

1 **Title:**

2 Diversification by CofC and control by CofD govern biosynthesis and evolution of
3 coenzyme F₄₂₀ and its derivative 3PG-F₄₂₀.

4 **Authors:**

5 Mahmudul Hasan¹, Sabrina Schulze², Leona Berndt², Gottfried J. Palm², Daniel Braga¹,
6 Ingrid Richter³, Daniel Last¹, Michael Lammers², and Gerald Lackner^{1#}

7

8 ¹Junior Research Group Synthetic Microbiology, Leibniz-Institute for Natural Product
9 Research and Infection Biology, Beutenbergstr. 11a, 07745 Jena, Germany

10 ²Institute for Biochemistry, Dep. Synthetic and Structural Biochemistry, Felix-Hausdorff-
11 Str. 4, University of Greifswald, 17487 Greifswald, Germany

12 ³Dept. Biomolecular Chemistry, Leibniz-Institute for Natural Product Research and
13 Infection Biology, Beutenbergstr. 11a, 07745 Jena, Germany

14

15 #Corresponding author: gerald.lackner@leibniz-hki.de

16

17 **Running title:**

18 Biosynthesis and evolution of coenzyme F₄₂₀ congeners

19 **Keywords:**

20 bacterial metabolism, biosynthesis, coenzyme, enzyme catalysis, substrate specificity, X-
21 ray crystallography

1 **Abstract**

2 Coenzyme F₄₂₀ is a microbial redox cofactor that is increasingly used for biocatalytic
3 applications. Recently, diversified biosynthetic routes to F₄₂₀ and the discovery of a
4 derivative, 3PG-F₄₂₀, were reported. 3PG-F₄₂₀ is formed via activation of 3-phospho-D-
5 glycerate (3-PG) by CofC, but the structural basis of substrate binding, its evolution, as
6 well as the role of CofD in substrate selection remained elusive.

7 Here, we present a crystal structure of the 3-PG-activating CofC from *Mycetohabitans* sp.
8 B3 and define amino acids governing substrate specificity. Site-directed mutagenesis
9 enabled bidirectional switching of specificity and thereby revealed the short
10 evolutionary trajectory to 3PG-F₄₂₀ formation. Furthermore, CofC stabilized its product,
11 thus confirming the structure of the unstable molecule, revealing its binding mode and
12 suggesting a substrate channeling mechanism to CofD. The latter enzyme was shown to
13 significantly contribute to the selection of related intermediates to control the specificity
14 of the combined biosynthetic CofC/D step. Taken together, this work closes important
15 knowledge gaps and opens up perspectives for the discovery, enhanced biotechnological
16 production, and engineering of coenzyme F₄₂₀ derivatives in the future.

17 **Importance**

18 The microbial cofactor F₄₂₀ is crucial for processes like methanogenesis, antibiotics
19 biosynthesis, drug resistance, and biocatalysis. Recently, a novel derivative of F₄₂₀ (3PG-
20 F₄₂₀) was discovered, enabling the production and use of F₄₂₀ in heterologous hosts.
21 By analyzing the crystal structure of a CofC homolog whose substrate choice leads to
22 formation of 3PG-F₄₂₀, we defined amino acid residues governing the special substrate
23 selectivity. A diagnostic residue enabled reprogramming of the substrate specificity,

1 thus mimicking the evolution of the novel cofactor derivative and successfully guiding
2 the identification of further 3-PG-activating enzymes.

3 Furthermore, a labile reaction product of CofC was revealed that has not been directly
4 detected so far and CofD was shown to provide as another layer of specificity of the
5 combined CofC/D reaction, thus controlling the initial substrate choice of CofC. The
6 latter finding resolves a current debate in the literature about the starting point of F₄₂₀
7 biosynthesis in various organisms.

8 **Introduction**

9 Organic cofactors are small molecules other than amino acids that are required for the
10 catalytic activity of enzymes (1). They are therefore crucial for the understanding of the
11 biochemical and physiological processes their dependent enzymes are involved in, as
12 well as the biotechnological exploitation of these enzymes. Coenzyme F₄₂₀ is a
13 specialized redox cofactor that was so far mainly identified in archaea and some
14 actinobacteria (2). In archaea, F₄₂₀ is a key coenzyme of methanogenesis (3). In
15 mycobacteria, F₄₂₀ plays a vital role in respiration (4,5), cell wall biosynthesis (6,7), as
16 well as the activation of medicinally relevant antimycobacterial (pro-)drugs. For
17 instance, the novel anti-tubercular drug pretomanid is activated by Ddn, an F₄₂₀-
18 dependent nitroreductase (8,9). In streptomycetes, F₄₂₀H₂ is used for reduction steps
19 during the biosynthesis of antibiotics like thiopeptins (10), lanthipeptides (11), or
20 oxytetracycline (12,13). Increasing interest in F₄₂₀ is also driven by the utilization of
21 F₄₂₀H₂-dependent reductases in biocatalysis, e.g., for asymmetric ene reductions (14-
22 18).

1 Intriguingly, F₄₂₀ also occurs in a few Gram-negative bacteria where it has been acquired
2 most likely by horizontal transfer of its biosynthetic genes from actinobacteria (19,20).
3 Initial studies have revealed that F₄₂₀ is indeed produced by some of these organisms
4 but their physiological role remains unknown (20,21). We have recently identified F₄₂₀
5 biosynthetic genes in the genome of *Mycetohabitans* (synonym: *Paraburkholderia*)
6 *rhizoxinica* (22), a symbiont that inhabits the hyphae and spores of the phytopathogenic
7 mold *Rhizopus microsporus* (23-26). Surprisingly, we discovered that the symbiont
8 produced a novel derivative of F₄₂₀, which we termed 3PG-F₄₂₀ (22). The cofactor
9 activity of 3PG-F₄₂₀ was comparable to classical F₄₂₀ and could serve as a substitute for
10 the latter in biocatalysis (22). Although this congener has not been described in any
11 other organism, it could also be detected in the microbiota of a biogas production plant,
12 thus demonstrating that it is not restricted to endofungal bacteria (22). The producers of
13 3PG-F₄₂₀ in these habitats, however, are unknown.

14 The biosynthesis of 3PG-F₄₂₀ (**Figure 1**) is generally similar to the biosynthesis of
15 classical F₄₂₀ (27). The pathway starts with the formation of the redox-active core
16 moiety 7,8-didemethyl-8-hydroxy-5-deazariboflavin (F₀) from L-tyrosine and 5-amino-
17 6-ribitylamino-uracil, a reactive metabolite of the flavin biosynthesis pathway. The F₀
18 core is then elongated by a chemical group that can formally be described as 2-phospho-
19 L-lactate (2-PL) before an oligoglutamate tail is added. The biosynthesis of the 2-PL
20 moiety has been the subject of several studies. Seminal work on archaea suggested that
21 it is directly formed from 2-phospho-L-lactate: Incubation of cell extracts of
22 *Methanosa*cina *thermophila* or *Methanocaldococcus jannaschii* with F₀, 2-PL, and GTP
23 led to the formation of F₄₂₀-0 (28). Biochemical assays with purified CofC and CofD
24 finally corroborated the model that the guanylyltransferase CofC catalyzes the reaction

1 of 2-PL and GTP to lactyl-2-phospho-guanosine (LPPG) (29), which is then passed on to
2 CofD to transfer the activated 2-PL moiety onto the precursor F₀. However, the unstable
3 nature of LPPG has prevented confirmation of its structure by NMR or mass
4 spectrometry so far. The last biosynthetic step leading to the mature coenzyme F₄₂₀ is
5 catalyzed by the F₄₂₀:glutamyl ligase CofE (30), which is responsible for the addition of
6 the γ -linked oligoglutamate moiety to the F₄₂₀-0 core, thus forming F₄₂₀-n, with n
7 indicating the number of glutamate residues.

8 In mycobacteria, CofE is not a free-standing enzyme but constitutes the *N*-terminal
9 domain of the FbiB protein (31). It was shown recently that mycobacteria utilize
10 phosphoenolpyruvate (PEP), but not 2-PL, to form F₄₂₀-0. Instead of LPPG, EPPG is
11 formed, which is converted into dehydro-F₄₂₀-0 (DF₄₂₀-0) by the action of FbiA, the
12 mycobacterial CofD homolog. DF₄₂₀-0 is then reduced to classical F₄₂₀-0 by the *C*-
13 terminal domain of FbiB, which belongs to the nitroreductase superfamily (32). We have
14 shown that a similar pathway is present in the thermophilic bacterium
15 *Thermomicrobium roseum* and related species (33). The formation of 3PG-F₄₂₀-0,
16 however, does not require any reduction step. Instead, enzyme assays revealed that 3-
17 phospho-D-glycerate (3-PG) is activated by CofC, presumably forming the short-lived
18 intermediate 3-(guanosine-5'-diphospho)-D-glycerate (GPPG), which is further
19 transferred to the F₀ core by the action of CofD.

20 However, it remained elusive, which amino acid residues within the CofC protein
21 conferred the specificity switch towards 3-PG and how genetic mutation might have led
22 to the evolution of 3PG-F₄₂₀ biosynthesis. Furthermore, the question persisted, why the
23 CofC CofD reaction only proceeds as a combined reaction and how reactive
24 intermediates like LPPG are stabilized. Another open question concerned the role of 2-

1 PL in the biosynthesis of F₄₂₀ in archaea. While our previous data (22) matched seminal
2 observations (29) of a substantial turnover of 2-PL by CofC enzymes of archaeal origin,
3 other studies raised doubts that 2-PL is a genuine substrate of archaeal CofC homologs
4 (32).

5 Here, we present a crystal structure of the 3-PG activating CofC from *Mycetohabitans* sp.
6 B3 and revealed the amino acid residues governing 3-PG activation. By site-directed
7 mutagenesis, we shed light on the evolution of 3PG-F₄₂₀. We furthermore bring to
8 attention that CofC strongly binds its product GPPG and collaborates closely with its
9 partner CofD to control the flux of intermediates into the F₄₂₀ biosynthesis pathway.

10 **Results**

11 **Assessment of substrate specificities of CofC enzymes from several organisms**

12 To gain a better understanding of CofC substrate specificities, we set out to identify
13 more homologs of CofC accepting 3-PG as a substrate. We reasoned that related bacteria,
14 harboring CofC homologs highly similar to the *M. rhizoxinica* enzyme (*Mrhiz*-CofC),
15 would have a similar substrate preference. Indeed, using LC-MS we detected 3PG-F₄₂₀
16 (**Figure 2A/B**) in cell extracts of *Mycetohabitans* sp. B3, a close relative of *M. rhizoxinica*
17 that shares the same lifestyle as a symbiont of a phytopathogenic *Rhizopus microsporus*
18 strain. Neither classical F₄₂₀ nor DF₄₂₀ was detectable.

19 Next, we produced CofC of *Mycetohabitans* sp. B3 (*MycB3*-CofC, accession number
20 KQH55_09515) as a hexahistidine fusion-protein in *E. coli*, purified the enzyme by Ni-
21 NTA affinity chromatography, and investigated its activity in a combined CofC/D assay
22 using CofD from *Methanocaldococcus jannaschii* (*Mjan*-CofD). When the substrates were
23 provided in equal concentrations in a competitive assay, *MycB3*-CofC (**Figure 2C**)

1 accepted 3-PG (65%), 2-PL (26.5%), and PEP (8.5%), a profile that was similar to the
2 one obtained previously for the *Mrhiz*-CofC (22).

3 To obtain an overview of the substrate specificities of CofC we re-assessed CofC enzymes
4 from well-studied F₄₂₀ producing organisms such as *Mycolicibacterium smegmatis*,
5 *Thermomicrobium roseum*, and *Methanosaeca mazaei* (22) and assayed CofCs from
6 further Gram-negative bacteria like *Paracoccus denitrificans*, *Oligotropha*
7 *carboxidovorans* as well as the uncultivable *Candidatus (Ca.) Entotheonella* factor TSY1
8 that is rich in genes encoding F₄₂₀-dependent enzymes (21). For all CofC-related
9 enzymes analyzed, 2-PL was used most efficiently from all substrates compared. We
10 observed PEP turnover in the range of 3.5% to 30% (**Figure 2C**). Generally, it can be
11 concluded that CofC assays cannot discriminate whether 2-PL or PEP is the relevant
12 substrate *in vivo*. The only CofC that accepted 3-PG to a certain extent (8%) was the
13 enzyme from *Ca. E. factor*. However, compared to the *Mycetohabitans* enzyme there was
14 no significant preference of 3-PG over PEP.

15 **Identification of 3-PG-binding residues of CofC**

16 Next, we aligned primary amino acid sequences of CofC homologs to identify the
17 residues that might be responsible for the altered substrate preference (**Figure 3A**). A
18 crystal structure of FbiD from *Mycobacterium tuberculosis* (*Mtb*-FbiD) in complex with
19 PEP (PDB: 6BWH) showed eight amino acid residues to be in close contact with PEP
20 suggesting a role in conferring the substrate specificity (32). Three of them are aspartate
21 residues (D116, D188, D190) that complex two Mg²⁺ ions which in turn interact with the
22 phosphate group of PEP. The remaining residues were supposed to bind the PEP
23 molecule via side-chain atoms (K17, L92, S166) or backbone amino groups (T148 and
24 G163). While most of these residues were highly conserved, two alignment positions

1 showed a deviation in those residues able to activate 3-PG, namely L92 and G163 of *Mtb*-
2 FbiD. While L92 is replaced by methionine (M91), the residue corresponding to FbiD-
3 G163 was replaced by serine (S162) in *Mrhiz*-CofC and *MycB3*-CofC. Homology modeling
4 further suggested H145 to be a potential critical residue for 3-PG binding and C95 to be
5 involved in the correct positioning of M91 (**Figure 3D, Figure S1**).

6 **Mutagenesis of CofC reveals S162 to be crucial for 3-PG activation**

7 To probe the role of the suggested residues, we performed site-directed mutagenesis in
8 *Mrhiz*-CofC (**Figure 3B**). Especially S162 turned out to be critical for 3-PG activation.
9 While the least invasive mutation, S162T, retained most of the activity of wild-type CofC
10 towards 3-PG, all other mutants of this residue preferentially turned over 2-PL and, to a
11 lesser extent, PEP. This finding suggested that the hydroxy group present in S162 of WT
12 and S162T might support the recruitment of 3-PG to the active site. M91L displayed
13 reduced activation of 3-PG, while M91A was not impaired in 3-PG activation. Possibly,
14 M91 controls the size of the substrate-binding pocket thus hindering (M91L) or
15 facilitating (M91A) access of the larger substrate 3-PG to the active site. The C95A
16 mutant approximately retained wild-type activity towards 3-PG, while C95L strongly
17 reduced 3-PG activation. This was an indication that C95 might indeed affect the
18 orientation of M91 and as a consequence 3-PG binding. Finally, the proposed interaction
19 of 3-PG with H145 was not reflected in altered specificity profiles of H145A and H145T
20 mutants of *Mrhiz*-CofC.

21 **Engineering *M. smegmatis* FbiD into a 3-PG activating enzyme**

22 Inspired by the finding that S162 of the *Mrhiz*-CofC is necessary for 3-PG activation we
23 wondered if mutation of the corresponding residue G169 of FbiD to serine (**Figure 3C**)
24 could turn FbiD from *M. smegmatis* (*Msmeg*-FbiD) into a 3-PG activating enzyme,

1 thereby imitating the molecular processes underlying the evolution of 3PG-F₄₂₀
2 biosynthesis. The G169S mutant, however, did not accept any 3-PG as substrate. We also
3 tried to mutate L98 to facilitate the entry of 3-PG into the active site. However, neither
4 the single mutant L98C, nor the double mutant G169S;L98C enabled 3-PG binding. Based
5 on the homology model we suspected a residue corresponding to H145 of the *Mrhiz*-
6 CofC might facilitate 3-PG binding. Indeed, while the single mutant T152H did not show
7 any significant effect, the double mutant G169S;T152H successfully turned over 3-PG.
8 The triple mutant (G169S;T152H;L98C) resulted in insoluble protein. Overall, these
9 results showed that substrate specificity of *Msmeg*-FbiD can be readily switched by
10 changing only two residues and again supported a prime role of the critical serine
11 residue for 3-PG recruitment.

12 **Structural insights into 3-PG activation by CofC.**

13 After several attempts had failed to crystallize recombinant *Mrhiz*-CofC we turned to
14 *MycB3*-CofC that was more soluble despite only minor differences in the amino acid
15 sequence. From diffraction data collected to 2.4 Å the crystal structure could be solved
16 by molecular replacement with a model of two superimposed structures (*Mtb*- FbiD and
17 *Methanosarcina mazei* CofC, *Mmaz*-CofC). The overall structure was similar to the known
18 homologs with the core of the single-domain protein being a six-stranded mixed β-sheet
19 (**Figure 4**).

20 Intriguingly, after initial refinement both molecules in the asymmetric unit
21 independently showed unambiguous difference density for GPPG, the reaction product
22 of GTP with 3-PG, completely immersed into the active site pocket. Each building block
23 of GPPG is bound by several interactions (**Figure 4B, Table S1**). Guanine is
24 distinguished from adenine by two H-bond donors to its oxygen (main-chain nitrogen of

1 V65 and P86) and two H-bond acceptors to its amino group (main-chain carbonyl group
2 of E89 and G90). The α - and β -phosphates are bound by two Mg^{2+} ions, which in turn are
3 positioned by three aspartates (D116, D193, D191) similarly as had been shown for
4 FbiD before (32). In the homologous structures the binding site for guanosine and ribose
5 is almost completely conserved. Hence, GPPG is tightly bound, in fact, it turned out to be
6 still quantitatively bound after purification of the protein from *E. coli* cell lysate, where
7 both substrates were present.

8 The 3-PG moiety is primarily bound by H-bonds with the side chain of S165 to the
9 carboxy group, apparently a highly conserved binding site of the carboxylate of the C3-
10 acid as known from *Mtb*-FbiD (32). More interactions with the carboxy group arise from
11 H-bonds with main-chain nitrogen atoms of T147 and S162, the latter being critical for
12 selective 3-PG activation. Although the mild effect of S162T suggested otherwise, S162
13 did not interact via its hydroxy group with the ligand. Furthermore, there are H-bonds
14 from the main-chain carbonyl groups of T147 and N148 to the 2-hydroxy group of 3-PG.
15 Surprisingly, these binding partners of the 3-PG hydroxy group are structurally very
16 well conserved and thus not likely to be involved in discrimination between 3-PG and 2-
17 PL/PEP. The residue corresponding to *Mtb*-FbiD-K17 (*Mrhiz*-CofC-K20) did not interact
18 with the substrate however, instead, K26 has taken over its role.

19 Taken together, the majority of the residues corresponding to the binding pocket
20 residues of the PEP-binding FbiD (32) were also found to be involved in 3-PG binding.
21 Although all C3-acids form the same hydrogen bonds, the carboxylate rotates by 36°
22 about the axis through the carboxylate defined by the oxygens. This moves the
23 phosphate by 2-3 Å. The hydrogen bonds for PEP have a more favorable geometry

1 (average out-of- π -plane distortion 0.74 Å) than those for 3-PG (1.38 Å) but the
2 phosphate group will not be positioned properly anymore to attack the α -phosphate of
3 GTP. 3-PG with one more bond (carboxylate-C2-C3-O-P compared to carboxylate-C2-O-P
4 in PEP or 2PL) can compensate for the new orientation of the carboxylate. The reason
5 why the 3-PG adopts a new orientation is the main-chain rotation of S162, the CA and CB
6 of which then squeeze the 3-PG into the productive conformation.

7 Based on the position of GPPG, the binding site for GTP is evident for the GMP moiety.
8 Differential scanning fluorimetry (nano-DSF) measurements further corroborated the
9 direct binding of GTP by CofC ($K_D < 20 \mu\text{M}$) (see **Supplementary Methods, Figure S3**).
10 The β -phosphate can either bind in the same position as the second phosphate of GPPG
11 or point outward into the solvent close to R28N. The latter conformation allows the
12 second substrate 3-PG to bind like PEP in FbiD. GTP and 3-PG are then positioned well
13 for the reaction, the nucleophilic attack of the 3-PG phosphate on the α -phosphate of
14 GTP (**Figure 4D**).

15 **Identification of further 3-PG accepting enzymes**

16 After establishing that serine or threonine in the position corresponding to *Mrhiz*-CofC-
17 S162 are linked to 3-PG formation, we hypothesized that the residue might be exploited
18 as a diagnostic residue to identify further 3-PG activating enzymes. Going beyond highly
19 related *Mycetohabitans* species, which can be expected to be 3PG-F₄₂₀ producers,
20 database searches revealed candidate proteins from as-yet uncultivated archaeal species
21 (**Figure S2A**) that contained serine or threonine at the critical alignment position.

22 Since their source organisms were not accessible, we obtained the coding sequences of
23 three of these candidate enzymes as synthetic genes and tested their substrate

1 specificities (**Figure S2B**). Interestingly, all of those enzymes accepted 3-PG as
2 substrates. The circumstance that 2-PL was the best substrate of all three enzymes does
3 not rule out the possibility that these enzymes are involved in 3PG-F₄₂₀ formation given
4 the fact that 2-PL is the default case for many enzymes examined in our assay system
5 even if PEP is the natural substrate. Notably, two of the enzymes tested did not accept
6 PEP as a substrate, a rather unusual finding. In the absence of 2-PL, this profile would
7 result in the production of 3PG-F₄₂₀. Taken together, S162 represents a diagnostic
8 residue to predict the specificity of CofC towards 3-PG.

9 **Evolution of 3-PG accepting enzymes**

10 To answer the question how 3-PG accepting enzymes might have evolved, we
11 constructed a phylogenetic tree of CofC enzymes examined in this study (**Figure 5**). The
12 *Mycetohabitans* CofC clade branched off early in the evolution of bacterial CofC enzymes
13 and is neither closely related to nor derived from actinobacterial CofC/FbiD nor to other
14 CofC enzymes found in Gram-negative bacteria. The archaeal 3-PG activating enzymes
15 represent a monophyletic clade within the archaeal proteins. Taken together, we
16 conclude that 3-PG activation has evolved independently in evolution at least twice from
17 an ancestral 2-PL/PEP activating enzyme.

18 **Role of CofD in substrate specificity of F₄₂₀ side-chain biosynthesis**

19 After gaining insights into the substrate specificity of CofC a few questions remained. For
20 instance, in almost all CofC homologs tested, 2-PL was the preferred substrate. This
21 finding contrasted previous results obtained by Bashiri and colleagues (32) for *Mtb*-FbiD
22 and CofC from *M. jannaschii* (*Mjan*-CofC), which was reported to accept exclusively PEP
23 (32). Furthermore, the residual activity of *Mrhiz*-CofC and *MycB3*-CofC towards PEP
24 could not explain why PEP-derived DF₄₂₀ was not found in their source organisms (22).

1 We therefore assumed that the choice of the CofD homolog used in the combined CofC/D
2 assay might have an impact on the overall outcome of the assay.

3 To test this hypothesis, we produced CofD homologs of several model species as
4 hexahistidine-fusion proteins and performed CofC/D assays using several combinations
5 of CofC and CofD. Strikingly, the choice of CofD homologs had a significant influence on
6 the product spectrum of the CofC/D pair (**Figure 6**). For instance, when *Mrhiz*-CofC and
7 *MycB3*-CofC were combined with their cognate CofD, the apparent substrate specificity
8 shifted almost completely towards 3-PG. The PEP- and 2-PL-derived products were only
9 produced in traces by the CofC/D reaction. Similarly, when *Msmeg*-FbiD was combined
10 with its natural partner *Msmeg*-FbiA instead of the homolog encoded in *M. jannaschii*
11 (*Mjan*-CofD), the apparent activity towards 2-PL was almost entirely abolished.
12 Obviously, CofD homologs have the ability to select between the pathway intermediates
13 LPPG (2-PL-derived), EPPG (PEP-derived), and GPPG (3-PG-derived). In bacterial
14 systems, CofD/FbiA appears to favor the intermediate that is known to be relevant in
15 their source organisms, i.e., EPPG for Mycobacteria or GPPG for *M.*
16 *rhizoxinica/Mycetohabitans* sp. B3. The archaeal *Mjan*-CofD, however, appears to prefer
17 its natural substrate LPPG but displays relaxed specificity towards EPPG and LPPG.

18

1 **Discussion**

2 **Structural basis of CofC specificity**

3 Extensive characterization of various CofC homologs, mutagenesis studies, and
4 crystallography enabled us to spot residues responsible for the unusual substrate choice
5 of CofC from *Mycetohabitans*. The crystal structure obtained from *MycB3*-CofC revealed
6 that most of the amino acid positions described for PEP-binding *Mtb*-FbiD play a role in
7 3-PG binding as well. However, rather than specific interactions with the free 2-hydroxy
8 moiety, it is the conformation of S162 that forces the substrate into a position from
9 which only the larger substrate 3-PG can undergo productive reaction of its phosphate
10 group with GTP. The effect of M91 and C95 on substrate specificity shows that indirect
11 influences on the overall conformation of the active site can be crucial for the correct
12 positioning of the substrate. Nevertheless, the S162 residue proved as diagnostic residue
13 to predict further 3-PG activating enzymes and even enabled engineering of *Msmeg*-FbiD
14 into a 3-PG activating enzyme. Notably, a bidirectional change of the substrate
15 specificity, i.e. from 3-PG to 2-PL/PEP in *Mrhiz*-CofC as well as from 2-PL/PEP towards
16 3-PG in *Msmeg*-FbiD FbiD as described here is a rather exceptional achievement.

17 **Evolution and occurrence of 3PG-F₄₂₀ in nature**

18 Interestingly, our mutagenesis study answers the question of how 3PG-F₄₂₀ might have
19 originated via mutation of 2-PL/PEP activating CofC on a molecular level. The
20 phylogenetic tree suggests that 3-PG activating enzymes have evolved two times
21 independently from an ancestral 2-PL/PEP activating CofC. Since DF₄₂₀ is less stable
22 than saturated forms, we suppose that the metabolic switching event has occurred to
23 enable the formation of a stable F₄₂₀-derivative in a metabolic background that lacked 2-

1 PL or the DF₄₂₀ reductase that is present in Actinobacteria (32) and Thermomicrobia
2 (33). Here, we showed that the exchange of two amino acids in CofC/FbiD is mainly
3 affecting substrate specificity and is thus sufficient to mimic this evolutionary process in
4 the laboratory.

5 Considering that 3PG-F₄₂₀ was detectable in biogas-producing sludge (22), there must
6 be microorganisms outside the monophylogenetic clade of endofungal bacteria
7 (*Mycetohabitans*) that produce 3PG-F₄₂₀. The exploitation of the critical S162 in the
8 active site indeed enabled us to identify further 3-PG activating enzymes, albeit from
9 uncultured organisms. The source organisms may be related to the 3-PG producing
10 organisms found in the digester. Efforts to isolate further 3PG-F₄₂₀ producers remain
11 ongoing and the knowledge gained here will facilitate this project in the future.

12 **CofD influences substrate specificity of CofC *in vivo* and *in vitro***

13 Another important key finding of this study is that CofC and CofD together contribute to
14 the substrate specificity of the combined reaction, where CofD of some bacteria seems to
15 represent a restrictive filter that acts after the more promiscuous CofC. This finding can
16 explain the before-mentioned inconsistencies between results obtained *in vivo* and *in*
17 *vitro*.

18 Even more importantly, we can now resolve the discrepancies concerning the results of
19 CofC/D assays performed with CofC from archaea existing in the literature. Seminal
20 work identified 2-PL to be a suitable substrate of archaeal CofC enzymes and thus
21 proposed the original biosynthetic pathway to start from 2-PL (29). In contrast, a more
22 recent study did not observe any turnover of 2-PL neither using the archaeal CofC from
23 *M. jannaschii*, nor using FbiD from *M. tuberculosis* and suggested a biosynthetic route

1 starting from PEP via EPPG, even for archaea (32). A follow-up study delivered further
2 evidence supporting the biosynthetic route via DF₄₂₀ in mycobacteria. The authors
3 showed the formation of DF₄₂₀ from PEP in cell-free extracts of *M. smegmatis* and could
4 reveal strong binding of DF₄₂₀ to the active site of FbiA (CofD) by X-ray crystallographic
5 studies of the enzyme (34).

6 Our previous study (22), however, found 2-PL to be the best substrate of the *M.*
7 *jannaschii* CofC enzyme - a result that we could reproduce here using CofC from *M.*
8 *mazei*. It is also confusing that even the *M. smegmatis* FbiD tested in this study preferred
9 2-PL as substrate, again challenging the hypothesis that PEP is the preferred substrate of
10 FbiD.

11 The solution to this perplexing situation comes with the herein defined influence of
12 CofD/FbiA on the overall specificity of the CofC/D pair. CofD selects between the
13 unstable pathway intermediates LPPG, EPPG, and GPPG (27). According to our results,
14 *Msmeg*-FbiA exclusively accepted EPPG to form DF₄₂₀. Bashiri *et al.* used the closely
15 related *Mtb*-FbiA to perform CofC/D assays (32). Since *Mjan*-CofC can accept PEP as a
16 minor substrate when combined with its cognate CofD (22), the assay resulted
17 exclusively in DF₄₂₀-formation when combined with FbiA, thus erroneously suggesting
18 that 2-PL was not accepted by CofC. Similarly, the activation of 2-PL by FbiD remained
19 “hidden” when assays were carried out with FbiA as a partner enzyme. Conversely, it is
20 plausible that the “supernatural” turnover of 2-PL observed in all our assays performed
21 with *Mjan*-CofD might be an artifact caused by the choice of a CofD homolog that might
22 preferably turn over its natural substrate LPPG. For future studies towards the
23 biosynthesis of novel F₄₂₀ derivatives we, therefore, suggest that only a combination of
24 CofC and its cognate CofD is suitable to reflect the *in-vivo* situation.

1 **The combined CofC and CofD reaction**

2 The X-ray structure of *MycB3*-CofC presented here included the reaction product GPPG,
3 while the previous crystal structure of FbiD was obtained in the presence of PEP only
4 (32). This is the first direct analytical evidence for the labile reaction product GPPG. So
5 far, the existence of its congener LPPG was only confirmed by chemical synthesis
6 followed by successful turnover by CofD (35). The fact that GPPG remains tightly bound
7 to the enzyme points to a substate-channeling mechanism where the GPPG molecule is
8 directly transferred to CofD to avoid degradation of the labile intermediate in the
9 absence of F₀. Substrate inhibition could also explain why any attempt to measure the
10 activity of CofC in the absence of CofD remained unsuccessful (27,32) and why direct
11 detection of GPPG or the related LPPG and EPPG from solution has failed so far.

12 The binding mode of GPPG also clearly revealed the GTP binding site of CofC. Notably, no
13 evidence for GTP binding could be obtained experimentally for *Mtb*-FbiD and it was
14 even speculated that GTP binding might require the presence of FbiA (32). This
15 mechanism could be disproved for CofC of *M. rhizoxiniae*.

16 **Conclusion**

17 Taken together, this study represents a significant advance in understanding the
18 flexibility of substrate specificity in CofC homologs and offers a molecular model for the
19 evolution of 3PG-F₄₂₀. By direct detection of the instable reaction product GPPG via X-ray
20 crystallography we gained insights into the structural basis of the combined CofC/D
21 reaction. The demonstration that CofC and CofD cooperate closely to control the entry of
22 central carbon metabolites into the biosynthetic route to F₄₂₀ derivatives also solved an
23 ongoing debate in the literature and thereby re-established 2-PL as the most likely
24 starting point of F₄₂₀ biosynthesis in archaea. Future perspectives are opened up to

1 investigate the cooperation of biosynthetic enzymes on a molecular level and to exploit
2 this knowledge gained here for enhanced biotechnological production of coenzyme F₄₂₀
3 and potentially novel derivatives.

4 **Materials and Methods**

5 **Chemicals and microbial strains and metabolite extraction**

6 Chemicals and media components were purchased from Acros Organics, Sigma Aldrich,
7 Alfa Aesar, Carl Roth, and VWR. The *E. coli* strains BL21 (DE3), LOBSTR-BL21 (Kerafast),
8 ETH101 and Top10 were grown routinely in Lysogeny Broth (LB). Axenic
9 *Mycetohabitans* sp. B3 (NCBI genome accession J AHLKN000000000) was grown in 50
10 ml MGY media (22) at 30 °C at 110 rpm for one week. Cultures (OD₆₀₀ = 2.5) were
11 lyophilized, extracted with 10 mL ice-cold methanol, sonicated for 20 min and incubated
12 at 250 rpm for 60 min. Cell debris was removed by centrifugation (4 °C, 10 min, 10000
13 rpm) and the supernatant was filtered into a round-bottom flask. The extract was dried
14 in a vacuum rotary evaporator at 40 °C and re-dissolved in 1 mL LC-MS-grade water.
15 Samples were analyzed using LC-MS as described before (22).

16 **Construction of expression vectors**

17 Unless stated otherwise, primers (**Table S2A**) were designed using the software tool
18 Geneious (36) and cloning was based on DNA recombination following the Fast Cloning
19 protocol (37). The *E. coli* Top10 strain was used to propagate plasmids. PCR reactions
20 were carried out using Q5 High-Fidelity polymerase (New England Biolabs) and
21 oligomers used for amplifications listed in Table S1. Constructed plasmids (**Table S2B**)
22 were confirmed by Sanger sequencing (Eurofins Genomics). CofC and CofD encoding
23 plasmids (pMH04, 05, 10, 18, 19, 20,) were purchased from BioCat as codon-optimized

1 synthetic gene construct cloned into pET28a+ between *Bam*HI and *Hind*III restriction
2 sites. Plasmid sequences of pMH43 (pACYCDuet backbone), pMH59 (pET28), and
3 pMH60 (pET28) were obtained by gene synthesis (BioCat), sequences are provided in
4 the Supplemental Material (**Appendix**).

5 **Site-directed mutagenesis**

6 Putative substrate-binding residues of CofC were subjected to site-directed mutagenesis
7 on DNA level using PCR (38). Amino acid numbering corresponds to the original residue
8 position in the native CofC protein.

9 To mutate CofC from *M. rhizoxinica* (pFS03) at position 162 (S162), primer pairs
10 oMH03/05, oMH05/06, oMH07/08, and oMH23/24 were used to generate S162E
11 (resulting plasmid pMH22), S162Y (pMH23), S162G (pMH24), S162A (pMH25) and
12 S162T (pMH26), respectively. Exchange of cysteine at position 95 (C95) was achieved
13 using primer pairs oMH102/103 (C95L, pMH66) oMH104/105 (C95A, pMH67). The
14 double-mutant C95L;S162G (plasmid pMH68) was obtained from pMH24 by
15 amplification with primer pair oMH106/107. Mutation of histidine at position 145
16 either to alanine (H145A) or threonine (H145T) was obtained using primer pairs
17 oMH116/117 (pMH74) or oMH118/119 (pMH75). Methionine at position 91 was
18 substituted with alanine (M91A) or leucine (M91L) to yield pMH76 and pMH77.

19 *M. smegmatis* *cofC* (pMH10) served as a template for substituting glycine residue in
20 position 169 (G169). Primer pairs (oMH33/34) or (oMH37/38) or (oMH45/46) or
21 oMH47/48 were used to obtain G169S (pMH32), G169Y (pMH33), G169A (pMH34) or
22 G169E (pMH35), respectively. Primer pairs oMH108/109 were used to substitute
23 leucine (L98) with cysteine (L98C, pMH70). The double-mutants G169S; L98C and

1 G169S;T152H were obtained from plasmid pMH32 using primer pair oMH110/111 and
2 oMH124/125, yielding plasmid pMH69 and pMH80, respectively. The triple mutation
3 was introduced in the pMH69 using oMH124/125 primers resulting in plasmid pMH81
4 (G169S;L98C;T152H).

5 **Heterologous protein production and purification**

6 Production condition and purification of all *N*-terminal hexahistidine (N-His₆) tagged
7 proteins (CofC and CofD) were similar as described before (22). Accession numbers of
8 native CofC/FbiD and CofD/FbiA proteins are listed in **Table S3**. In short, chemical
9 competent *E. coli* BL21 (DE3) or LOBSTR-BL21 cells were transformed with individual
10 CofC/CofD encoding plasmids and the respective antibiotic (kanamycin 50 µg/mL or
11 chloramphenicol 25 µg/mL) was used to maintain selection pressure. Correct positive
12 clones were grown overnight at 37 °C and 180 rpm and used to inoculate fresh 100 mL
13 cultures (1:100). Upon reaching late exponential growth phase (OD₆₀₀ = 0.7), expression
14 of the gene was induced by the addition of 1 mM IPTG and incubated (18 °C, 180 rpm)
15 following 18 hours for protein production. After harvesting, cells were disrupted with
16 pulsed sonication. The clear cell lysate was loaded onto a Ni-NTA affinity column to
17 separate N-His₆ tagged protein. Later on, the protein was eluted with a higher
18 concentration of imidazole (500 mM) and re-buffered in a PD-10 column.

19 **Combined CofC/D assay**

20 Distinct derivatives of F₄₂₀-0 were produced via biochemical reaction of purified CofC
21 and CofD proteins in a combined assay (22,29). In-vitro reaction conditions were
22 analogous to Braga *et al.* (22) and 50 µl reaction consisted of 100 mM HEPES buffer (pH
23 7.4), 2 mM GTP, 2 mM MgCl₂, 0.14 nM Fo, 34 µM CofD, and 0.5 mM of substrates (3-
24 phospho-D-glyceric acid, phosphoenolpyruvic acid, and 2-phospho-L-lactate). The

1 reactions were initiated upon the addition of 26 μ M CofC. Reactions were quenched with
2 one volume of acetonitrile and formic acid (20%). Production of F₄₂₀-0 derivatives was
3 monitored in LC-MS. Technical set up, method, conditions for LC-MS analysis were
4 similar as described before (22). Data analysis followed extraction of ion
5 chromatograms (XICs), calculation of area under the curve (AUC), normalization of AUC,
6 plotting area against time, and product formation was calculated for a linear time range
7 (0 to 20 min). Quantification of relative product formation was determined from three
8 biological replicates (n = 3) and plotted as bar charts. Standard deviations (SD) were
9 used as error bars.

10 **CofC sequence alignment and phylogenetic tree inference**

11 Multiple protein sequences of CofC from different F₄₂₀ producing organisms (**Table S3**)
12 were retrieved from the NCBI database and primary sequences were aligned based on
13 their predicted structure using Expresso (T-Coffee) (39). For phylogenetic tree inference
14 as implemented in Geneious Prime (36), the MUSCLE algorithm (40) was used to align
15 sequences and a Maximum Likelihood tree was inferred using PhyML 3.0 (41) with the
16 LG model for protein evolution and a gamma distribution of rates. Support values
17 (Shimodaira-Hasegawa-like branch test) were computed and are shown above
18 branches. Trees were visualized in Geneious Prime.

19 **Structural Modeling**

20 The Phyre2.0 web portal was used to obtain a structural model for *M. rhizoxinica* CofC
21 (PrCofC) (42). This enabled the identification of three enzymes (PDB: C3GX, PDB: 2I5E,
22 PDB: 6BWH) that were used as a template for structural modeling resulting in models
23 with 100% confidence in the fold. The enzyme FbiD from *M. tuberculosis* H37Rv (PDB:
24 6BWH) was used for further analyses. Initial structural alignment based on short

1 fragment clustering of *M. rhizoxinica* CofC and FbiD was performed by program GESAMT
2 (General Efficient Structural Alignment of Macromolecular Targets) from the CCP4i2
3 V1.0.2 program suite (43-45). This superimposed a total of 189 residues with an rmsd of
4 1.687 Å. FbiD binds to two Mg²⁺ ions that are important for catalysis and PEP binding
5 (32). To place the Mg²⁺ ions in CofC, the CofC model was superposed on FbiD using
6 residues surrounding the Mg²⁺ binding site. This model was used as a template for
7 molecular docking of GTP into the CofC model using AutoDoc Vina. The input PDBQT
8 files for AutoDoc Vina were generated with AutoDoc Tools V1.5.6 (46). The PRODRG
9 server was used to generate the three-dimensional coordinates for GPPG from two-
10 dimensional coordinates (47). The 3-PG was manually modeled in CofC•GTP using COOT
11 (48). Representations of structures were prepared using PyMOL Molecular Graphics
12 System (Schrödinger, LCC). The Adaptive Poisson-Boltzmann Solver (APBS)
13 electrostatics plugin in PyMOL was used for the electrostatic surface representation
14 (49).

15 **Crystallization and data collection**

16 *Mycetohabitans* B3 CofC was further purified on a size exclusion column (Superdex75,
17 16/600, Cytiva) and concentrated in 50 mM Tris pH 7.4, 100 mM NaCl, 5 mM MgCl₂, 2 mM
18 mercaptoethanol (SEC buffer) to 8.2 mg/mL. Sitting drop crystallization trials were set
19 up with screens Wizard I and II (Rigaku), PEG/Ion (Hampton Research), and JBScreen
20 (Jena Bioscience) using 0.3 µL protein solution and 0.3 µL reservoir. Crystals appeared
21 after 2 weeks with reservoir 10% PEG 3000, 200 mM MgCl₂, 100 mM sodium cacodylate
22 pH 6.5. After briefly soaking in a reservoir with 20% glucose added crystals were
23 cryocooled in liquid nitrogen. Diffraction data were collected at BESSY, beamline 14.1.

1 Data collection parameters are given in **Table S4**. Data sets were processed with
2 XDSAPP (50).

3 **Structure solution and refinement**

4 Programs used for this part were all used as provided by CCP4 (45). Sequence search
5 against the PDB revealed structures of two related proteins, FbiD from *Mycobacterium*
6 *tuberculosis* and CofC from *Methanosa**cina mazae* with sequence identities of 30%.
7 Molecular replacement by PHASER with 6BWG (*Mtb*-FbiD) or 2I5E (*Mmaz*-CofC) did not
8 solve in the automatic mode, an assembly of both structures superimposed and
9 truncated (residues 9-82, 91-173, 178-211 of monomer A of 6BWG and aligned residues
10 of monomer A of 2I5E) solved the phase problem with an LLG of 233. The 6BWG
11 structure was used as starting model for replacing the sequence with the
12 *Mycetohabitans* sequence, one round of automatic model building with BUCANEER (51)
13 and iterative rounds of refinement with REFMAC5 (52) and manual model building with
14 COOT (53) completed the two protein chains.

15 Unambiguous water molecules were added when R_{free} reached 0.315 and a GDP moiety
16 with two Mg^{2+} ions was built into the difference density in the active site. Pertaining
17 difference density was connected to the β -phosphate suggesting a covalently bound PEP
18 or 3-PG. Only the latter refined without residual difference density above $\pm 2\sigma$. For final
19 refinement, non-crystallographic symmetry was not used. TLS refinement was applied
20 with one group per monomer. Refinement statistics are given in **Table S4**.

21 The C3-acid substrates are fixed in the active site by three H-bonds of their carboxylate
22 group. For H-bonds to π -systems (peptide bond, carboxylate group) the partner should

1 lie in the same plane. To characterize deviations from favorable H-bonding geometry we
2 calculated the average distance of the partner to the π -plane.

3 **Data deposition**

4 The crystal structure of CofC from *Mycetohabitans* sp. B3 was deposited at the protein
5 database PDB (PDB code: 7P97).

6

7 **Author contributions**

8 MH, DB (cloning, CofC/D assays, mutagenesis, LC-MS analysis), and DL (diagnostic
9 sequence analysis) performed research and analyzed data, and contributed to writing
10 the manuscript. IR performed research (cultivation and metabolite extraction). SS and
11 LB performed research (crystallization, GTP binding assays). GJP and ML analyzed data
12 (X-ray crystallography, structural modeling, GTP binding assays) and contributed to
13 writing the manuscript. GL performed research (phylogenetic reconstruction), designed
14 the study together with ML, and wrote the manuscript.

15 **Acknowledgments**

16 GL, MH and DB thank the Carl Zeiss Foundation, GL and DL the German Research
17 Foundation (DFG; Deutsche Forschungsgemeinschaft) Grant LA4424-1/1 (Project:
18 408113938) and the Leibniz Association for funding. ML and SS thank the DFG Grant
19 LA2984-5/1 (Project: 389564084) and LA2984-6/1 (Project: 449703098) for funding.
20 We declare there are no conflicts of interest. IR is grateful for financial support from the
21 European Union's Horizon 2020 Research and Innovation Programme under the Marie
22 Skłodowska-Curie grant agreement No. 794343.

23 **References**

1. Fischer, J. D., Holliday, G. L., Rahman, S. A., and Thornton, J. M. (2010) The structures and
2. physicochemical properties of organic cofactors in biocatalysis. *J Mol Biol* **403**, 803-824
3. 2. Greening, C., Ahmed, F. H., Mohamed, A. E., Lee, B. M., Pandey, G., Warden, A. C., Scott, C.,
4. Oakeshott, J. G., Taylor, M. C., and Jackson, C. J. (2016) Physiology, biochemistry, and
5. applications of F₄₂₀⁻ and F_o-dependent redox reactions. *Microbiol Mol Biol Rev* **80**, 451-493
6. 3. DiMarco, A. A., Bobik, T. A., and Wolfe, R. S. (1990) Unusual coenzymes of methanogenesis.
7. *Annu Rev Biochem* **59**, 355-394
8. 4. Gurumurthy, M., Rao, M., Mukherjee, T., Rao, S. P. S., Boshoff, H. I., Dick, T., Barry, C. E., and
9. Manjunatha, U. H. (2013) A novel F420-dependent anti-oxidant mechanism protects
10. *Mycobacterium tuberculosis* against oxidative stress and bactericidal agents. *Mol Microbiol*
11. **87**, 744-755
12. 5. Lee, B. M., Harold, L. K., Almeida, D. V., Afriat-Jurnou, L., Aung, H. L., Forde, B. M., Hards, K.,
13. Pidot, S. J., Ahmed, F. H., Mohamed, A. E., Taylor, M. C., West, N. P., Stinear, T. P., Greening,
14. C., Beatson, S. A., Nuermberger, E. L., Cook, G. M., and Jackson, C. J. (2020) Predicting
15. nitroimidazole antibiotic resistance mutations in *Mycobacterium tuberculosis* with protein
16. engineering. *PLoS Pathog* **16**, e1008287
17. 6. Purwantini, E., Daniels, L., and Mukhopadhyay, B. (2016) F₄₂₀H₂ is required for phthiocerol
18. dimycocerosate synthesis in *Mycobacteria*. *J Bacteriol* **198**, 2020-2028
19. 7. Purwantini, E., and Mukhopadhyay, B. (2013) Rv0132c of *Mycobacterium tuberculosis*
20. encodes a coenzyme F₄₂₀-dependent hydroxymycolic acid dehydrogenase. *PloS One* **8**,
21. e81985
22. 8. Singh, R., Manjunatha, U., Boshoff, H. I., Ha, Y. H., Niyomrattanakit, P., Ledwidge, R., Dowd,
23. C. S., Lee, I. Y., Kim, P., Zhang, L., Kang, S., Keller, T. H., Jiricek, J., and Barry, C. E., 3rd. (2008)
24. PA-824 kills nonreplicating *Mycobacterium tuberculosis* by intracellular NO release. *Science*
25. **322**, 1392-1395
26. 9. Manjunatha, U. H., Boshoff, H., Dowd, C. S., Zhang, L., Albert, T. J., Norton, J. E., Daniels, L.,
27. Dick, T., Pang, S. S., and Barry, C. E., 3rd. (2006) Identification of a nitroimidazo-oxazine-
28. specific protein involved in PA-824 resistance in *Mycobacterium tuberculosis*. *Proc Natl Acad
29. Sci U S A* **103**, 431-436
30. 10. Ichikawa, H., Bashiri, G., and Kelly, W. L. (2018) Biosynthesis of the thiopeptins and
31. identification of an F₄₂₀H₂-dependent dehydropiperidine reductase. *J Am Chem Soc* **140**,
32. 10749-10756
33. 11. Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J. L., Wang, Y. M., Xu, L. J., Chen, X. F.,
34. Huang, S. X., Lin, S. J., Deng, Z. X., and Tao, M. F. (2020) Functional genome mining reveals a
35. class V lanthipeptide containing a D-amino acid introduced by an F₄₂₀H₂-dependent
36. reductase. *Angew Chem, Int Ed* **59**, 18029-18035
37. 12. McCormick, J. R. D., and Morton, G. O. (1982) Identity of Cosynthetic Factor 1 of
38. *Streptomyces aureofaciens* and Fragment Fo from Co-Enzyme F420 of *Methanobacterium*
39. Species. *J Am Chem Soc* **104**, 4014-4015
40. 13. Wang, P., Bashiri, G., Gao, X., Sawaya, M. R., and Tang, Y. (2013) Uncovering the enzymes
41. that catalyze the final steps in oxytetracycline biosynthesis. *J Am Chem Soc* **135**, 7138-7141
42. 14. Greening, C., Jirapanjawat, T., Afroze, S., Ney, B., Scott, C., Pandey, G., Lee, B. M., Russell, R.
43. J., Jackson, C. J., Oakeshott, J. G., Taylor, M. C., and Warden, A. C. (2017) Mycobacterial
44. F₄₂₀H₂-dependent reductases promiscuously reduce diverse compounds through a common
45. mechanism. *Front Microbiol* **8**, 1000
46. 15. Mathew, S., Trajkovic, M., Kumar, H., Nguyen, Q. T., and Fraaije, M. W. (2018) Enantio- and
47. regioselective ene-reductions using F₄₂₀H₂-dependent enzymes. *Chem Commun* **54**, 11208-
48. 11211
49. 16. Drenth, J., Trajkovic, M., and Fraaije, M. W. (2019) Chemoenzymatic synthesis of an
50. unnatural deazaflavin cofactor that can fuel F₄₂₀-dependent enzymes. *ACS Catal* **9**, 6435-6443

1 17. Shah, M. V., Antoney, J., Kang, S. W., Warden, A. C., Hartley, C. J., Nazem-Bokaee, H.,
2 Jackson, C. J., and Scott, C. (2019) Cofactor F₄₂₀-dependent enzymes: an under-explored
3 resource for asymmetric redox biocatalysis. *Catalysts* **9**, 868

4 18. Martin, C., Tjallinks, G., Trajkovic, M., and Fraaije, M. W. (2021) Facile stereoselective
5 reduction of prochiral ketones by using an F₄₂₀-dependent alcohol dehydrogenase.
6 *ChemBioChem* **22**, 156-159

7 19. Selengut, J. D., and Haft, D. H. (2010) Unexpected abundance of coenzyme F(420)-dependent
8 enzymes in *Mycobacterium tuberculosis* and other actinobacteria. *J Bacteriol* **192**, 5788-5798

9 20. Ney, B., Ahmed, F. H., Carere, C. R., Biswas, A., Warden, A. C., Morales, S. E., Pandey, G.,
10 Watt, S. J., Oakeshott, J. G., Taylor, M. C., Stott, M. B., Jackson, C. J., and Greening, C. (2017)
11 The methanogenic redox cofactor F₄₂₀ is widely synthesized by aerobic soil bacteria. *ISME J*
12 **11**, 125-137

13 21. Lackner, G., Peters, E. E., Helfrich, E. J., and Piel, J. (2017) Insights into the lifestyle of
14 uncultured bacterial natural product factories associated with marine sponges. *Proc Natl
15 Acad Sci U S A* **114**, E347-E356

16 22. Braga, D., Last, D., Hasan, M., Guo, H., Leichnitz, D., Uzum, Z., Richter, I., Schalk, F.,
17 Beemelmanns, C., Hertweck, C., and Lackner, G. (2019) Metabolic pathway rerouting in
18 *Paraburkholderia rhizoxinica* evolved long-overlooked derivatives of coenzyme F₄₂₀. *ACS
19 Chem Biol* **14**, 2088-2094

20 23. Moebius, N., Uzum, Z., Dijksterhuis, J., Lackner, G., and Hertweck, C. (2014) Active invasion of
21 bacteria into living fungal cells. *eLIFE* **3**, e03007

22 24. Partida-Martinez, L. P., and Hertweck, C. (2005) Pathogenic fungus harbours endosymbiotic
23 bacteria for toxin production. *Nature* **437**, 884-888

24 25. Partida-Martinez, L. P., Monajembashi, S., Greulich, K. O., and Hertweck, C. (2007)
25 Endosymbiont-dependent host reproduction maintains bacterial-fungal mutualism. *Curr Biol*
26 **17**, 773-777

27 26. Lackner, G., Moebius, N., Partida-Martinez, L. P., Boland, S., and Hertweck, C. (2011)
28 Evolution of an endofungal lifestyle: deductions from the *Burkholderia rhizoxinica* genome.
29 *BMC Genomics* **12**, 210

30 27. Bashiri, G., and Baker, E. N. (2020) Convergent pathways to biosynthesis of the versatile
31 cofactor F₄₂₀. *Curr Opin Struct Biol* **65**, 9-16

32 28. Graupner, M., and White, R. H. (2001) Biosynthesis of the phosphodiester bond in coenzyme
33 F(420) in the methanarchaea. *Biochemistry* **40**, 10859-10872

34 29. Grochowski, L. L., Xu, H., and White, R. H. (2008) Identification and characterization of the 2-
35 phospho-L-lactate guanylyltransferase involved in coenzyme F₄₂₀ biosynthesis. *Biochemistry*
36 **47**, 3033-3037

37 30. Li, H., Graupner, M., Xu, H., and White, R. H. (2003) CofE catalyzes the addition of two
38 glutamates to F₄₂₀-O in F₄₂₀ coenzyme biosynthesis in *Methanococcus jannaschii*. *Biochemistry*
39 **42**, 9771-9778

40 31. Bashiri, G., Rehan, A. M., Sreebhavan, S., Baker, H. M., Baker, E. N., and Squire, C. J. (2016)
41 Elongation of the poly-gamma-glutamate tail of F₄₂₀ requires both domains of the
42 F₄₂₀:gamma-glutamyl ligase (FbiB) of *Mycobacterium tuberculosis*. *J Biol Chem* **291**, 6882-
43 6894

44 32. Bashiri, G., Antoney, J., Jirgis, E. N. M., Shah, M. V., Ney, B., Copp, J., Stuteley, S. M.,
45 Sreebhavan, S., Palmer, B., Middleditch, M., Tokuriki, N., Greening, C., Scott, C., Baker, E. N.,
46 and Jackson, C. J. (2019) A revised biosynthetic pathway for the cofactor F₄₂₀ in prokaryotes.
47 *Nat Commun* **10**, 1558

48 33. Braga, D., Hasan, M., Kröber, T., Last, D., and Lackner, G. (2020) Redox coenzyme F₄₂₀
49 biosynthesis in Thermomicrobia involves reduction by stand-alone nitroreductase
50 superfamily enzymes. *Appl Environ Microbiol* **86**

1 34. Grinter, R., Ney, B., Brammananth, R., Barlow, C. K., Cordero, P. R. F., Gillett, D. L., Izore, T.,
2 Cryle, M. J., Harold, L. K., Cook, G. M., Taiaroa, G., Williamson, D. A., Warden, A. C.,
3 Oakeshott, J. G., Taylor, M. C., Crellin, P. K., Jackson, C. J., Schittenhelm, R. B., Coppel, R. L.,
4 and Greening, C. (2020) Cellular and structural basis of synthesis of the unique intermediate
5 dehydro-F420-0 in mycobacteria. *mSystems* **5**
6 35. Graupner, M., Xu, H., and White, R. H. (2002) Characterization of the 2-phospho-L-lactate
7 transferase enzyme involved in coenzyme F(420) biosynthesis in *Methanococcus jannaschii*.
8 *Biochemistry* **41**, 3754-3761
9 36. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S.,
10 Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A.
11 (2012) Geneious Basic: an integrated and extendable desktop software platform for the
12 organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649
13 37. Li, C., Wen, A., Shen, B., Lu, J., Huang, Y., and Chang, Y. (2011) FastCloning: a highly
14 simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC*
15 *Biotechnol* **11**, 92
16 38. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion,
17 single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91
18 39. Di Tommaso, P., Moretti, S., Xenarios, I., Orobio, M., Montanyola, A., Chang, J. M., Taly, J. F.,
19 and Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of
20 protein and RNA sequences using structural information and homology extension. *Nucleic*
21 *Acids Res.* **39**, W13-17
22 40. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
23 throughput. *Nucleic Acids Res* **32**, 1792-1797
24 41. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010)
25 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
26 performance of PhyML 3.0. *Syst Biol* **59**, 307-321
27 42. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2
28 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858
29 43. Potterton, L., Aguirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan,
30 R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M.,
31 Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., von Delft, F.,
32 Waterman, D., Wilson, K., Winn, M., and Wojdyr, M. (2018) CCP4i2: the new graphical user
33 interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* **74**, 68-84
34 44. Krissinel, E. (2012) Enhanced fold recognition using efficient short fragment clustering. *J Mol*
35 *Biochem* **1**, 76-85
36 45. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R.
37 M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S.,
38 Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the
39 CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
40 46. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of
41 docking with a new scoring function, efficient optimization, and multithreading. *J Comput*
42 *Chem* **31**, 455-461
43 47. van Aalten, D. M., Bywater, R., Findlay, J. B., Hendlich, M., Hooft, R. W., and Vriend, G. (1996)
44 PRODRG, a program for generating molecular topologies and unique molecular descriptors
45 from coordinates of small molecules. *J Comput Aided Mol Des* **10**, 255-262
46 48. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of
47 Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
48 49. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of
49 nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**,
50 10037-10041

1 50. Krug, M., Weiss, M. S., Heinemann, U., and Mueller, U. (2012) XDSAPP: a graphical user
2 interface for the convenient processing of diffraction data using XDS. *J Appl Crystallogr* **45**,
3 568-572

4 51. Cowtan, K. (2012) Completion of autobuilt protein models using a database of protein
5 fragments. *Acta Crystallogr., Sect. D: Struct. Biol.* **68**, 328-335

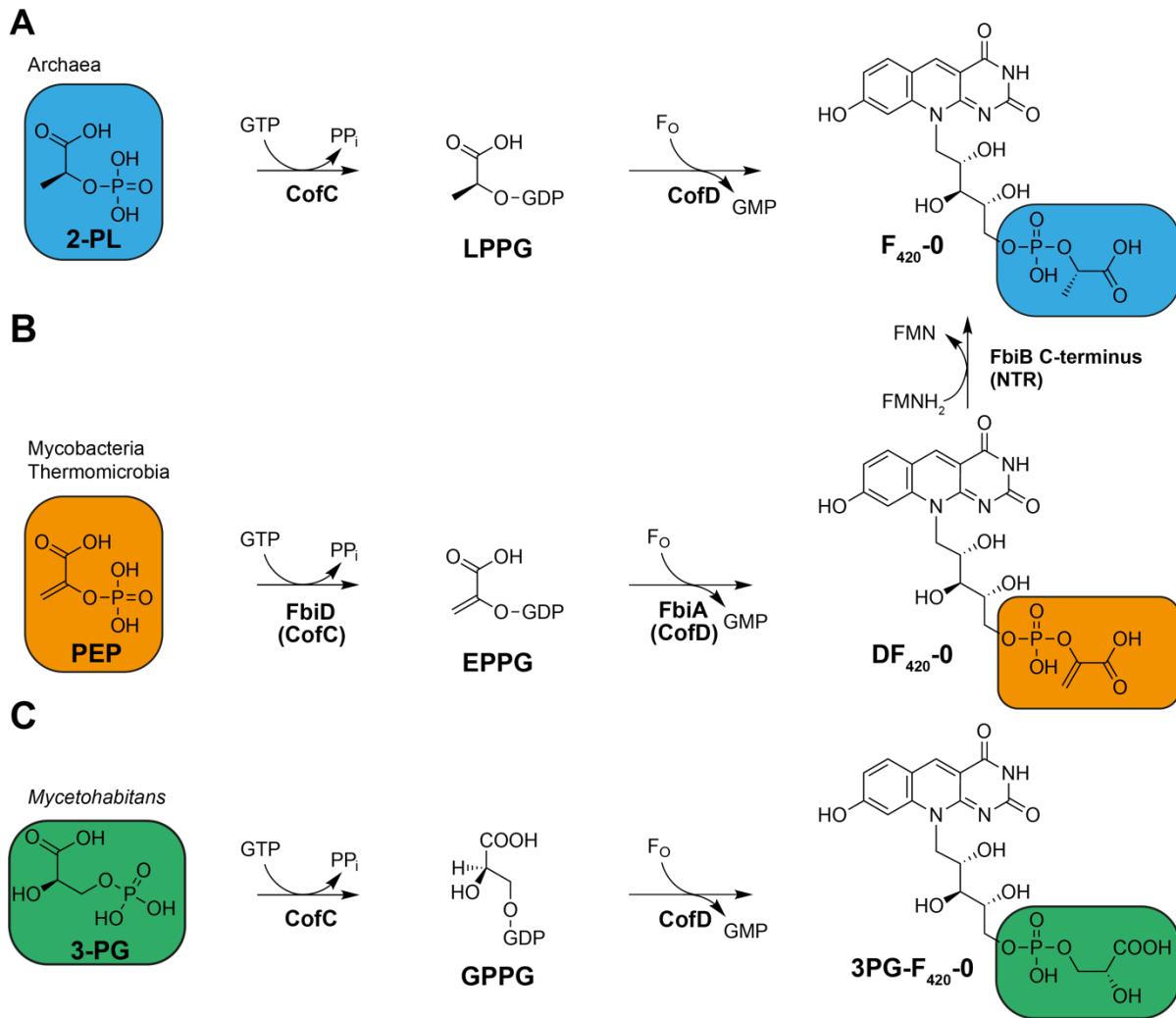
6 52. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A.,
7 Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of
8 macromolecular crystal structures. *Acta Crystallogr D* **67**, 355-367

9 53. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of
10 Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501

11

12

1 Figures



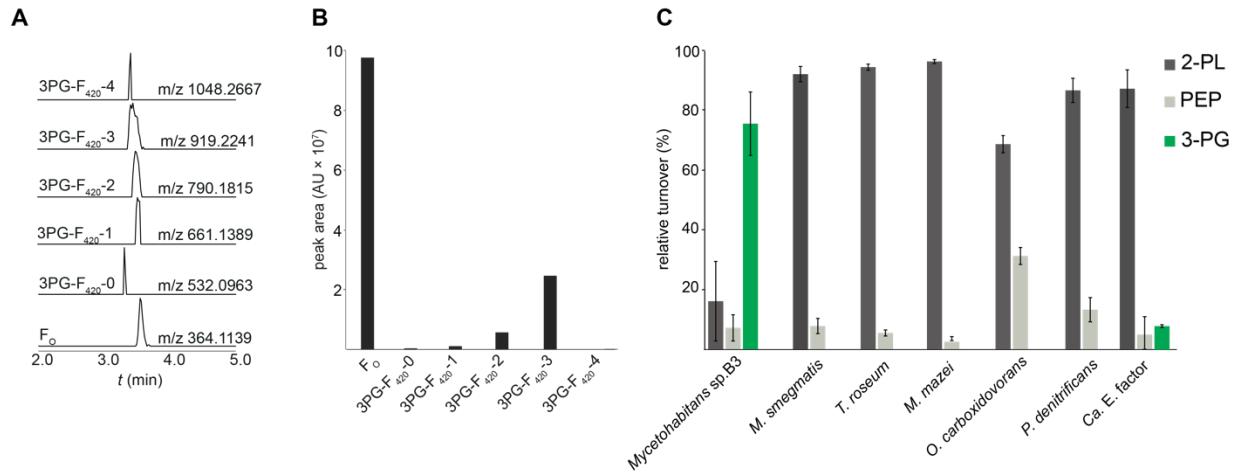
2

3 **Figure 1: Biosynthetic steps performed by CofC and CofD during the formation of**
 4 **F₄₂₀-species. (A) Formation of F₄₂₀-0 from 2-PL as proposed for archaea. (B) Formation**
 5 **of the F₄₂₀ precursor DF₄₂₀, a pathway intermediate of F₄₂₀ found in mycobacteria and**
 6 **Thermomicrobia. DF₄₂₀ is further reduced to F₄₂₀ by a nitroreductase (NTR)-like enzyme**
 7 **(C) Biosynthesis of 3PG-F₄₂₀-0 in *M. rhizoxinica* and related endofungal bacteria. 2-PL: 2-**
 8 **phospho-L-lactate, 3-PG: 3- phospho-D-glycerate, EPPG: enolpyruvyl-2-diphospho-5'-**
 9 **guanosine, FMN: flavin mononucleotide, F_o: 7,8-didemethyl-8-hydroxy-5-**
 10 **deazariboflavin, GMP: guanosine 5'-monophosphate, GPPG: 3-(guanosine-5'-**

1 diphospho)-D-glycerate, GTP: guanosine 5'-triphosphate, LPPG: L-lactyl-2-diphospho-5'-

2 guanosine, NTR: nitroreductase, PEP: phosphoenolpyruvate, PP_i: pyrophosphate.

3



4

5 **Figure 2: Production of 3PG-F₄₂₀ and substrate specificity of CofC/FbiA enzymes A)**

6 Extracted ion chromatograms (XICs) of deazaflavin species extracted from
7 *Mycetohabitans* sp. B3 (peaks are scaled to the same height). **(B)** Areas under the peaks
8 (arbitrary units) depicted in panel A. **(C)** Substrate specificity assay of CofC/FbiA from
9 various source organisms. The CofC of *Mycetohabitans* sp. B3 showed strong 3-PG
10 activation while all other homologs preferred 2-PL. *Mjan*-CofD was combined with all
11 CofC homologs to perform the CofC/D assay. Error bars represent the standard
12 deviation of three biological replicates. Relative turnover of 3-PG, 2-PL, and PEP are
13 reflected by the rate of 3PG-F₄₂₀-0, F₄₂₀-0, and DF₄₂₀-0 formation, respectively.

14

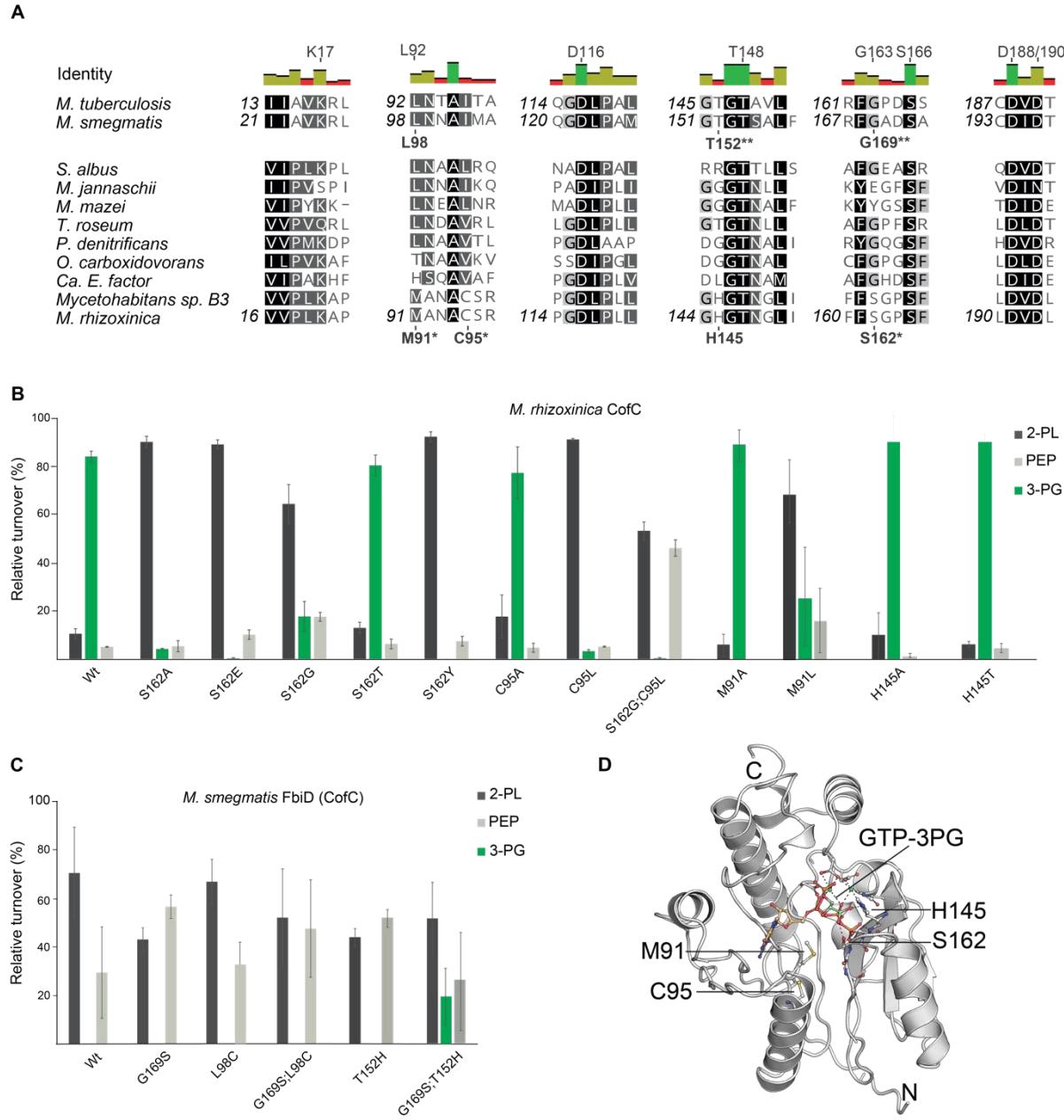


Figure 3: Residues determining substrate specificities of CofC homologs. (A)
 Multiple sequence alignment of CofC proteins from selected source organisms. Amino acids of *Mtb*-FbiD suggested previously to be involved in PEP binding residues (32) are indicated above the identity graph. Residues tested by mutagenesis are shown below sequences. Asterisks: crucial for 3-PG activation in *Mrhiz*-CofC, double asterisk: enabled 3-PG activation by *Msmeg*-FbiD. **(B)** Substrate specificity of *Mrhiz*-CofC after site-directed mutagenesis. Substitution of S162, C95, and M91 by residues occurring in 2-

1 PL/PEP activating enzymes led to reduction or abolishment of 3-PG activation. **(C)**
2 Substrate specificity of *Msmeg*-FbiD after site-directed mutagenesis. Gly169, Leu98, and
3 Thr152 were exchanged by amino acids found in 3-PG activating homologs. While single
4 mutations did not result in 3-PG activation, the combination G169S;T152H enabled 3-PG
5 turnover. Error bars represent the standard deviation of three biological replicates. **(D)**
6 Homology model of *Mrhiz*-CofC in complex with GTP (placed by molecular docking) and
7 3-PG (placed manually).

8

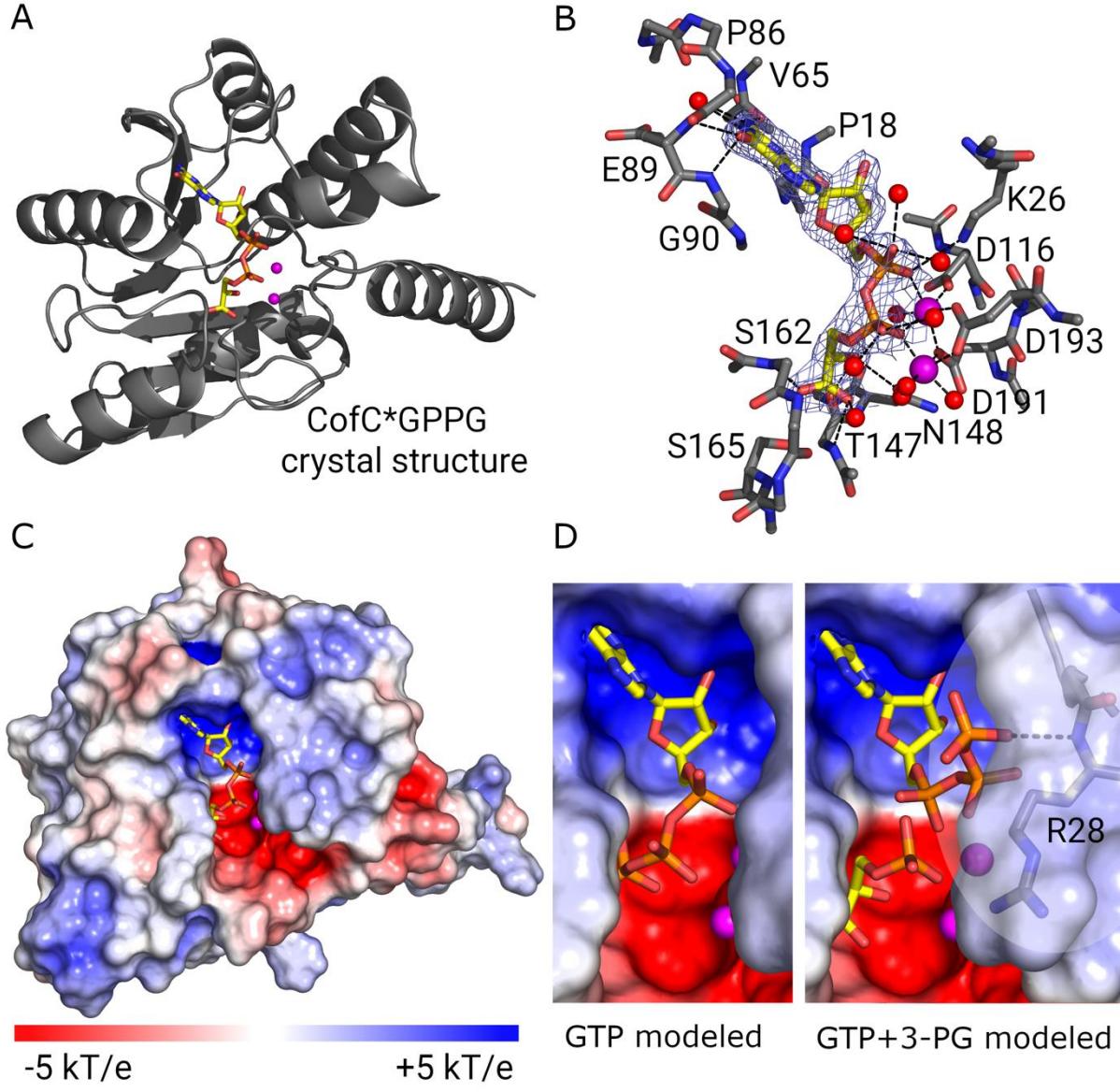
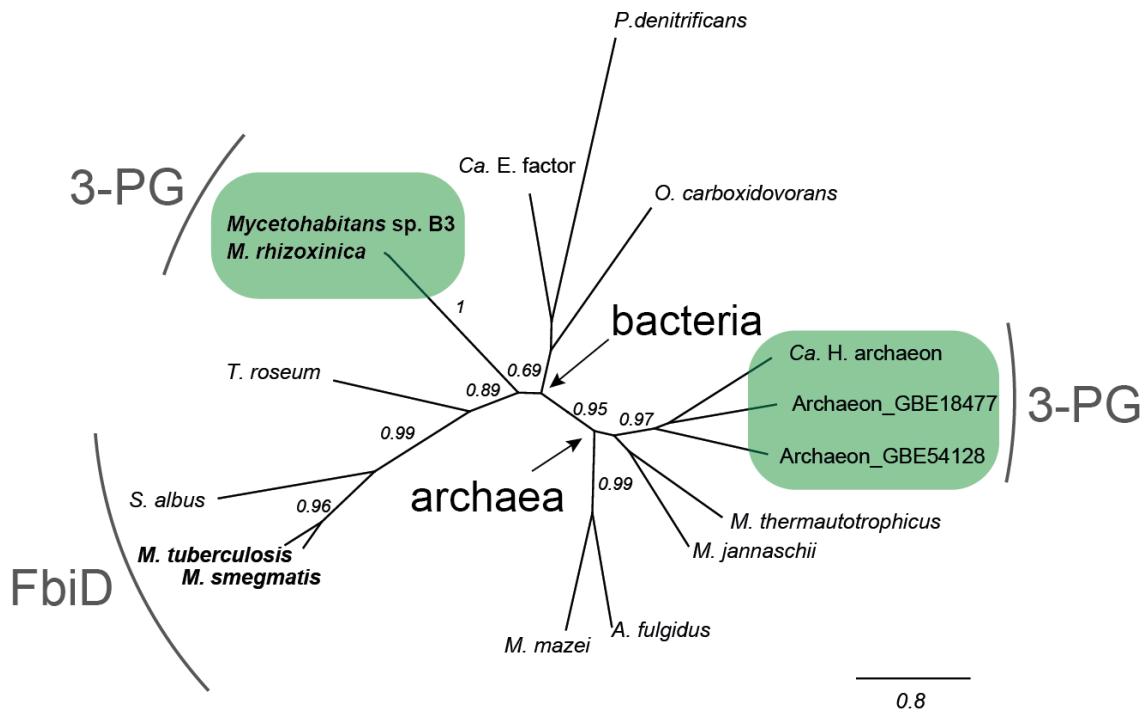


Figure 4: Structures of MycB3-CofC in complex with GPPG (crystal structure) and educts (model). **(A)** Ribbon diagram of the crystal structure of CofC with its product GPPG. **(B)** Final electron density of GPPG ($2F_0F_c$ at 1.5σ) and H-bonds for GPPG binding. Side chains are only shown for residues contacting the ligand. **(C)** Electrostatic surface representation of the protein shows a negatively charged deep pocket for GPPG. Two Mg^{2+} -ions allow binding of the charged phosphates of GPPG. The electrostatic surface potential was calculated using APBS. Red: negative charge; blue: positive charge. **(D)**

1 Possible educt conformations. GTP could position its α - and β -phosphates where GPPG
2 binds to the Mg^{2+} -ions (left side). With 3-PG binding in the same mode as the 3-PG
3 moiety of GPPG the β - and γ -phosphates of GTP have to rotate out of the binding pocket,
4 though (right side). The phosphate moiety of 3-PG is well poised for nucleophilic attack
5 on the α -phosphate of GTP via a trigonal bipyramidal transition state. Figures were
6 produced with PyMOL. A window around R28 is transparent to show the potential H-
7 bond of GTP to the amide nitrogen. Electrostatic surface representations were made
8 using the APBS electrostatics plugin in PyMOL. Carbon: yellow, oxygen: red, nitrogen:
9 blue, magnesium: magenta.



10

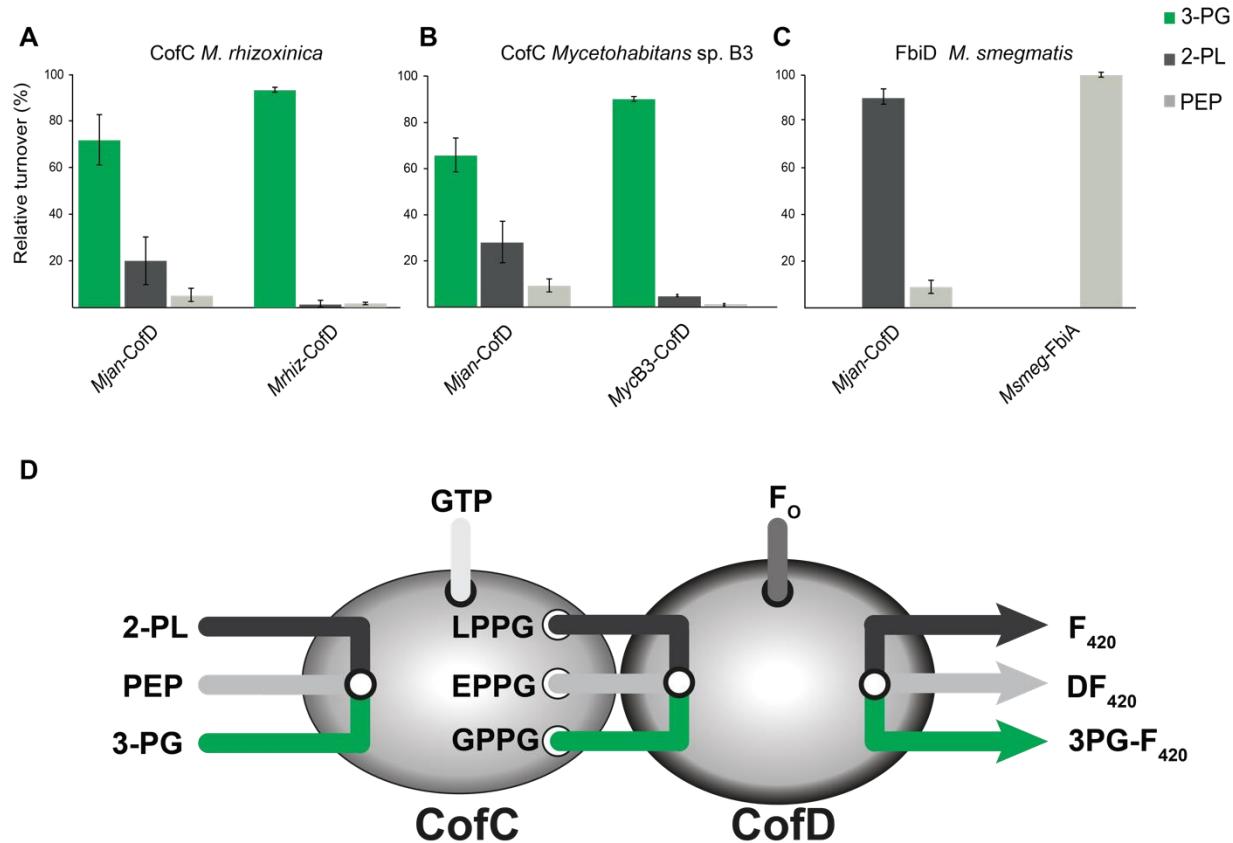
11 **Figure 5. Phylogenetic tree of CofC/FbiD enzymes.** The maximum likelihood method
12 implemented in PhyML 3.0 was used to reconstruct the phylogeny. 3-PG activating
13 enzymes (highlighted in green boxes) have evolved two time independently. Branch
14 labels indicate SH-like support values. The scale bar represents substitutions per site.

1 Accession numbers and source organisms of primary amino acid sequences are available

2 in **Table S3**.

3

1



2

3 **Figure 6: Influence of CofD/FbiA on the biosynthesis of F_{420} -0 derivatives. (A) CofC**
4 from *M. rhizoxinica* **(B) CofC from *Mycetohabitans* sp. B3, and (C) FbiD from *M.***
5 *smegmatis* were tested with either *M. jannaschii* CofD (standard assay) or together with
6 their cognate CofD enzyme from the same source organism **(D) Schematic drawing**
7 illustrating the channeling of the labile metabolites LPPG, EPPG, GPPG from CofC to CofD.
8 While CofC selects primary substrates, CofD is able to select between these
9 intermediates. For abbreviations see Figure 1.

10

11

12

13

1

2