)⁺ | 11.4 (23.4)⁺     |
| Rₚ (%)                                   | 6.9 (23.5)⁺ |           |                   |
| Refinement                               |             |           |                   |
| Resolution range (Å)                     | 20–1.8     | 20–2.0    | 20–2.2            |
| R/Bₚ (%)                                 | 15.9/19.1⁺ | 14.6/18.8⁺ | 17.7/20.4⁺       |
| r.m.s.d. bonds (Å)                       | 0.02       | 0.034     | 0.016             |
| r.m.s.d. angles (Å)                      | 1.7        | 2.0       | 1.6               |
| Mean B-factor (Å²)                       | 19.1       | 20.5      | 42.4              |
| PROCHECK zones distribution (%)          | 92.4/7.0/0.5/0.0 | 92.9/6.5/0.3/0.3 | 91.2/8.2/0.4/0.3 |

⁺ Values for the last resolution shell are in parentheses.

Fig. 2. Crystal structure of the YdfW gene product. A, view of the monomeric secondary structure color-identified by domain: 1, orange; 2, yellow; 3, blue; and 4, green. The N and C termini are identified by red and blue circles, respectively, and the helices are numbered 1–18. B, view of the dimer in the binary complex. The first subunit molecular surface is colored according to surface curvature with convex surfaces in green and concave surfaces in white, and the second subunit, represented as a white tube, passes through the hole of the first monomer. The AcCoA molecule is partially visible along the monomers interface; its sulfur atom is hidden in a surface crevice. The second AcCoA molecule is partially visible through the hole of the first monomer. C, schematic representation of the dimer-interlaced rings. The domains have been approximately located.
Structure of the E. coli YfdW Gene Product

Arg$^{364}$ and Ala$^{415}$ at the N- and C-terminal ends, respectively. The salt bridges are established between Asp$^{51}$ and Arg$^{364}$ and between Arg$^{68}$ and Glu$^{399}$.

**Dimer Association**—The YfdW dimer has overall dimensions of 64 $\times$ 56 $\times$ 48 Å (Fig. 2B). The dimeric nature of YfdW was observed in solution by gel filtration and dynamic light scattering (14). Superimposition of the native main chains A and B leads to an r.m.s.d. of 0.14 Å, indicating that both monomers have identical structures. The YfdW dimer is formed by two elongated rings that interpenetrate as the two first rings of a chain (Fig. 2). As a consequence of this topology, the dimer is tightly bound by an extensive interface of 3540 Å$^2$ of a total surface of 13,300 Å$^2$/monomer, accounting for 27% of the surface area of each of the monomers. Indeed, the biological relevance of the dimer is rather obvious considering its structure.

The interactions between each monomer involve numerous van der Waals and hydrophobic contacts, which, at the dimer level, account for a well packed hydrophobic core. The dimer is further stabilized by 36 direct hydrogen bonds/monomer spanning residues 128 to 378. Although the majority of the contacts involve residues located in coil or loop regions, some of them are made by residues located in secondary structure elements. Helices $\alpha$9 and $\alpha$10, which protrude from domains 1, yield 13 hydrogen bonds and play a major role in the intermolecular contacts. They interact with strand $\beta$11 through a network of 7 hydrogen bonds involving residues Arg$^{209A}$ and Gln$^{216A}$ of a monomer and Leu$^{368B}$, Thr$^{369B}$, and Val$^{370B}$ of the other monomer. In addition Arg$^{209A}$, a key residue of the dimer architecture, is involved in two additional hydrogen bonds with Gln$^{248B}$ and Met$^{62B}$ located in helix $\alpha$1 and $\alpha$3, respectively. Helices $\alpha$6 and $\alpha$7, which are close to helix $\alpha$9, are also involved in intermolecular contacts with loop and coil regions of the C terminus of the other monomer. Furthering these interactions, residues Thr$^{197A}$ and Thr$^{195A}$ of strand $\beta$6 interact with Lys$^{375B}$ and Ser$^{377B}$ and provide, along with Phe$^{376B}$, an extra $\beta$-strand to the $\alpha/\beta$ structure.

**Acetyl-CoA and Acetyl-CoA-Oxalate Complexes**—Because formyl-CoA is an unstable molecule, the YfdW gene product was co-crystallized with CoA and with AcCoA, a surrogate ligand of formyl-CoA. Although crystallization with AcCoA yielded crystals containing the cofactor, CoA could never be identified. Electron density consistent with bound AcCoA molecules appears at two sites of the dimer, one in each subunit, separated by a distance of $\sim$25 Å.

The apo and AcCoA complex structures can be superposed with an r.m.s.d. of 0.8 Å. This deviation is significantly above background because monomers A and B of the native and complex dimer structures display r.m.s.d. values of only 0.15 Å. The same low value (0.14 Å) is also observed when the AcCoA and the AcCoA/oxalate structures are superimposed, indicating that they are identical and that AcCoA is the trigger of protein conformational change, whereas oxalate has no global effect.

The deviations between the apo and the binary or ternary complexes are observed in several places: between residues 97 and 118, a surface helix, with amplitudes of 1–2 Å; between residues 247 and 251, the GGGQQP motif (Fig. 1), with amplitudes up to 3 Å; then in a long segment between residues 272 and 335 with small deviations of 0.5–1 Å; finally, at the C terminus, residues 397–416, with the largest deviations up to 3.5 Å. The conformational change of the 97–118 and 247–251 segments is clearly linked to AcCoA binding, as favorable contacts are established between these segments and AcCoA. AcCoA displays a Z-shaped conformation. It is located at the interface of both monomers (Fig. 2B) in a crevice of the open twist $\alpha/\beta$ structure with its adenine base and ribose sugar pointing toward the solvent, whereas the thioester group is buried deep inside the protein, 10 Å from the protein surface. AcCoA rides the $\beta$1 and $\beta$4 strands, with the walls of the cavity formed by loops between strands and helices $\beta$2–$\alpha$2, $\beta$3–$\alpha$4, and $\beta$5–$\alpha$7, and between helices $\alpha$5–$\alpha$6 and $\alpha$1–$\alpha$7.

AcCoA and CoA have been found to be flexible molecules as illustrated by the large number of different conformations observed in protein complexes (27). This property is reflected here, as the electron density map reveals the coexistence of two discrete conformations for the carbonyl group of the acetyl group of the molecules. The acetyl group of AcCoA is likely to interact with catalytically relevant residues. Its two conformers are located roughly on either side of Asp$^{169}$, bringing the acetate moiety into contact with it (3.0–3.4 Å). In one of the conformers, the carbonyl of acetate binds Glu$^{17}$ NH, whereas in the other, Glu$^{140}$ NH is the hydrogen bond provider. In this latter case, the carbonyl occupies the site of a phosphate ion in the apo-structure. Active site side chains moved upon complexation, providing closer contacts with AcCoA. This is the case for tyrosines 139 and 59 (1- and 2-Å deviation at OH) and to a lesser extent for other residues.

The binary and ternary complex structures are virtually identical; the main chains are superimposed, and the active site side chains as well as the AcCoA molecules occupy identical positions. The oxalate molecule is clearly defined in the electron density map of the ternary complex, although with high B-factors (79 versus 42 Å$^2$ for the protein). It is located 8.2 and 10.0 Å away from each carbonyl group of the two conformations.
of AcCoA. This position is not favorable for catalysis and should be considered as a "resting" position along the reaction pathway. Oxalate occupies the position of Gly\textsuperscript{249} (from the other monomer) in the structure of the apo-enzyme. Gly\textsuperscript{249} belongs to a stretch of residues (\textsuperscript{246}GGGGQQP\textsuperscript{253}), which moves considerably in the binary and ternary complex with respect to the apo-enzyme. This loop is kept in position by a hydrogen bond between Tyr\textsuperscript{59} OH and Gly\textsuperscript{248} carbonyl. The relocation of this loop is the major contribution of active site reshaping in the complex structures. The oxalate resting site is formed, therefore, on one side by the relocated Gly\textsuperscript{247}–Pro\textsuperscript{251} loop of the other monomer, and on the other side by side chains of Leu\textsuperscript{49} and Gln\textsuperscript{48} with which it forms a hydrogen bond (Fig. 3A). One carboxyl group of the oxalate is directed toward the bulk solvent and the other toward a channel containing a water molecule and abutting AcCoA. This channel is large enough for the oxalate molecule to diffuse along it. An active position of oxalate can readily be modeled 6.0 Å away from the experimental resting position, in contact with the AcCoA carbonyl group. The remote carboxyl group of the active oxalate occupies a molecular "hot spot," populated by a water molecule in the ternary complex and by a phosphate ion in the apo-enzyme. The active oxalate modeled is located in a cavity lining Tyr\textsuperscript{139} on one side and comprising Val\textsuperscript{146}, Gln\textsuperscript{48}, and Tyr\textsuperscript{59} on the other side. Its remote carbonyl can establish hydrogen bonds with Gln\textsuperscript{146}, Tyr\textsuperscript{139}, and Gln\textsuperscript{250B} (Fig. 3A). Its proximal carboxyl group is poorly locked instead; its closest potential hydrogen bond provider is Tyr\textsuperscript{59}. It is very close to the acetyl moiety of AcCoA, this latter being sandwiched between oxalate and Asp\textsuperscript{169} (Fig. 3A).

A Putative Catalytic Mechanism—The description of the active site, thanks to the ternary complex, makes it possible to postulate the reaction mechanism. The global reaction can be described by the equation CoA-CHO + COO-COO → CoA-CO-COO + HCO\textsubscript{2}. During catalysis, CoA-CHO should give the formyl moiety to an acceptor, liberating a CoA molecule. In a subsequent step, the CoA should perform a nucleophilic attack on an activated oxalate, yielding CoA-oxalate. The complex structure suggests that Asp\textsuperscript{169} plays a pivotal role in catalysis as a nucleophile.

In a first step, Asp\textsuperscript{169} may perform a nucleophilic attack on the carbonyl of the formyl-CoA, which after the breakage of the thioester bond would release CoA, yielding an Asp-formyl adduct. The reaction intermediate forms an oxanion, which should be stabilized by an appropriate molecular arrangement called oxanion hole (28) (Fig. 3B, top). Depending on the AcCoA conformer, the oxanion hole might be formed by Gln\textsuperscript{140} NH or by Gln\textsuperscript{48} and Ser\textsuperscript{14} NH atoms. This latter oxanion hole geometry with two NH groups seems to be preferred in nature (29). Oxalate is positioned in the structure at a resting position far from the acetyl group of AcCoA, whereas the modeled active oxalate position should be close to the formyl group of the Asp-formyl adduct. Oxalate should thus slide along the active site, establish a contact with Tyr\textsuperscript{50} at a position adequate to form a contact with Tyr\textsuperscript{59} (Fig. 3B, bottom). Indeed, each of these steps is probably driven by electrostatics and requires proper positioning of the reactive groups. For example, at step 2, an oxalate attack on the formyl carbonyl would stop the reaction, as would an attack of CoA on the Asp carbonyl at step 3. Therefore, in addition to the catalytic Asp\textsuperscript{169}, Tyr\textsuperscript{59} is well placed to properly position oxalate in the active site and help extract it from its resting position, identified in the ternary complex structure.

Contrary to the nucleophilic residue in catalytic triads, Asp\textsuperscript{169} has no direct strong hydrogen bond with partners. A weak hydrogen bond is established with the side chain of Gln\textsuperscript{17} (3.3 Å, poor directionality), which might contribute somewhat, however, to orientation of the carbonyl. More interesting is the close proximity (3.2 Å) between two charged residues, Asp\textsuperscript{169} and Gln\textsuperscript{48} (conserved). Gln\textsuperscript{48} is kept in position by a hydrogen bond with a Gln\textsuperscript{146} side chain (also conserved). The closeness of Asp\textsuperscript{169} and Gln\textsuperscript{48} may dramatically enhance the density of electrons on Asp\textsuperscript{169} and therefore increase its nucleophilicity.

Conclusion—The two interleaved rings structure of the YfdW gene product must provide an incredible stability to the enzyme dimer, probably larger than that observed in domain-swapped dimer structures (30). As in the latter case, the folding of both monomers should rely on a tightly concerted process.

The conservation of the catalytic and active site residues supports the likelihood of a conserved mechanism among all formyl-CoA transferases. Considering the sequence identity with paralogs genes coding for other CoA transferases, it is likely that the interleaved ring structure is the hallmark of a wide family of CoA transferases.

Acknowledgments—The ESRF is gratefully acknowledged for beam-time allocation. We thank the structural genomics team of the AFMB laboratory for technical assistance.

Note Added in Proof—The structure of the formyl-CoA transferase from O. formigenes was published recently by Ricagno et al. (Ricagno, S., Jonsson, S., Richards, N. & Lindqvist, Y. (2003) EMBO J. 22, 3210–3219).