Published in final edited form as:
J Mol Biol. 2007 September 14; 372(2): 456–469. doi:10.1016/j.jmb.2007.06.063.

Structure of an Amide Bond Forming F_{420}:γ-glutamyl Ligase from *Archaeoglobus Fulgidus* - A Member of a New Family of Non-ribosomal Peptide Synthases

B. Nocek[1], **E. Evdokimova**[1,2], **M. Proudfoot**[1,2], **M. Kudritska**[1,2], **L. L. Grochowski**[3], **R. H. White**[3], **A. Savchenko**[1,2], **A. F. Yakunin**[2], **A. Edwards**[1,2], and **A. Joachimiak**[1,4]

1*Midwest Center for Structural, Genomics and Structural, Biology Center, Biosciences, Argonne National Laboratory, 9700 South Cass Avenue, Building 202, Argonne, IL 60439, USA*

2*Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada M5G 1L6*

3*Department of Biochemistry, Virginia Polytechnic, Institute and State, University, Blacksburg, VA 24061-0308, USA*

4*The University of Chicago, Department of Biochemistry and Molecular Biology, University of Chicago, 920 E. 58th St., Chicago, IL 60637, USA*

Abstract

F_{420} is a flavin-like redox-active coenzyme commonly used by archaea and some eubacteria in a variety of biochemical reactions in methanogenesis, the formation of secondary metabolites, the degradation of nitroaromatic compounds, activation of nitroimidazofurans, and F_{420}-dependent photolysis in DNA repair. Coenzyme F_{420}-2 biosynthesis from 7,8-didemethyl-8-hydroxy-5-deazariboflavin (Fo) and lactaldehyde involves six enzymatic steps and five proteins (CofA, CofB, CofC, CofD, and CofE). CofE, a F_{420}-0:γ-glutamyl ligase, is responsible for the last two enzymatic steps; it catalyses the GTP-dependent addition of two L-glutamate residues to F_{420}-0 to form F_{420}-2. CofE is found in archaea, the aerobic actinomycetes, and cyanobacteria. Here, we report the first crystal structure of the apo-F_{420}-0:γ-glutamyl ligase (CofE-AF) from *Archaeoglobus fulgidus* and its complex with GDP at 2.5 Å and 1.35 Å resolution, respectively. The structure of CofE-AF reveals a novel protein fold with an intertwined, butterfly-like dimer formed by two-domain monomers. GDP and Mn^{2+} are bound within the putative active site in a large groove at the dimer interface. We show that the enzyme adds a glutamate residue to both F_{420}-0 and F_{420}-1 in two distinct steps. CofE represents the first member of a new structural family of non-ribosomal peptide synthases.

Keywords

γ-glutamyl ligase; coenzyme F_{420} biosynthesis; amide bond-forming enzyme; metal-dependent; new fold

Introduction

Most enzymes exploit small molecules such as metals and cofactors, to enhance their catalytic properties.[1] These molecules can aid catalysis and expand the range of available enzymatic functions. In keeping with their central importance, cells dedicate a large set of genes to

Correspondence to: A. Joachimiak, andrzejj@anl.gov.

synthesize and regenerate cofactors at a considerable metabolic cost. Many important and broadly utilized cofactors, including nicotinamide adenine dinucleotide, coenzyme A, pyridoxal phosphate, flavin adenine nucleotide, biotin, 5′-deoxyadenosyl cobalamin, thiamine pyrophosphate, tetrahydrofolate, S-adeno-sylmethionine, iron clusters and metal ions have been extensively characterized.

However, there are still a large number of newly discovered cofactors (e.g. nitrogenase P-cluster and FeMo-cofactor,[2,3] the hydrogenase H-cluster,[4] and (6*R*)-5,6,7,8-tetrahydro-L-biopterin (BH_4), a cofactor of nitric oxide synthetase[5]) and some post-translational modifications of proteins that await more detailed characterization.[6] Recent studies of prokaryotic organisms have revealed important insights into the remarkable chemistry and mechanisms of biosynthesis of cofactors such as coenzyme M, methanofuran, methanopterin, coenzyme B and coenzyme F_{420}.[7–13] Here, we focus on coenzyme F_{420}-2 (F_{420}-0-glutamyl-glutamate) biosynthesis (Figure 1). F_{420} was first discovered in 1960 in myco-bacteria and 12 years later was isolated from Methanobacterium.[14–17]

F_{420} coenzyme has an important role in the metabolic pathways of methanogenic archaea, bacteria and some eubacteria such as *Streptomyces, Rhodococcus, Nocardioides, Mycobacterium* spp., and their relatives.[18] The coenzyme has been recognized to participate in a variety of biochemical reactions, including formation of different secondary metabolites, the degradation of nitroaromatic compounds, activation of nitroimidazofurans, and is involved in F_{420}-dependent photolyase functions in DNA repair.[19–23]

Despite its structural resemblance to flavin-like coenzymes, F_{420} is functionally similar to the nicotinamide adenine dinucleotide cofactors NAD and NADP. F_{420} is an obligate two-electron acceptor, which is involved only in hydride transfer, with redox potential in the range of −340 mV to −350 mV.[11] Upon reduction, 1,5-dihydro-coenzyme F_{420} is formed with a prochiral center at C5. Thus far, only a few coenzyme-F_{420}-dependent enzymes have been discovered and all of these enzymes have been shown to be F_{420} *si*-face stereo-specific.[24–26] Different forms of the F_{420} cofactor have been observed including γ-glutamyl linked F_{420} and γ-glutamyl F_{420} capped with a single α-linked L-glutamate (such as α-F_{420}-3). In *Archaeoglobus fulgidus*, γ-glutamyl linked F_{420} is the predominant form with F_{420}-2 being the end product, whereas in *Methanocaldococcus jannaschii* (formally *Methanococcus jannaschii*), the α-F_{420}-3 is the major form. The α-F_{420}-3 is produced by CofF from γ-F_{420}-2.[27]

Studies of the coenzyme F_{420} biosynthesis pathway revealed that the F_{420}-0 (L-lactyl phosphodiester of 7,8-didemethyl-8-hydroxy-5-deazariboflavin) intermediate is synthesized in four steps requiring four enzymes (CofA, CofB, CofC, CofD) (Figure 1). In the first three steps, lactaldehyde is converted to L-lactate by CofA[28] and to 2-phospho-L-lactate by CofB. 2-Phospho-L-lactate is then converted by CofC to lactyl (2) diphospho-(5′) guanosine. In the next step, CofD transfers 2-phospholactate moiety from lactyl (2) diphospho-(5′) guanosine to 7,8-dide-methyl-8-hydroxy-5-deazariboflavin (Fo) forming the F_{420}-0. The fifth and sixth steps require the GTP-dependent additions of two successive L-glutamate residues to the F_{420}-0 and F_{420}-1 to form F_{420}-2. Both of these reactions are catalyzed by the enzyme F_{420}-0:γ-glutamyl ligase (Figure 1), also known as CofE.

Sequence analysis revealed ~70 homologs of F_{420}-0:γ-glutamyl ligases found in the methanogenic archaea and Gram-positive bacteria with high G+C content.[29] Biochemical analysis of CofE from *M. jannaschii* showed that the enzyme forms dimers in solution and that the addition of multiple γ-linked L-glutamates requires GTP, divalent (Mg^{2+}, Mn^{2+}) and monovalent (K^+) metal cations. [30] The CofE protein was selected as one of the targets by the Midwest Center for Structural Genomics, since it was predicted to belong to a structurally uncharacterized family of enzymes that shows no sequence similarity to any known protein.

Here, we report the crystal structures of the apo- and GDP-bound forms of F_{420}-0:γ-glutamyl ligase (CofE-AF) from *A. fulgidus* strain DSM 4304. The structures were determined and refined at 2.5 Å and 1.35 Å resolutions, respectively. The structure of the CofE-AF reveals a novel protein fold, with two-domain monomers forming tight butterfly-like dimers. The γ-glutamyl ligase activity of CofE has been confirmed in a biochemical experiment with the F_{420}-0 substrate and a putative active site has been identified at the dimer interface. The GDP, which is a co-product of the reaction, and two manganese cations have been found at this site.

Results and Discussion

Recombinant CofE-AF shows F_{420}-0:γ-glutamyl ligase activity *in vitro*

The CofE-AF protein coded by the gene AF2256 (gi:2648269) shares 46% sequence identity with the biochemically characterized F_{420}-0:γ-glutamate ligase CofE (MJ0768) from *M. jannaschi*.[30] The latter enzyme catalyzes the addition of two L-glutamate residues to F_{420}-0 to form the cofactor F_{420}-2 (Figure 1).

CofE-AF, under the reaction conditions reported for *M. jannaschii* CofE (CofE-MJ), demonstrated high F_{420}-0:γ-glutamate ligase activity (Figure 2). In the presence of 0.5 μM F_{420}-0 at 50 °C, the specific activity of CofE-AF was 22.7 nmol/min mg^{-1} protein (calculated from the decrease in the F_{420}-0 concentration) which is approximately ten times higher than the V_{max} calculated for CofE-MJ. In the first step of the reaction (first 10 min), CofE-AF predominantly used F_{420}-0 as a substrate and accumulated the F_{420}-1 product (Figure 2). The addition of the second glutamyl residue to F_{420}-1 started when its concentration became close to that of F_{420}-0. Thus, similar to CofE-MJ (MJ0768), the protein encoded by AF2256 catalyzes the addition of one glutamate residue to F_{420}-0 and one glutamate residue to F_{420}-1 in two independent steps, confirming the function of CofE-AF as a F_{420}-0:γ-glutamate ligase.

The solution of affinity chromatography-purified recombinant CofE-AF expressed in *Escherichia coli* displayed a yellow color and its absorption spectrum showed the presence of two peaks at 364 nm and 454 nm (and two shoulders at 435 nm and 486 nm) (data not shown). The yellow color disappeared upon the addition of dithionite, a strong chemical reductant, and the absorbance at a range of 400–500 nm decreased markedly. The yellow color was restored upon exposure of the solution to air (data not shown). Very similar absorption spectra were reported for FMN or FAD-containing proteins,[31,32] suggesting the presence of such cofactors in CofE-AF. HPLC analysis of small molecules extracted from CofE-AF revealed the presence of pterin, FAD/FMN, riboflavin as well a small amount of lumizine, an intermediate of riboflavin biosynthesis. Thus, it appears that ~10% of the *E. coli*-expressed CofE-AF molecules retained flavin-like (and structurally similar to F_{420}-0 substrate) metabolites during purification. These small molecules are not observed in the crystal structure either due to low occupancy or due to the fact that they were not incorporated into the crystal.

Crystal structure of CofE-AF and quality of the models

The structure of apo-CofE-AF was determined by synchrotron-based X-ray crystallography using the single anomalous diffraction (SAD) phasing method from a seleno-L-methionine (SeMet)-substituted enzyme to 2.5 Å resolution. The structure was refined using REFMAC,[33] yielding a crystallo-graphic R_{work} of 19.4% and R_{free} of 25.5%. The data collection and refinement statistics are summarized in Table 1. The final model contains a dimer in the asymmetric unit and accounts for 233 out of 249 possible amino acid residues and 49 water molecules per monomer. The electron density map was generally of good quality; most of the atoms could be positioned, except the N-terminal methionine of chain B, which was only partially visible and was modeled as alanine, and two disordered loops in both chains (177–

188 and 208–211). Also, cloning artifacts QGH at the N terminus and GS at the C terminus were not observed in the density maps.

Subsequently, the crystals of CofE-AF with the co-product GDP and Mn^{2+} cations were obtained and the structure of the complex was determined using molecular replacement. The structure was refined with REFMAC[33] to nearly atomic resolution (1.35 Å) (Table 1) with R_{work} of 16.0% and R_{free} of 19.0%. The resulting electron density maps were of outstanding quality, allowing modeling of all 248 residues (except N-terminal methionine residues, which were disordered in this crystal form) in both chains A and B. In this case, the N and C-terminal cloning artifacts were not visible in the density maps. The final model contains two molecules of GDP, four Mn^{2+} cations, three molecules of ethylene glycol (EG), four acetate anions, one molecule of glycerol and 586 water molecules.

Overall structure of the monomer

The structure of CofE described below is based on the near-atomic resolution (1.35 Å) model of CofE-AF in complex with GDP and Mn^{2+}. The CofE-AF is a homodimer with 249 amino acid residues per monomer. The overall shape of the dimer resembles a butterfly with each monomer contributing to a butterfly wing. A representation of the dimer is shown in Figure 3 (a). Structurally, the monomer could be further sub-divided into two domains: a large N-terminal domain (residues 1–48 and 121–232) and a smaller C-terminal domain (residues 49–120 and 233–249) (Figure 3(b)). The N-terminal domain consists of seven β-strands and four α-helices (Figure 3(b)). The β-strands are organized into a highly twisted, five-stranded β-sheet (resembling nearly half of a cylinder) and a 25 Å long hairpin structure (residues 174–194) that crosses over to the other subunit. The top three strands (β1, β12, and β2) of the β-sheet are antiparallel with respect to each other, while the bottom two strands (β8, β9) form an antiparallel hairpin motif. The β-sheet is flanked by three α-helices on the concave side (surface-exposed) and one α-helix on the convex site (dimer interface). The two-stranded hairpin (β10 and β11) is connected *via* a ~15 Å hook-shaped loop that extends and coils with the opposite subunit of the dimer. The small C-terminal domain consists of five β-strands, three α-helices and two 3_{10} helices. The central feature of this domain is an anti-parallel, three-stranded β-sheet flanked by two helices, a small two-stranded linker at the N-terminal side of the domain, and an α-helix at the C-terminal part of the domain.

Organization of CofE homodimer

Size-exclusion chromatography (data not shown) suggests that CofE-AF is a dimer in solution. This same oligomeric state was reported for *M. jannaschii* CofE.[30] The crystal structure of CofE-AF also strongly supports this observation. The dimeric state is well accounted for by tight packing and extensive interactions between the subunits observed in the crystal structures of both the apo-protein and the complex with GDP. In both cases, the asymmetric unit of the crystal contains two monomers (A and B) related by a 2-fold, non-crys-tallographic axis. In the GDP complex structure, monomers superimpose with a root-mean-squared deviation (RMSD) of 0.35 Å over all 248 C^{α} atoms.[34] As expected, the largest differences are observed in the highly flexible hook-like loop regions (residues 177–178 and 184–188 of both monomers). These residues are disordered in the apo-CofE structure.

The CofE-AF dimer interface is extensive and buries an accessible surface area (*ASA*) of 3670 $Å^2$ per monomer, representing 27% of the monomer's surface, and falls within the highest range for a complex formation (interface score of 1 as assessed by PISA server, where the scores range from 0 to 1 as interface relevance to complex formation increases).[35] In the N-terminal domain, the five-stranded β-sheet with four turn helix 6 of one subunit interacts with the β-sheet and equivalent helix 6 of the opposing subunit forming a ten-stranded anti-parallel β-sheet barrel at the dimer interface. The β-sheet barrel encompasses both α-helices forming a

unique design. On a broader level, a comparable fold can be observed in a so-called hotdog fold,[36] where typically a seven-stranded antiparallel β-sheet wraps around a five turn α-helix. Although, α-helices occur frequently on the concave surface of a β-sheet, particularly in the hot-dog fold and its variants,[37] to the best of our knowledge this is the first example of dimerization involving the formation of a two helix bundle wrapped by an antiparallel β-sheet barrel. This motif strengthens the dimer through numerous hydrophobic and some hydrophilic interactions. The antiparallel helices 6A and 6B make dipole–dipole interactions with each other at an angle of 30°. Most significantly, these helices contribute to a hydrophobic core through interactions with the concave surface of the β-barrel. The major feature of this surface comprises an array of three pairs of valine residues from each monomer: V3 and V5, the conserved V216 and the semi-conserved V218, V35, and V37. In addition, two symmetry-related hairpins interact tightly on the side of the β-barrel.

At the center of the dimer interface the two subunits interact through long loops that swap to the neighboring subunit and are involved in interactions with GDP (this region is disordered in the absence of GDP in the apo-CofE structure). Because of the interaction of the loops with the nucleotide and their close proximity to the presumed active site, we speculate that this interaction is important in dimer stabilization and in catalysis (see Nucleotide and metal ion binding site section, below). An interaction of short helices 7A and 7B provide additional contacts within the dimer interface.

At the C-terminal end, the CofE-AF dimer appears to be further stabilized by three inter-subunit disulfide bonds: C155(A)/C155(B), C244(A)/C248(B), and C248(A)/C244(B). In both crystal structures the side-chains of cysteine residues are observed in double conformations, suggesting partial disulfide bonds. Thus, it is not clear if the disulfide bonds formed as a result of oxidation during the purification and crystallization of the enzyme or if they play a functional role in the stabilization of the CofE-AF dimer. Nevertheless, the Cys residues are in favorable positions to make disulfide bonds. Sequence alignment analysis indicates that these residues are not conserved in the CofE family (Figure 4). Thus, we believe they may be structurally important in some species, for example in thermophiles such as *A. fulgidus*. Apart from these disulfide bridges, the interface is stabilized by almost 30 direct and solvent-mediated hydrogen bonds. There are eight inter-subunit salt-bridges between R154(A)/D196(B), R157(A)/D196 (B), R158(A)/E238(B), and R179(A)/D211(B), and their corresponding counterparts. We can conclude that as a result of these interactions and the extended dimer interface, the CofE-AF dimer is well adapted to function at elevated temperatures.[38–40]

CofE-AF structure reveals a novel protein fold

A search for structural homologues of CofE-AF in the Protein Data Bank (PDB) using the DALI server was performed for the entire monomer (248 residues), as well as for the two individual domains.[41] The search showed no significant structural similarity with any known protein. The closest similarity was diaminopimelate epimerase from *Hemophilus influenzae* (PDB code 1BWZ, RMSD of 3.4 Å and Z-score of 3.6).[42] This enzyme shares only 8% sequence similarity with CofE and shows only slight overlap (80 out of 249 residues) between the structural motif α3, β12, β11 and α4 of diaminopimelate epimerase and α5 β8 β9 α6 corresponding elements of the N-terminal domain of CofE-AF. Similarly distant structural matches between CofE-AF and $F_{420}H_2$:NADP+ oxidoreductase (PDB code 1JAX, RMSD 3.46 Å, Z-score 3.0)[26] and delta(1)-pyrroline-5-carboxylate reductase (PDB code 2AMF, RMSD 3.15 Å, Z-score 2.6)[43] were identified using protein secondary structure-matching algorithm as implemented in the SSM program of the ProFunc programs suite.[44] These results indicate strongly that the structure of CofE-AF lacks any structural homologues in the current PDB and therefore represents a novel protein fold. Thus, this enzyme represents the first member of a new structural family of non-ribosomal peptide synthases.

Catalytic site of CofE

The structure of the apo-CofE-AF dimer unveiled the presence of large symmetry-related deep grooves at the dimer interface extending from the N-terminal domain toward the C-terminal domain. Many highly conserved and potentially catalytic residues in CofE enzymes line up along the molecular surface of these grooves (T41, D109, S111, N112, D150, and T151). Therefore, these grooves are proposed to harbor the active site (Figure 5(a)). The shape of the grooves at each subunit resemble an inverted Y-connector (Figure 5(a)) and could be divided into three sub-cavities based on their position, depth and electrostatic potential. The top cavity T (12 Å × 10 Å × 8 Å) is located at the middle of the N-terminal domain and is contoured by helix 2, a loop positioned between strand 1 and helix 1, and a long loop placed between strand 12 and helix 6 (Figure 3(b)). This cavity has blended electrostatic charge with the top and bottom part being positively and negatively charged, respectively. All residues of this cavity are provided by a single chain. The bioinformatic comparative surface pattern analysis using the GPSS[45] server indicated that this part of the groove shows a high level of similarity to the GTP-binding cavity of GTP cyclohydrolase I,[46] and that it might bind GTP (further confirmed by the co-crystallization with the GDP co-product, see below).

The other two cavities are located at the interface of the N and C-terminal domains with one facing the dimer interface (an interface sub-cavity I) and the second extending symmetrically along the monomer surface (convex sub-cavity C). The interface cavity I with the dimensions of 16 Å × 8 Å × 9 Å is outlined by strands 5 and 6 of the C-terminal domain and β-hairpin region formed by strands 8 and 9 of the N-terminal domain. The electrostatic analysis of this cavity shows that this cavity is positively charged (Figure 5(a)), with the charges contributed mainly by conserved residues R158 and R234. In turn, the convex cavity C is formed by helix 2, helix 4, strand 7, and part of the long loop connecting strand 6 and strand 7, and the loops connecting the N and C-terminal domains. This cavity is mainly hydrophobic, with partial negative charge, and is considerably shallower in comparison to the other two with dimension of 8 Å × 7 Å × 4 Å deep. Three conserved residues, D109, S111, and T151, line at the intersection of these three sub-cavities and represent potential catalytic residues.

Nucleotide and metal ion-binding sites

It has been shown that CofE-MJ requires GTP and divalent metal cations such as Mn^{2+} or Mg^{2+} for enzymatic activity.[30] In order to better define the active site cavities involved in substrates, cofactor and metal binding, the CofE-AF was co-crystallized with GDP and Mn^{2+} and the near-atomic resolution structure of the complex was determined. A close inspection of the electron density maps revealed excellent electron density for GDP and Mn^{2+}. There are two Mn^{2+}-binding sites per monomer within close proximity of the GDP α and β-phosphate groups (Figures 3(a), 5(b) and (c)). This structure probably represents a state of the reaction after GTP hydrolysis, L-glutamate transfer and dissociation of F_{420} product.

The GDP molecules and Mn^{2+} cations are bound to the cavity T of the groove within the N-terminal domain. The Fourier difference density maps (2*Fo*–*Fc*, *Fo*–*Fc*) of the nucleotide and neighboring Mn^{2+} cations were of excellent quality, allowing highly accurate refinement of the ligand positions (Figure 5(b)). The GDP molecule interacted extensively with protein residues from the loop located between β1 and α1, $\alpha 2,3_{10}$ helix, η2, the loop located between β6 and β7 and the loop placed between α6 and β10. The β phosphate group extends to the region where all three sub-cavities join. The interaction of the GDP molecule with protein residues is summarized in Figure 5(c). The guanine ring of GDP binds to a highly hydrophobic cavity composed of residues L11, P12, L13, I14, V42, I213, L182 (the L182 is contributed by the loop of subunit B), and K45 (the only charged residue in the guanine-binding site). While the majority of the hydrophobic residue side-chains point toward the bottom and the roof of the cavity, the middle of the cavity is surrounded by the backbone carbonyl oxygen atoms,

peptide bond amide groups of these residues, and the amino group (N^{ε}) of the conserved K45. These backbone residues interact with the guanine ring of the nucleotide by the extensive network of hydrogen bonds illustrated by Figure 5(c). Interestingly, this main chain hydrogen bond network seems to be able to discriminate guanine from adenine and other bases. The ribose moiety of GDP, which displays a C2′-*endo* pucker, interacts with protein through the hydroxyl groups of the ribose. The 2′ hydroxyl group forms hydrogen bonds with the peptide bond amide hydrogen of G212 and the peptide bond carbonyl group of G209. The 3′ hydroxyl group of the ribose moiety of GDP interacts also with the peptide bond carbonyl group of M206 and the backbone amide group of I213. Once more the main chain hydrogen bond network seems to be able to discriminate ribose from 2-deoxyribose. Therefore, GTP is the preferred cofactor based on the interaction with main chain atoms. This is in agreement with data reported in the case of CofE-MJ.[30] The diphosphate moiety of GDP forms hydrogen bonds with highly conserved residues S40, T41, and N112 (Figure 5(c)) and additional interactions with protein are mediated *via* metal ions and water molecules.

Two manganese ions bind to the GDP α and β-phosphate groups. Mn1 has distorted octahedral coordination geometry with six oxygen atoms: $O1^{\alpha}$ and $O2^{\beta}$ of the GDP diphosphate moiety, OE1 of E208, OG1 of T151, and two water molecules. Mn2 is located near the β-phosphate group of GDP in the region where all three sub-cavities merge and therefore its position appears to be of critical importance to catalysis. Mn2 is coordinated by $O3^{\beta}$ of the GDP and two OD1 atoms of the aspartate residues D109 and D150. The partly distorted octahedral coordination environment of Mn2 is completed by three water molecules (Figure 5(d)).

Structural implication for catalysis

Studies of nucleotide-dependent ligases such as tetrahydrofolate: L-glutamate γ-ligase (FPGS; EC 6.3.2.17),[47] and UDP-N-acetylmuramoyl-L-alanine: glutamate ligase (MurD; EC 6.3.2.9) [48] have shown that the amide bond-forming enzymes follow a common mechanism. In the initial step, the acid carboxylate group of the substrate is activated by phosphorylation with subsequent hydrolysis of the nucleotide triphosphate. Consequently, the carbonyl carbon of the resulting acyl phosphate undergoes nucleophilic attack by the amine, leading to a tetrahedral intermediate, which breaks down to the final product and inorganic phosphate.[48,49] It is proposed that CofE uses the same strategy.[30] The formation of F_{420}-1 or F_{420}-2 by CofE-MJ requires the activation of the free carboxylate group of either F_{420}-0 or F_{420}-1 by GTP and creation of respective acyl-phosphate intermediates. This intermediate undergoes nucleophilic attack by the amino group of the incoming glutamate residue, leading to formation of the respective monoglutamate or diglutamate form of F_{420}-0. Given that the CofE-AF catalyzes the same reaction and shares a high level of sequence similarity with CofE-MJ, it is expected that CofE-AF follows the same mechanism.

The crystal structures of apo- and GDP-bound forms of CofE-AF provide some structural insights into the catalytic mechanism. Structural studies of FPGS and MurD showed two different structural designs of nucleotide-dependent ligases with FPGS having two domains, and MurD having three domain organizations, respectively. In both structures, the common feature is the presence of a deep cleft formed between the domains. On a broad level, the structure of the apo form of the enzyme reveals similar organization with the presence of the deep grooves in each monomer. However, in contrast to the enzymes listed above, CofE-AF shows extensive grooves with three sub-cavities (the top cavity T, the convex cavity C and the interface cavity I) organized in a Y-connector shape (Figure 5(a)).

In order for ligation to occur, GTP and F_{420}-0 as well as L-glutamate should be brought close together and should be oriented properly. It is reasonable to assume that the Y-connector design would allow for the binding of each of the three ligands into the specific cavities in orientation consistent with the mechanism outlined above. Along these lines, the structure CofE-AF

complexed with GDP revealed that cavity T binds GDP and two manganese ions. At this point, it is unclear in which sub-cavity F_{420}-0 and L-glutamate binds. Structural analysis of the dimer interface and the convex pockets in the apo and GDP-bound CofE-AF suggests that F_{420}-0 could bind in the interface pocket rather than in the convex pocket. This observation relies solely on features such as larger size of the interface cavity in comparison with the much smaller convex pocket, presence of numerous conserved residues and the positive electrostatic potential of this pocket. These features make the interface pocket better suited to bind a large substrate, such as F_{420}-0, in comparison with the convex pocket. The dimension and elongated shape of this cavity would allow the binding of this substrate in extended conformation such that the ring moiety could be placed at the hydrophobic region (F92, L94, V104, F156), whereas the negatively charge 2-phospho-L-lactate moiety of F_{420}-0 could be stabilized by the positively charged patch in the center of the groove.

Given the Y-shaped architecture of the active center of the enzyme, the geometric constraints that the proposed mechanism imposes on all three substrates, it is reasonable to hypothesize that the L-glutamate has to be positioned on the side of the convex cavity. While we await structural results to confirm this hypothesis, some conclusions can be drawn from the analysis of the convex pocket environment in both the apo and complexed structures of CofE-AF. In both structures, the center of the convex pocket is mainly hydrophobic and only potential catalytic residues that could interact with charged L-glutamate are positioned at the edge of the convex cavity. Conserved K71 and N105 residues are about 5 Å apart and could provide a favorable environment for binding L-glutamate and enforce its proper orientation for attack on the acyl phosphate. Supportive evidence for this hypothesis comes from the structural studies of other members of the amide bond ligases superfamily (MurD, FPGS). Comparable architecture of L-glutamate-binding sites was reported previously for FPGS and MurD. In MurD K348, T321 and S415 were implicated in glutamate-binding, whereas in FPGS, residues K346, H316, and S412 were proposed to have a role in the binding of a glutamate substrate.[47,48]

The comparison of the active site structures of CofE-GDP-Mn with MurD-ADP-Mg and FPGS-AMPPCP-Mg complexes shows different arrangements of the metal-binding pockets. While in CofE-AF, the Mn1 is bound in the pocket showing a similar metal-binding environment (α and β-phosphoryl oxygen atoms of the GDP, OE1 of E208 and OG1 of T151, and two ordered water molecules) to MurD (side-chain oxygen atom of S116 and E157, β-phosphoryl oxygen atom of the ADP, and water molecules) and FPGS (the backbone oxygen atom of S73 and side-chain oxygen atom of E157, β-phosphoryl oxygen atom of the ADP, and water molecules), the positioning of the pockets appears to be distinctly different. In MurD and FPGS the divalent metal ion is positioned between the β and γ-phosphate groups of the nucleotide (so called classical magnesium site),[48,49] while in CofE-AF the α and β-phosphate oxygen atoms coordinate the metal ion. With respect to the position and the coordination environment of Mn2 in the structure of CofE-GDP-Mn, no similarity could be found in the FPGS and MurD enzymes. In CofE-AF the Mn2 is strongly chelated by the carboxyl groups of D109 and D150 and O3 β-phosphoryl oxygen atom of the GDP. In FPGS, the second magnesium ion is bound by γ and δ-phosphoryl oxygen atoms of the observed adenosine tetraphosphate.[49] In the MurD structure complex, the second metal ion interacts with the carboxyl oxygen atom of the L-alanine moiety of the UMA substrate (UDP-MurNAc-L-Ala) and nitrogen atom of H183, and water molecules.[47]

In conclusion, CofE-AF represents the first member of a new structural family of non-ribosomal peptide synthases. The apo and complex structures reveal, for the first time, the highly detailed snapshot of the active site architecture of this enzyme and the location of the GDP cofactor and manganese ions. Although the CofE structure is unique and represents a new protein fold, the enzyme shows some similarity in the overall organization of the active site with other nucleotide-dependent ligases. It appears that these enzymes utilize similar

design, where the active sites are positioned at the domain interfaces, with specialized sub-regions involved in binding of substrates and cofactors. In this regard, CofE-AF shows a conceptually similar organization that utilizes a Y-shaped design with three sub-cavities. Enzymatic studies revealed that CofE-AF synthesizes F_{420}-2 from F_{420}-0 in two independent binding events. It is puzzling how the enzymes can accommodate both substrates F_{420}-0 and F_{420}-1 in the same active site to perform the addition of glutamate. Crystallographic and biochemical studies are in progress to understand this interesting feature of the enzyme and the details of its catalysis.

Materials and Methods

Materials

All reagents (except for F_{420}-0) were purchased from Sigma.

Gene cloning, protein expression and purification

The open reading frame of CofE-AF (gi | 2648269) was amplified by PCR from *A. fulgidus* strain DSM 4304 genomic DNA (ATCC). The gene was cloned into the NdeI and BamHI sites of a modified pET15b cloning vector (Novagen) in which the TEV protease cleavage site replaced the thrombin cleavage site, and a double stop codon was introduced downstream from the BamHI site. This construct provides an N-terminal His_6 tag separated from the gene by a TEV protease recognition site (ENLYFQ↓G). The fusion protein was overexpressed in *E. coli* BL21-Gold (DE3 (Stratagene) harboring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile). The cells were grown in LB at 37 °C to an A_{600} of ~0.6, and protein expression was induced with 0.4 mM IPTG. After induction, the cells were incubated with shaking overnight at 15°C. The harvested cells were resuspended in binding buffer (50 mM Hepes (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol, 5 mM imidazole), flash-frozen in liquid N_2 and stored at −70°C. The thawed cells were lysed by sonication after the addition of 0.5% (v/v) NP-40, 1 mM PMSF and 1 mM benzamidine. The lysate was clarified by centrifugation (30 min at 27,000*g*) and passed through a DE52 column pre-equilibrated in binding buffer. The flow-through fraction was applied to a metal chelate affinity column charged with Ni^{2+}. The His_6-tagged protein was eluted from the column in elution buffer (50 mM Hepes (pH 7.5), 500 mM NaCl, 5% glycerol, 500 mM imidazole) and the tag then cleaved from the protein by treatment with recombinant His_6-tagged TEV protease. The cleaved protein was resolved from the cleaved His_6 tag and the His_6-tagged protease by passage of the mixture through a second Ni^{2+}-column.

The CofE-AF protein was dialyzed against 10 mM Hepes (pH 7.5), 500 mM NaCl, and concentrated using a BioMax concentrator (Millipore). Before crystallization, any particulate matter was removed from the sample by passage through a 0.2 mm pore size Ultrafree-MC centrifugal filter (Millipore). For the preparation of SeMet-labeled protein, CofE-AF was expressed in the *E. coli* methionine auxotroph strain B834(DE3) (Novagen) and purified under the same conditions as the native protein in supplemented M9 medium. The reducing reagent β-mercaptoethanol was added (5 mM final concentration) to all purification buffers.

Synthesis of F_{420}-0

F_{420}-0 was synthesized enzymatically using recombinantly produced MJ1117 gene product and CofD (MJ1256)[27] from *M. jannaschii*. Each 1 ml reaction mixture contained 50 mM Tris–HCl (pH 7.5), 0.1 mM GTP, 11 mM lactyl phosphate, 0.08 mM Fo, 1 mM $MnCl_2$, 16 μg of MJ1117 gene product and 40 μg of CofD. Since this enzyme system is inhibited by high concentrations of GTP, 5 mM phosphoenol pyruvate and 17 units of pyruvate kinase (Sigma P-1381) were included in order to regenerate GTP from the GMP formed in the reaction. The reactions were incubated at 37°C for 12 h, after which approximately 80% of the Fo had been

converted to F_{420}-0. Proteins were precipitated by the addition of 1.5 volumes of methanol and removed by centrifugation (14,000*g*, 10 min). The reaction mixture was concentrated by evaporation with a stream of nitrogen gas and loaded onto a C_{18} column (1 cm × 25 cm) equilibrated with water. Under these chromatographic conditions, Fo binds to the column and F_{420}-0 elutes. The fractions containing the purified F_{420}-0 were combined and concentrated by evaporation with a stream of nitrogen gas at 70°C. The concentration of the sample was determined from the absorbance at 421 nm. Purity of the sample was determined by HPLC on a C_{18} column (Varian Pursuit XRs-C18; 4.6 mm × 250 mm, 5 μm film thickness) with a linear gradient from 5% (v/v) to 80% methanol in 95% (v/v) 25 mM sodium acetate (pH 6.0) at 0.5 ml/min over 40 min (start after 5 min). Fo and F_{420}-0 were detected with excitation at 420 nm and emission at 480 nm.

Enzymatic activity of CofE-AF

F_{420}-0:γ-glutamyl ligase enzymatic activity was measured using HPLC as described.[30] The reaction mixture (50 μl final volume) contained 50 mM Ches-Na buffer (pH 8.5), 0.2 M NaCl, 5 mM $MgCl_2$, 5 mM $MnCl_2$, 5 mM GTP, 10 mM L-glutamate, 0.5 μM F_{420}-0, 0.1 μg of CofE-AF. The reaction was incubated at 50 °C for the indicated time and terminated by the addition of EDTA (10 mM final concentration). Quantitative analysis of the reaction products was performed on a Varian Prostar HPLC using a C_{18} reversed-phase chromatography column (AXXI-CHROM ODS; 4.6 mm × 250 mm, 5 μm film thickness) as described. The elution was monitored by the deazariboflavin fluorescence (excitation wavelength 420 nm, emission wavelength 480 nm). Under our chromatographic conditions, the F_{420}-0 derivatives showed the following retention times: F_{420}-0, 23.8 min; F_{420}-1, 19.6 min; and F_{420}-2, 11.7 min.

Crystallization

CofE-AF crystallization was performed by the sitting-drop and hanging-drop methods using vapor diffusion at 18 °C. Diffraction-quality crystals were obtained by mixing equal (1 μl) volumes of the protein and a solution containing a mixture of the reservoir solution (0.2 M ammonium acetate, 0.1 M sodium citrate (pH 5.6), 25% (w/v) PEG 5K MME). The colorless crystals appeared typically within five to ten days, and reached a typical size of 0.3 mm × 0.3 mm × 0.3 mm. CofE-AF crystallizes in the primitive tetragonal space group $P4_12_12$, with cell dimensions of a = 100.39 Å, b = 100.39 Å, c = 92.96 Å, α = β = γ = 90°. The crystal asymmetric unit contains two molecules, with the V_M value of 2.32 $Å^3$/Da, which corresponds to a solvent content of 46.6% (v/v). Before data collection, crystals were soaked for few seconds in a cryoprotectant consisting of 20% (v/v) ethylene glycol in the crystal mother liquor and then flash-frozen in liquid nitrogen.

In order to determine the structure with GDP and manganese ions, the protein was incubated with 10 mM GDP and 10 mM $MnCl_2$ for 20 min before the crystal setup. The conditions described above were used to grow the crystals. The GDP, $MnCl_2$-treated protein crystals resembled the previously obtained crystals showing the same morphology and size, as well as space group. The only difference observed was the unit cell dimensions, which changed by about 2 Å in each dimension to a = 98.37 Å, b = 98.37 Å, c = 94.35 Å, α = β = γ = 90°.

Data collection

A SAD experiment was carried out on the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source.[50] The experimental set-up involved taking two test diffraction images 90° apart, from which we evaluated the optimal exposure time, oscillation range, orientation of the crystal and resolution of the data to be collected. Diffraction data were collected from a SeMet-substituted crystal of CofE-AF protein and native protein (no SeMet) co-crystallized with GDP and $MnCl_2$ using inverse-beam geometry at the selenium absorption peak ((12.6603 keV (0.9794 Å)) to 2.5 Å and 1.35 Å, respectively. In both cases, before data

collection, X-ray florescence spectrums were recorded from the samples, which identified the presence of Se and Mn in the proteins, respectively. The SBCCollect program was used for data collection and visualization.[50] All data were integrated and scaled with the software package HKL3000.[51] Data collection and processing statistics are shown in Table 1.

Structure determination

The structure of the CofE-AF apo-enzyme has been determined by SAD phasing, density modification, and initial protein model building as implemented in the HKL3000 software package.[51–57] The initial model (~40% of the C^{α} trace built) was further extended manually by using the program COOT.[57] Several cycles of manual model rebuilding using COOT and refinement with REFMAC 5.2 were carried out.[52] The final model was refined against all reflections in the resolution ranges of 40.0−2.5 Å, except for the 5% randomly selected reflections that were used for monitoring R_{free}. The final round of refinement was carried out by using translation/libration/screw (TLS) refinement with 20 TLS groups per molecule. The TLS operators were calculated using the TLSMD server,[58] which indicated that partitioning of chains A and B into 20 flexible pieces may be optimal and should lead to maximal reduction in R/R_{free}.

The structure of CofE-AF-GDP-Mn complex has been phased by the molecular replacement method using the previously determined structure of apo-CofE-AF dimer as a search model, since the substrate-soaked crystals were non-isomorphous with the apo-enzyme crystal. Molecular replacement searches were completed with MOLREP (*R*-factor 0.44, correlation coefficient 0.48)[59] of the CCP4 suite. Only a few rounds of rebuilding and readjusting in the program COOT were required to obtain the final model. The final model was refined against all reflections in the resolution ranges of 40.0−1.35 Å, except for the 5% randomly selected reflections that were used for monitoring Rfree. The quality of the structure models was checked with PROCHECK.[60] The final refinement statistics are presented in Table 1.

Size-exclusion chromatography

The molecular mass of CofE-AF apo-protein in solution was evaluated by size-exclusion chromatography on a Superdex-200 10/30 column (Pharmacia) by using crystallization buffer and calibrated with ribonuclease A (13.7 kDa), ovalbumin (43 kDa), albumin (67 kDa), aldolase (158 kDa), catalase (232 kDa), and thyroglobulin (669 kDa) as standards. The calibration curve of K_{av} *versus* log molecular mass was prepared by using the equation:

$$K_{av}=(V_e - V_0)/(V_t - V_0)$$

where V_e is the elution volume for the protein, V_0 is the column void volume, and V_t is the total bed volume.

Protein Data Bank accession codes

The atomic coordinates and structure factors for apo-CofE-AF and for CofE-AF complexed with GDP and Mn^{2+} have been deposited in the RCSB Protein Data Bank with accession codes 2G9I and 2PHN, respectively.

Abbreviations used

Fo, 7,8-didemethyl-8-hydroxy-5-deazariboflavin
F_{420}, the 7,8-didemethyl-8-hydroxy-5-deazariboflavin
F_{420}, *N*-(*N*-L-lactyl-γ-L-glutamyl)-L-glutamic acid phosphodiester of 7,8-didemethyl-8-hydroxy-5-deazariboflavin

F_{420}-0, F_{420} with no glutamic acid
F_{420}-1, F_{420} with one glutamic acid
F_{420}-2, F_{420} containing two glutamate residues
SAD, single anomalous diffraction
TLS, translation/libration/screw
ASA, accessible surface area
RMSD, root-mean-square deviation
MR, molecular replacement

Acknowledgements

We are grateful to Dr Z. Dauter, Dr S. Rajan, and Dr J. Osipiuk for helpful discussions. We thank Dr T. Binkowski for the bioinformatic studies, M. Chodkiewicz-Nocek for assistance with the Figures and A. Cipriani for help in preparing the manuscript. We thank all members of the Structural Biology Center at Argonne National Laboratory and at the Ontario Centre for Structural Proteomics for their help in conducting these experiments. This work was supported by National Institutes of Health grant GM074942 and by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357 and by the U. S. National Science Foundation grant MCB 0231319 (to R.H.W.), and by Genome Canada (through the Ontario Genomics Institute).

References

1. Begley TP. Cofactor biosynthesis: an organic chemist's treasure trove. Nature Prod. Rep 2006;23:15–25. [PubMed: 16453030]
2. Peters JW, Fisher K, Newton WE, Dean DR. Involvement of the P-Cluster in intramolecular electron-transfer within the nitrogenase MoFe protein. J. Biol. Chem 1995;270:27007–27013. [PubMed: 7592949]
3. Chan MK, Kim JS, Rees DC. The nitrogenase FeMo-cofactor and P-cluster pair- 2.2 angstrom resolution structures. Science 1993;260:792–794. [PubMed: 8484118]
4. Peters JW, Lanzilotta WN, Lemon BJ, Seefeldt LC. X-ray crystal structure of the Fe-only hydrogenase (Cpl) from *Clostridium pasteurianum* to 1.8 angstrom resolution. Science 1998;282:1853–1858. [PubMed: 9836629]
5. Gorren ACF, Mayer B. Tetrahydrobiopterin in nitric oxide synthesis: a novel biological role for pteridines. Curr. Drug Metab 2002;3:133–157. [PubMed: 12003347]
6. Walsh CT, Garneau-Tsodikova S, Gatto GJ. Protein posttranslational modifications: the chemistry of proteome diversifications. Angew. Chem. Int. Ed 2005;44:7342–7372.
7. White RH. Biosynthesis of the methanogenic cofactors. Vitam. Horm 2001;61:299–337. [PubMed: 11153270]
8. Ensign SA, Allen JR. Aliphatic epoxide carboxylation. Annu. Rev. Biochem 2003;72:55–76. [PubMed: 12524213]
9. Weiss DS, Thauer RK. Methanogenesis and the unity of biochemistry. Cell 1993;72:819–822. [PubMed: 8458077]
10. Nocek B, Jang SB, Jeong MS, Clark DD, Ensign SA, Peters JW. Structural basis for CO_2 fixation by a novel member of the disulfide oxidoreductase family of enzymes, 2-ketopropyl-coenzyme M oxidoreductase/carboxylase. Biochemistry 2002;41:12907–12913. [PubMed: 12390015]
11. Dimarco AA, Bobik TA, Wolfe RS. Unusual coenzymes of methanogenesis. Annu. Rev. Biochem 1990;59:355–394. [PubMed: 2115763]
12. Shima S, Warkentin E, Thauer RK, Ermler U. Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen. J. Biosci. Bioeng 2002;93:519–530. [PubMed: 16233244]
13. Graham DE, White RH. Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. Nature Prod. Rep 2002;19:133–147. [PubMed: 12013276]
14. Cousins FB. The prosthetic group of a chromoprotin from mycobacteria. Biochim. Biophys. Acta 1960;40:532–534. [PubMed: 13812439]
15. Sutton WB. Properties of a new TPN-like electron transport component from *Mycobacterium phlei*. Biochem. Biophys. Res. Co 1964;15:414–419.

16. Cheeseman P, Toms-Wood A, Wolfe RS. Isolation and properties of a fluorescent compound factor 420, from *Methanobacterium* strain M.o.H. J. Bacteriol 1972;112:527–531. [PubMed: 5079072]
17. Eirich LD, Vogels GD, Wolfe RS. Proposed structure for coenzyme-F_{420} from *Methanobacterium*. Biochemistry 1978;17:4583–4593. [PubMed: 728375]
18. Daniels L, Bakhiet N, Harmon K. Widespread distribution of a 5-deazaflavin cofactor in actinomyces and related bacteria. Syst. Appl. Microbiol 1985;6:12–17.
19. Ebert S, Rieger PG, Knackmuss HJ. Function of coenzyme F420 in aerobic catabolism of 2,4,6-trinitrophenol and 2,4-dinitrophenol by *Nocardioides simplex* FJ2-1A. J. Bacteriol 1999;181:2669–2674. [PubMed: 10217752]
20. Stover CK, Warrener P, VanDevanter DR, Sherman DR, Arain TM, Langhorne MH, et al. A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature 2000;405:962–966. [PubMed: 10879539]
21. Peschke U, Schmidt H, Zhang HZ, Piepersberg W. Molecular characterization of the lincomycin-production gene-cluster of *Streptomyces lincolnensis* -78-11. Mol. Microbiol 1995;16:1137–1156. [PubMed: 8577249]
22. Graupner M, Xu HM, White RH. Characterization of the 2-phospho-L-lactate transferase enzyme involved in coenzyme F_{420} biosynthesis in *Methanococcus jannaschii*. Biochemistry 2002;41:3754–3761. [PubMed: 11888293]
23. Hausinger RP, Ormejohnson WH, Walsh C. Factor-390 chromophores- phosphodiester between Amp or Gmp and methanogen factor-420. Biochemistry 1985;24:1629–1633. [PubMed: 2988607]
24. Kunow J, Schworer B, Setzke E, Thauer RK. Si-face stereospecificity at C5 of coenzyme F_{420} for F_{420}-dependent N5,N10-methylenetetrahydro-methanopterin dehydrogenase, F_{420}-dependent N5, N10-methylenetetrahydromethanopterin reductase and F_{420} H_2: dimethylnaphthoquinone oxidoreductase. Eur. J. Biochem 1993;214:641–646. [PubMed: 8319675]
25. Seedorf H, Kahnt J, Pierik AJ, Thauer RK. Si-face stereospecificity at C5 of coenzyme F_{420} for F_{420} H_2 oxidase from methanogenic Archaea as determined by mass spectrometry. FEBS J 2005;272:5337–5342. [PubMed: 16218963]
26. Warkentin E, Mamat B, Sordel-Klippert M, Wicke M, Thauer RK, Iwata M, et al. Structures of F_{420} H_2: NADP(+) oxidoreductase with and without its substrates bound. EMBO J 2001;20:6561–6569. [PubMed: 11726492]
27. Li H, Xu HM, Graham DE, White RH. Glutathione synthetase homologs encode alpha-L-glutamate ligases for methanogenic coenzyme F_{420} and tetrahydrosarcinapterin biosyntheses. Proc. Natl Acad. Sci. USA 2003;100:9785–9790. [PubMed: 12909715]
28. Grochowski LL, Xu HM, White RH. Identification of lactaldehyde dehydrogenase in *Methanocaldococcus jannaschii* and its involvement in production of lactate for F_{420} biosynthesis. J. Bacteriol 2006;188:2836–2844. [PubMed: 16585745]
29. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res 1997;25:3389–3402. [PubMed: 9254694]
30. Li H, Graupner M, Xu HM, White RH. CofE catalyzes the addition of two Glutamates to F_{420}-0 in F_{420} coenzyme biosynthesis in *Methanococcus jannaschii*. Biochemistry 2003;42:9771–9778. [PubMed: 12911320]
31. Cafaro V, Scognamiglio R, Viggiani A, Izzo V, Passaro I, Notomista E, et al. Expression and purification of the recombinant subunits of toluene/o-xylene monooxygenase and reconstitution of the active complex. Eur. J. Biochem 2002;269:5689–5699. [PubMed: 12423369]
32. Klugkist J, Voorberg J, Haaker H, Veeger C. Characterization of 3 different flavodoxins from *Azotobacter vinelandii*. Eur. J. Biochem 1986;155:33–40. [PubMed: 3948879]
33. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallog. sect. D 1997;53:240–255.
34. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallog. sect. D 2004;60:2256–2268.
35. Krissinel, E.; Henrick, K. Detection of protein assemblies in crystals. In: Berthold, MR.; Glen, R.; Diederichs, K.; Kohlbacher, O.; Fischer, I., editors. CompLife. Vol. LNBI 3695. Berlin: Springer-Verlag; 2005. p. 163-174.

36. Leesong M, Henderson BS, Gillig JR, Schwab JM, Smith JL. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. Structure 1996;4:253–264. [PubMed: 8805534]
37. Dillon SC, Bateman A. The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. BMC Bioinformatics 2004;5(109):14. [PubMed: 15053846]
38. Vetriani C, Maeder DL, Tolliday N, Yip KSP, Stillman TJ, Britton K, et al. Protein thermostability above 100 degrees C: a key role for ionic interactions. Proc. Natl Acad. Sci. USA 1998;95:12300–12305. [PubMed: 9770481]
39. Arnott MA, Michael RA, Thompson CR, Hough DW, Danson MJ. Thermostability and thermoactivity of citrate synthases from the thermophilic and hyperthermophilic archaea, *Thermoplasma acidophilum* and *Pyrococcus furiosus*. J. Mol. Biol 2000;304:657–668. [PubMed: 11099387]
40. Karshikoff A, Ladenstein R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a ‘traffic rule’ for hot roads. Trends Biochem. Sci 2001;26:550–556. [PubMed: 11551792]
41. Holm L, Sander C. Protein-structure comparison by alignment of distance matrices. J. Mol. Biol 1993;233:123–138. [PubMed: 8377180]
42. Cirilli M, Zheng RJ, Scapin G, Blanchard JS. Structural symmetry: the three-dimensional structure of *Haemophilus influenzae* diaminopimelate epimerase. Biochemistry 1998;37:16452–16458. [PubMed: 9843410]
43. Nocek B, Chang C, Li H, Lezondra L, Holzle D, Collart F, Joachimiak A. Crystal structures of Delta (1)-pyrroline-5-carboxylate reductase from human pathogens *Neisseria meningitides* and *Streptococcus pyogenes*. J. Mol. Biol 2005;354:91–106. [PubMed: 16233902]
44. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucl. Acids Res 2005;33:89–93.
45. Binkowski TA, Joachimiak A, Liang J. Protein surface analysis for function annotation in high-throughput structural genomics pipeline. Protein Sci 2005;14:2972–2981. [PubMed: 16322579]
46. Rebelo J, Auerbach G, Bader G, Bracher A, Nar H, Hosl C, et al. Biosynthesis of pteridines. Reaction mechanism of GTP cyclohydrolase I. J. Mol. Biol 2003;326:503–516. [PubMed: 12559918]
47. Sheng Y, Sun XL, Shen Y, Bognar AL, Baker EN, Smith CA. Structural and functional similarities in the ADP-forming amide bond ligase superfamily: implications for a substrate-induced conformational change in folylpolyglutamate synthetase. J. Mol. Biol 2000;302:427–440. [PubMed: 10970743]
48. Bertrand JA, Auger G, Fanchon E, Martin L, Blanot D, van Heijenoort J, Dideberg O. Crystal structure of UDP-N-acetylmuramoyl-L-ala-nine: D-glutamate ligase from Escherichia coli. EMBO J 1997;16:3416–3425. [PubMed: 9218784]
49. Sun XL, Cross JA, Bognar AL, Baker EN, Smith CA. Folate-binding triggers the activation of folylpolyglutamate synthetase. J. Mol. Biol 2001;310:1067–1078. [PubMed: 11501996]
50. Rosenbaum G, Alkire RW, Evans G, Rotella FJ, Lazarski K, Zhang R, et al. The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. J. Synchrotron Radiat 2006;13:30–45. [PubMed: 16371706]
51. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution- from diffraction images to an initial model in minutes. Acta Crystallog. sect. D 2006;62:859–866.
52. Collaborative Computing Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallog. sect. D 1994;50:760–763.
53. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallog. sect. A 1991;47:110–119.
54. Cowtan KD, Main P. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. Acta Crystallog. sect. D 1993;49:148–157.
55. Terwilliger TC. Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. Acta Crystallog. sect. D 2003;59:1174–1182.

56. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallog. sect. D 2002;58:1772–1779.
57. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallog. sect. D 2004;60:2126–2132.
58. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallog. sect. D 2006;62:439–450.
59. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallog. sect. D 2000;56:1622–1624.
60. Laskowski RA. PDBsum: summaries and analyses of PDB structures. Nucl. Acids Res 2001;29:221–222. [PubMed: 11125097]
61. Graupner M, Xu HM, White RH. The pyrimidine nucleotide reductase step in riboflavin and F_{420} biosynthesis in Archaea proceeds by the eukaryotic route to riboflavin. J. Bacteriol 2002;184:1952–1957. [PubMed: 11889103]
62. Corpet F. Multiple sequence alignment with hierarchical-clustering. Nucl. Acids Res 1988;16:10881–10890. [PubMed: 2849754]
63. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 1999;15:305–308. [PubMed: 10320398]

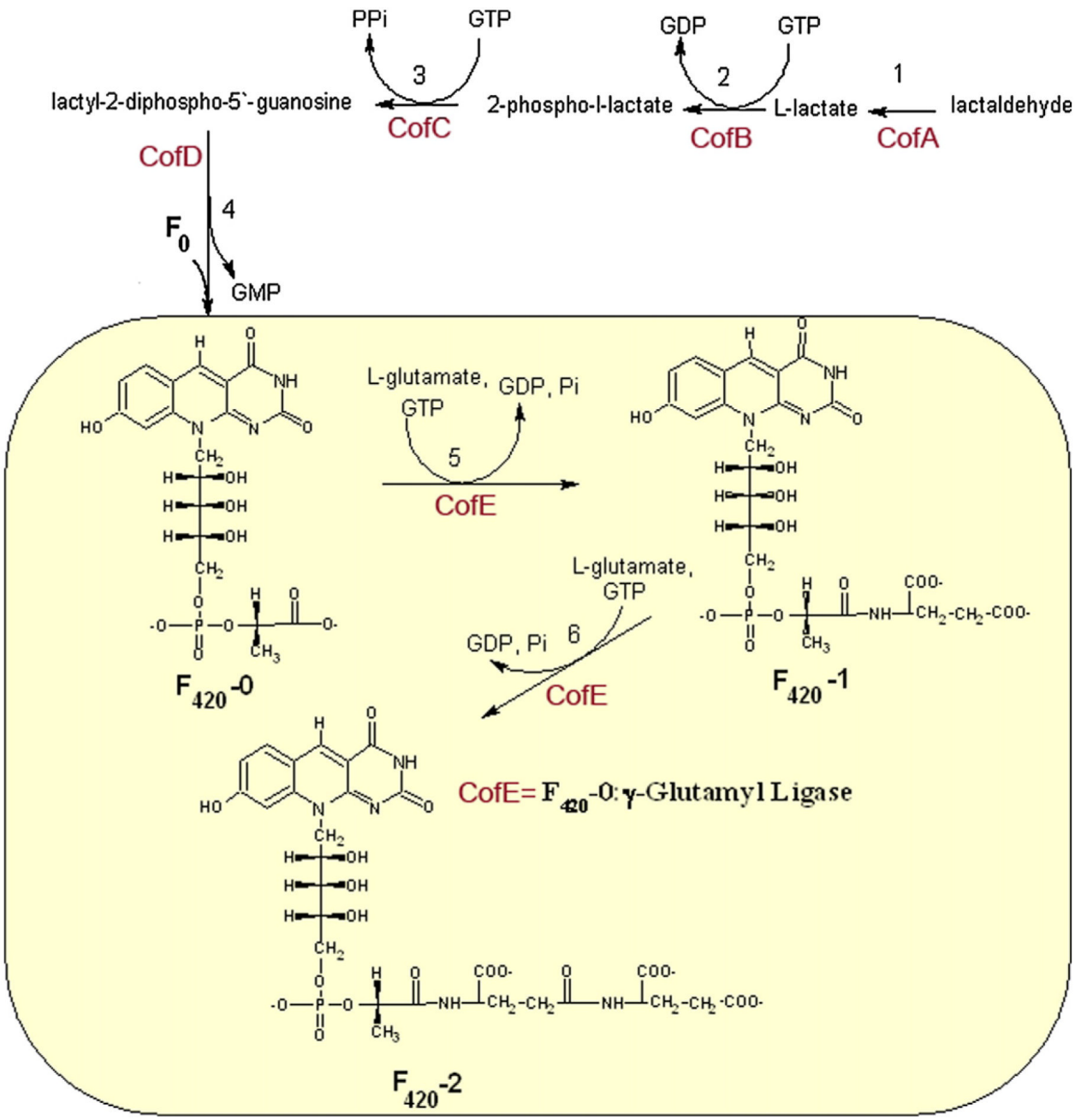

Figure 1.

Proposed pathway of the biosynthesis of coenzyme F_{420} based on the studies of a representative Archaea, *M. jannaschi*..[27,61] The reaction catalyzed by the enzyme F_{420}-0:γ-L-glutamyl ligase (CofE) is highlighted in the yellow box.

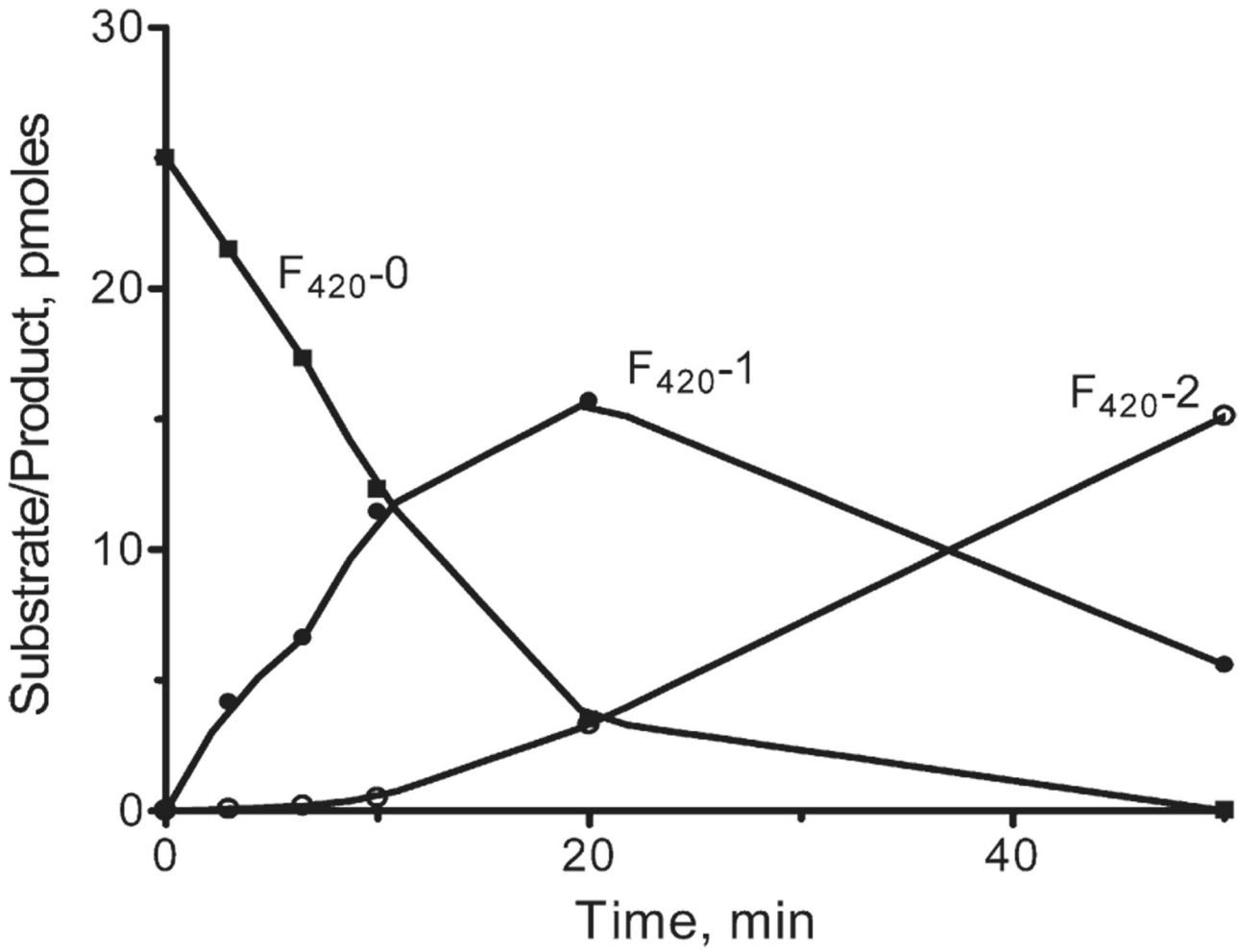

Figure 2.
CofE-AF activity assay showing the time-course of substrate F_{420}-0 depletion and the formation of products F_{420}-1 and F_{420}-2) in the presence of L-glutamate and GTP.

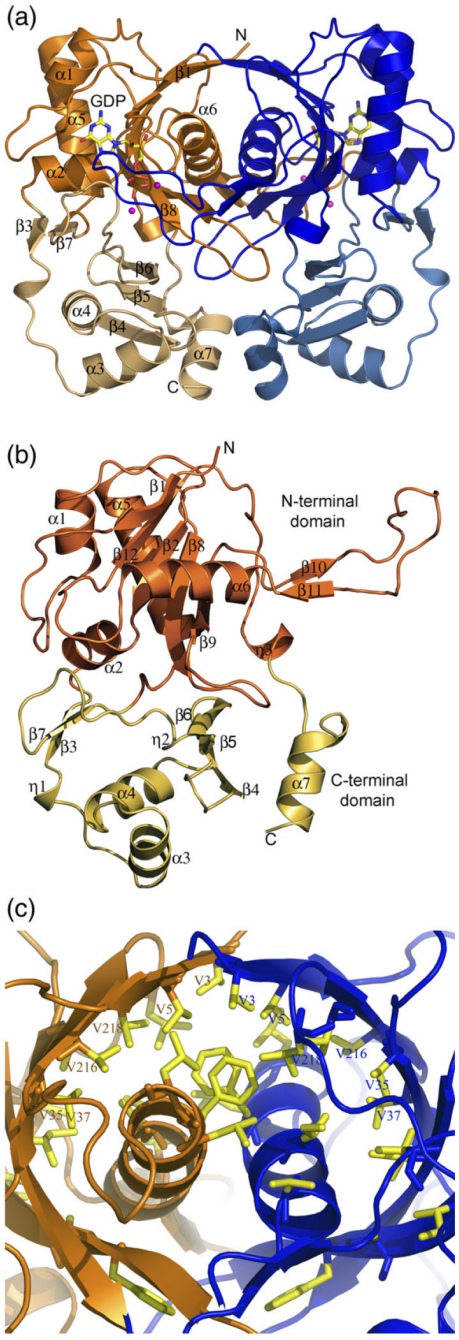

Figure 3.

Structure of the CofE-AF. (a) Ribbon diagram of the homodimer of CofE-AF. The dimer resembles a butterfly and occupies a volume of 73 × 52 × 38 $Å^3$. Chain A is colored orange and chain B is colored blue. Molecules of GDP are shown as yellow, red and blue sticks, and Mn^{+2} ions are shown as magenta spheres. Secondary structure elements, GDP site and N and C termini are labeled. (b) Overall view of the monomer of CofE-AF showing two domains: the N-terminal domain highlighted in orange and C-terminal domain highlighted in light yellow. Secondary structure elements are labeled. (c) Close-up view of the unique dimerization interactions involving formation of an antiparallel two helices bundle wrapped by an antiparallel ten-stranded β-sheet barrel. Non-polar side-chains are shown as yellow sticks

including six pairs of Val residues. These Figure and Figure 5 were prepared using PyMOL [http://www.pymol.org].

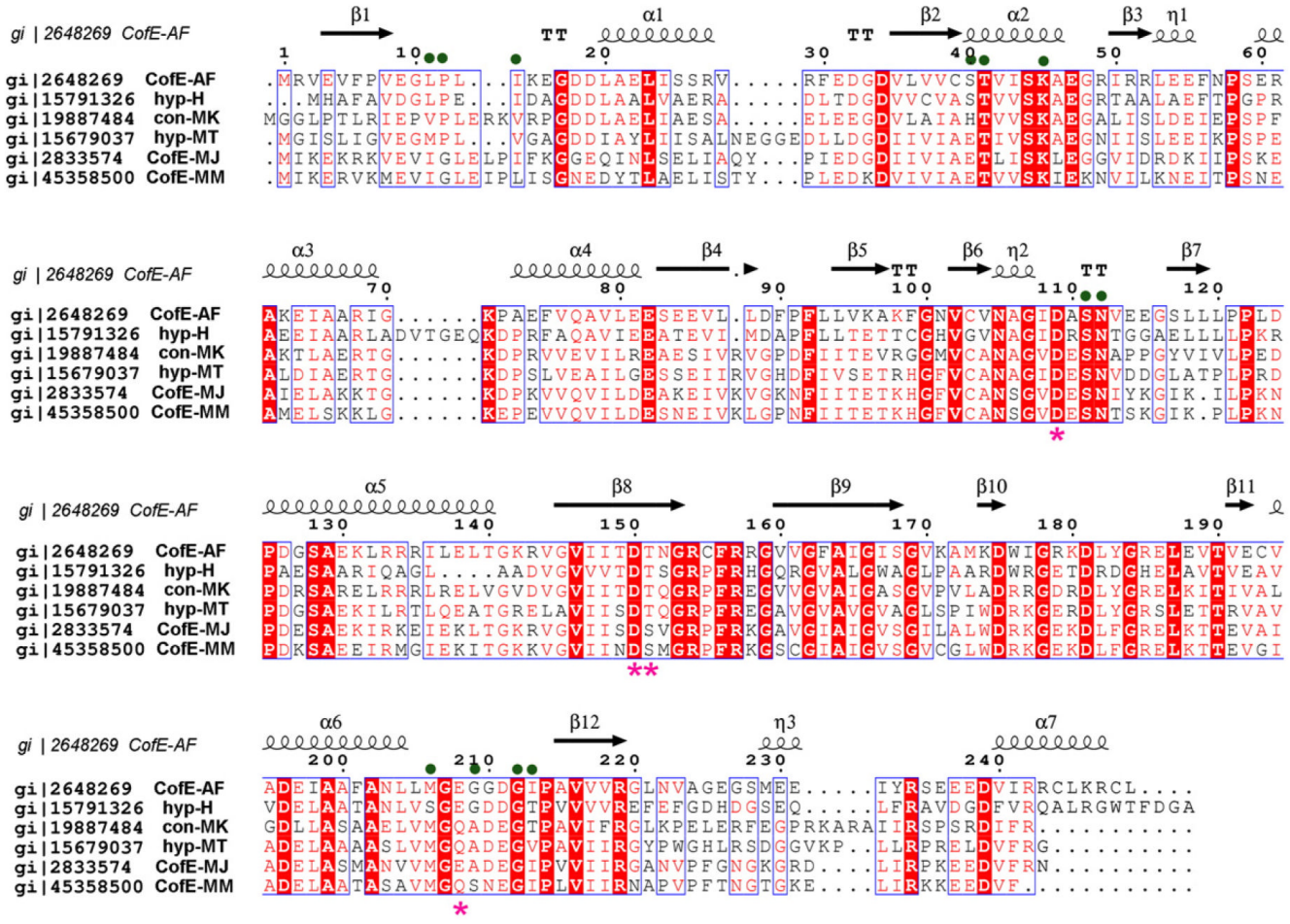

Figure 4.

Multiple sequence alignment of selected representatives of the F_{420}-0:γ-L-glutamate ligase family and its homologues from microorganisms known to produce coenzyme F_{420}. Sequence identities are highlighted in red and similarities are shown as red letters. The corresponding secondary structure elements of CofE-AF are shown on the top (black). Magenta stars indicate residues that coordinate metal ions, and green dots indicate residues interacting with a nucleotide (GDP). The Gene Bank accession numbers correspond to the following proteins: gi|2648269 CofE-AF from *A. fulgidus DSM 4304*, gi | 19887484 uncharacterized conserved protein *Methanopyrus kandleri AV19*, gi|15791326 hypothetical protein VNG2586C *Halobacterium* sp. *NRC-1*, gi|15679037 hypothetical protein MTH1019 *M. thermautotrophicus str. Delta H*, gi|2833574 F_{420}-0:γ-glutamyl ligase *M. jannaschii* MJ0768, gi|45358500 F_{420}-0:γ-L-glutamate ligase from *M. maripaludis S2*. The sequence alignment was performed using the program Multalin[62] and the Figure was generated using ESPript.[63]

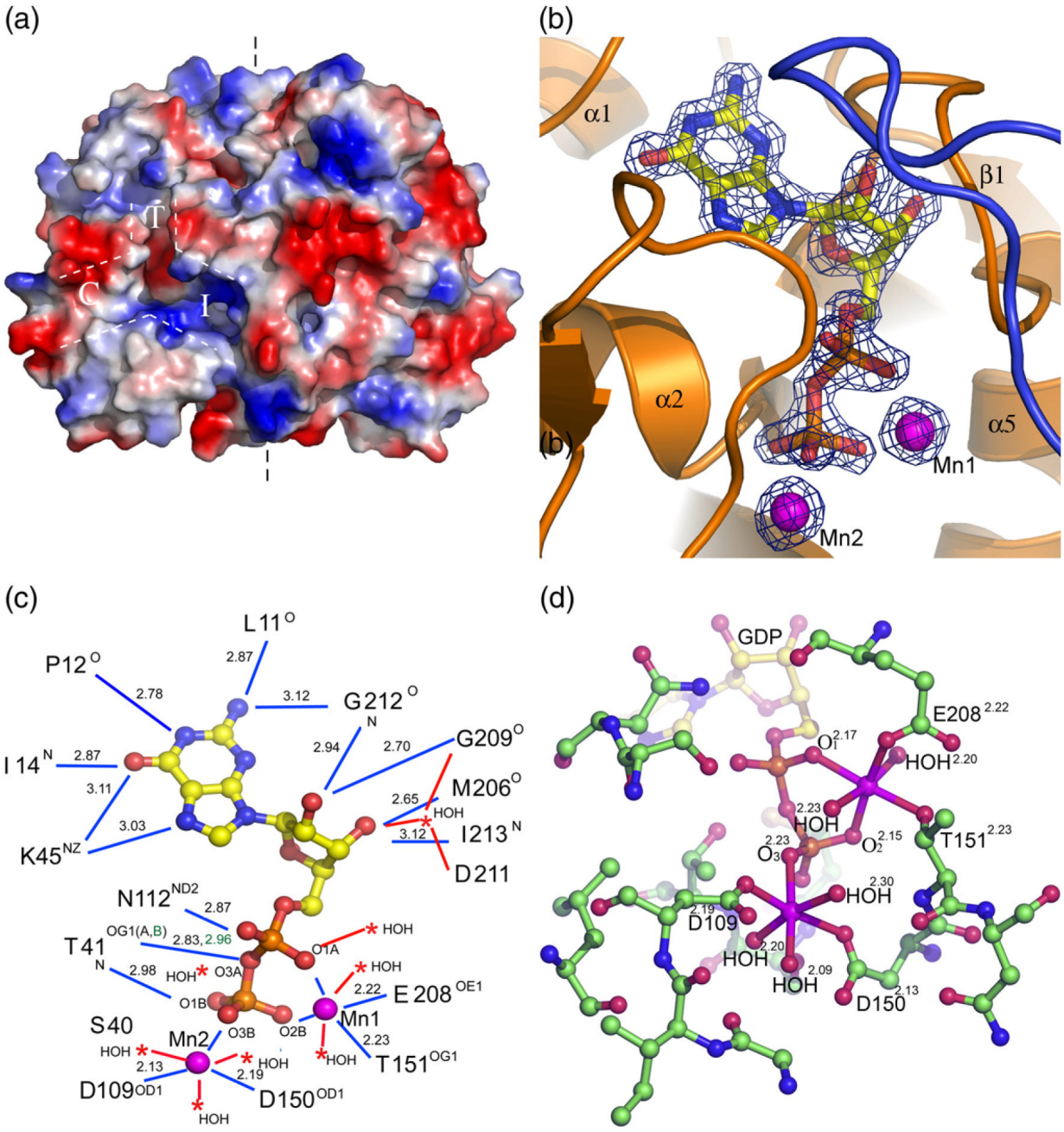

Figure 5.

Active site of the CofE-AF. (a) Surface rendering of the apo-Cof-AF structure. White broken lines highlight a region of the putative active site, resembling an inverted Y-connector. (b) $2F_o–F_c$ electron density map contoured at 3σ covering the molecule of GDP and two magnesium ions (magenta). (c) A representation of GDP interactions with the surrounding protein residues (blue lines) and solvent molecules (red lines) in molecule A of CofE-AF. Distances are shown along with bond donors and acceptors. The alternative conformation interactions are highlighted in green. (d) Close-up view of the Mn^{2+}-binding site. Manganese ions are represented by magenta balls. The coordination geometry of each Mn^{2+} is illustrated as red lines with the coordinating atoms labeled and distances (in Å) marked as superscripts.

Table 1

Data and refinement statistics

A. *Data collection statistics*	CofE	CofE+GDP+Mn
Space group	$P4_12_12$	$P4_12_12$
Unit cell dimensions		
a (Å)	100.39	98.37
b (Å)	100.39	98.37
c (Å)	92.96	94.35
Wavelength (Å)	0.9794	0.9794
Resolution (Å)	40–2.50	40–1.35
No. observed reflections	34,373	823,035
No. unique reflections	16,992	190,592
R_{merge} [a] (%)	8.9 (43.6)	5.5 (48.9)
Completeness (%)	99.9 (99.2)	98.7 (97.9)
I/Σ	22.8 (4.0)	22.3 (4.7)
B. *Phasing*		
Phasing method	SAD	MR
Resolution range (Å)	40–3.0	40–3.0
No. SeMet residues	8	0
Phasing statistics	Phasing power 1.53	*R*-factor 0.44; CC [b] 0.48
C. *Refinement statistics*		
Resolution range (Å)	40.0–2.50	40.0–1.35
R_{cryst} (%)	19.4	16.0
R_{free} (%)	25.5	19.0
No. non-hydrogen atoms	3704	4784
No. protein atoms	3603	4076
GDP ions	—	2
Mn ions	—	4
Solvent ions	98	586
RMSD from target values		
Bond lengths (Å)	0.020	0.012
Bond angles (deg)	1.83	1.56
Average *B* factors Protein main chain (Å^2)	25.2	17.6
Protein side-chain (Å^2)	28.3	20.5
Mn	—	12.3
GDP	—	11.9
H_2O	44.3	33.9

Numbers in parentheses are values for the highest-resolution bin.

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i \mid I_i - \langle I \rangle / \Sigma_{hkl}\Sigma_i \mid \langle I \rangle \mid$, where I_i is the intensity for the *i*th measurement of an equivalent reflection with indices *h*, *k*, and *l*.

[b] CC, correlation coefficient.