

ARTICLE

Received 16 Oct 2014 | Accepted 11 Mar 2015 | Published 17 Apr 2015

DOI: 10.1038/ncomms7895

# Protein-pyridinol thioester precursor for biosynthesis of the organometallic acyl-iron ligand in [Fe]-hydrogenase cofactor

Takashi Fujishiro<sup>1</sup>, Jörg Kahnt<sup>1</sup>, Ulrich Ermler<sup>2</sup> & Seigo Shima<sup>1,3</sup>

The iron-guanylylpyridinol (FeGP) cofactor of [Fe]-hydrogenase contains a prominent iron centre with an acyl-Fe bond and is the only acyl-organometallic iron compound found in nature. Here, we identify the functions of HcgE and HcgF, involved in the biosynthesis of the FeGP cofactor using structure-to-function strategy. Analysis of the HcgE and HcgF crystal structures with and without bound substrates suggest that HcgE catalyses the adenylation of the carboxy group of guanylylpyridinol (GP) to afford AMP-GP, and subsequently HcgF catalyses the transesterification of AMP-GP to afford a Cys (HcgF)-S-GP thioester. Both enzymatic reactions are confirmed by *in vitro* assays. The structural data also offer plausible catalytic mechanisms. This strategy of thioester activation corresponds to that used for ubiquitin activation, a key event in the regulation of multiple cellular processes. It further implicates a nucleophilic attack onto the acyl carbon presumably via an electron-rich Fe(O)- or Fe(I)-carbonyl complex in the Fe-acyl formation.

<sup>1</sup>Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany. <sup>2</sup>Max Planck Institute for Biophysics, Max-von-Laue-Strasse 3, 60438 Frankfurt am Main, Germany. <sup>3</sup>PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho Kawaguchi, 332-0012 Saitama, Japan. Correspondence and requests for materials should be addressed to S.S. (email: shima@mpi-marburg.mpg.de).

Nature expands the catalytic capability of proteins by developing metal cofactors<sup>1</sup>. Some of the most attractive representatives are those of [NiFe]-, [FeFe]- and [Fe]-hydrogenases, which catalyse the production and consumption of H<sub>2</sub> in microorganisms<sup>2–5</sup>. The unusual Fe centres of [NiFe]- and [FeFe]-hydrogenases contain thiolate, CO and CN<sup>−</sup> ligands. Instead of CN<sup>−</sup> ligands, the Fe centre of the iron-guanylylpyridinol (FeGP) cofactor of [Fe]-hydrogenase has a guanylylpyridinol (GP), which is bidentately chelated to Fe by its nitrogen and acyl-carbon<sup>6,7</sup>. The Fe-acyl bond constitutes the only stable acyl-organometallic compound found in nature, although an acyl-Ni species has been reported to be an intermediate in the acetyl-CoA formation reaction catalysed by acetyl-CoA synthase<sup>8,9</sup>. The metal cofactors of hydrogenases and their biosynthesis have been extensively explored to unravel their fascinating chemistry<sup>10–12</sup> and to use their H<sub>2</sub> activation capabilities for biotechnological applications<sup>13,14</sup>. In contrast to studies of [NiFe]- and [FeFe]-hydrogenases, convincing hypotheses and experimental data have not been reported for the biosynthesis of the Fe centre of [Fe]-hydrogenase, including the formation of the unique acyl-group, the greatest challenge in FeGP cofactor biosynthesis.

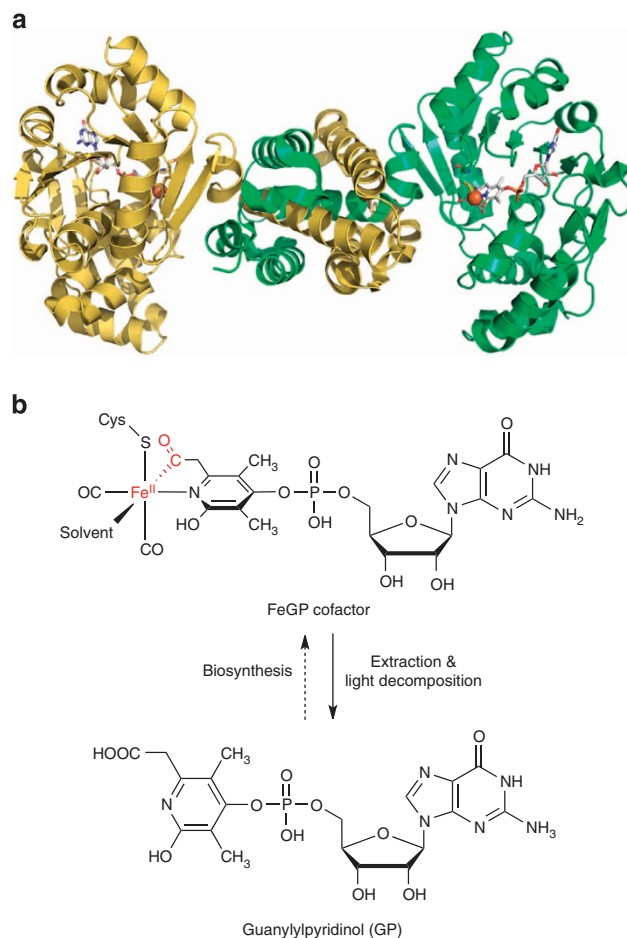
The seven conserved *hmd* co-occurring (*hcgA-G*) genes in the genomes of hydrogenotrophic methanogens are involved in the biosynthesis of the FeGP cofactor<sup>15,16</sup>. Recent findings using structure-to-function strategy<sup>17,18</sup> indicated that the formation of 6-carboxymethyl-guanylylpyridinol from GTP and a pyridinol compound are catalysed by HcgB, which strongly suggested that GP is an intermediate in the FeGP biosynthetic pathway and, consequently, that its carboxy group is converted to the acyl-iron ligand in subsequent enzymatic reactions<sup>19</sup>.

In the current study, we find that HcgE and HcgF catalyse the formation of a thioester-activated acyl-group using a two-state mechanism reminiscent of that of ubiquitin activation<sup>20</sup>. HcgE catalyses the adenylation of the carboxy group of GP to afford adenosine 5'-monophosphate (AMP)-GP. HcgF catalyses the transesterification of AMP-GP to afford a Cys-S-GP thioester. On the basis of the findings, possible Fe-acyl formation mechanisms are proposed.

## Results

**Structural identification of substrates of HcgE.** The functional characterization of HcgE was inspired by the relationship between the primary structures of HcgE and E1-like ubiquitin-activating enzymes (E1 enzymes) (Supplementary Figs 1 and 2)<sup>20</sup>. E1 enzymes activate the C-terminal carboxy group of ubiquitin or ubiquitin-like proteins and of a precursor heptapeptide of an antibiotic by adenylation with adenosine 5'-triphosphate (ATP). This essential activation reaction is performed by various eukaryotic E1 enzymes/ubiquitin-like proteins<sup>20,21</sup>, and by the prokaryotic protein pairs MoeB/MoaD<sup>22</sup>, ThiF/ThiS<sup>23</sup> and MccB/MccA<sup>24</sup>, which are involved in the modulation of protein function in the regulation of multiple cellular processes and in the biosynthesis of molybdopterin, thiamine and the antibiotic microcin C7, respectively. Likewise, HcgE might therefore catalyse the adenylation of a carboxy group of an unknown compound, that is, GP, in FeGP cofactor biosynthesis. Structures of [Fe]-hydrogenase, the FeGP cofactor and GP are shown in Fig. 1.

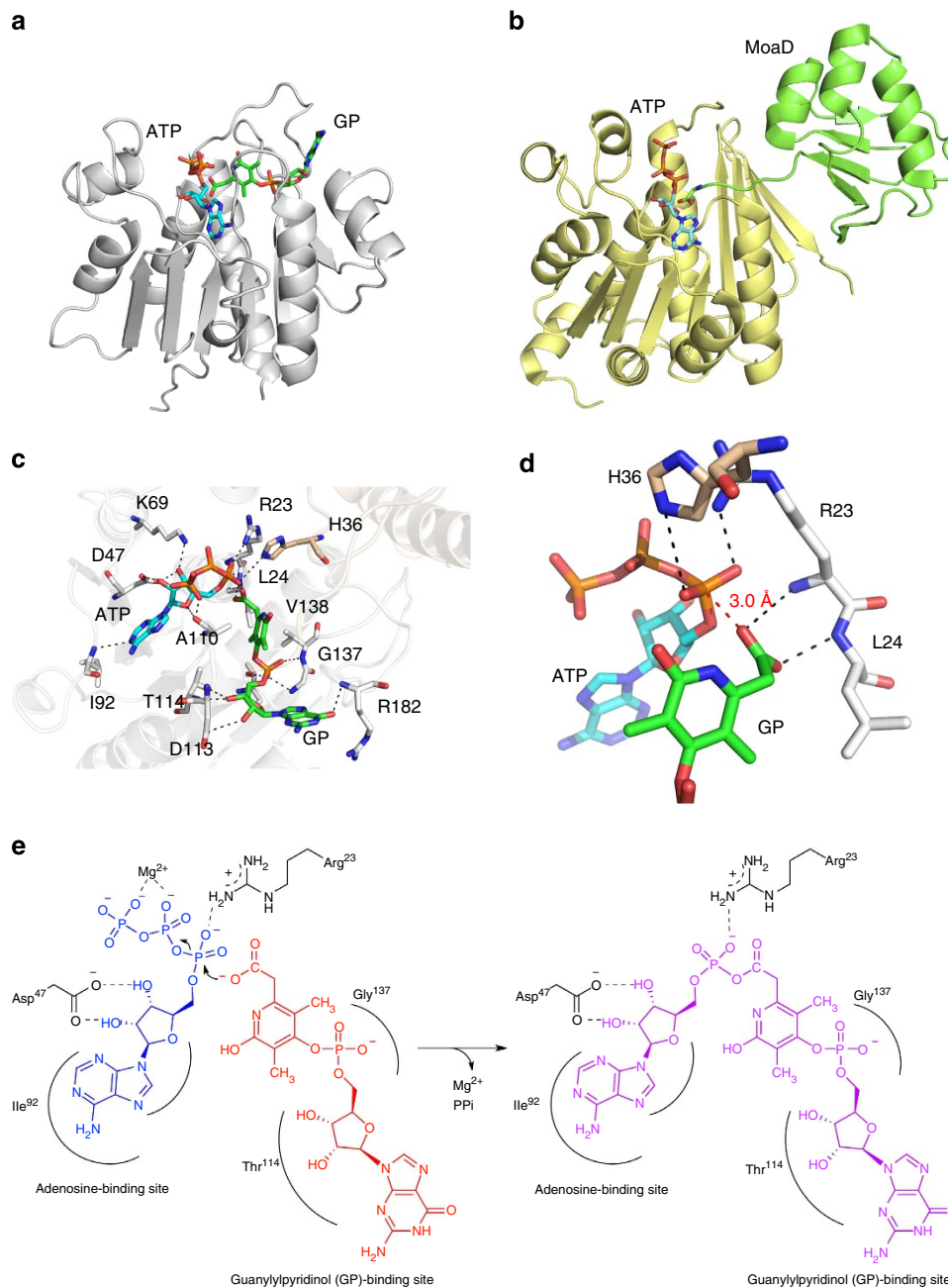
To experimentally explore this hypothesis, we determined the crystal structure of homotrimeric HcgE and of the HcgE-ATP complexes from *Methanothermobacter marburgensis* (Fig. 2a; Supplementary Figs 3 and 4) at 1.6 and 1.8 Å resolution, respectively. ATP binds to the canonical ATP-binding site of



**Figure 1 | [Fe]-hydrogenase and FeGP cofactor.** (a) Crystal structure of dimeric [Fe]-hydrogenase with bound FeGP cofactor shown as a ball-and-stick model. This enzyme is involved in the methanogenic energy metabolism of many hydrogenotrophic methanogens. (b) Biosynthesis and light decomposition of the FeGP cofactor. The FeGP cofactor can be extracted with methanol in the presence of 2-mercaptoethanol and decomposed by light exposure to Fe<sup>2+</sup>, CO and GP, in which the acyl-group is hydrolysed to a carboxy group<sup>34</sup>. GP appears to be a natural intermediate in the FeGP cofactor biosynthesis pathway of the cells<sup>19</sup>. An acyl-iron bond (red) is only found in the FeGP cofactor.

E1 enzymes located in the conserved N-terminal half. In contrast, the fold of the C-terminal half substantially differs between HcgE and the structurally known E1 enzymes, resulting in a completely altered substrate-binding site (Fig. 2b; Supplementary Fig. 3). We identified GP as a possible substrate of HcgE on the basis of the ternary HcgE-ATP-GP complex structure determined at 2.8 Å resolution (Fig. 2c,d). Notably, GP binding was only achievable when HcgE and GP were crystallized in the presence of ATP. Despite different substrate-binding sites, the carboxy group of GP in HcgE and the carboxy group of the C-terminal glycine of MoeB/MoaD are in equivalent position.

**In vitro activity assay of HcgE.** To complement the structure-guided functional prediction, we analysed a reaction mixture containing HcgE, ATP, GP and Mg<sup>2+</sup> by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MS) (Supplementary Fig. 5). The obtained peak at *m/z* 872.2 corresponded to the expected product AMP-GP (calculated mass = 871.2) and its fragmentation peaks at *m/z* 348.1 and 525.1 corresponded to AMP and GP, respectively. HcgE activity

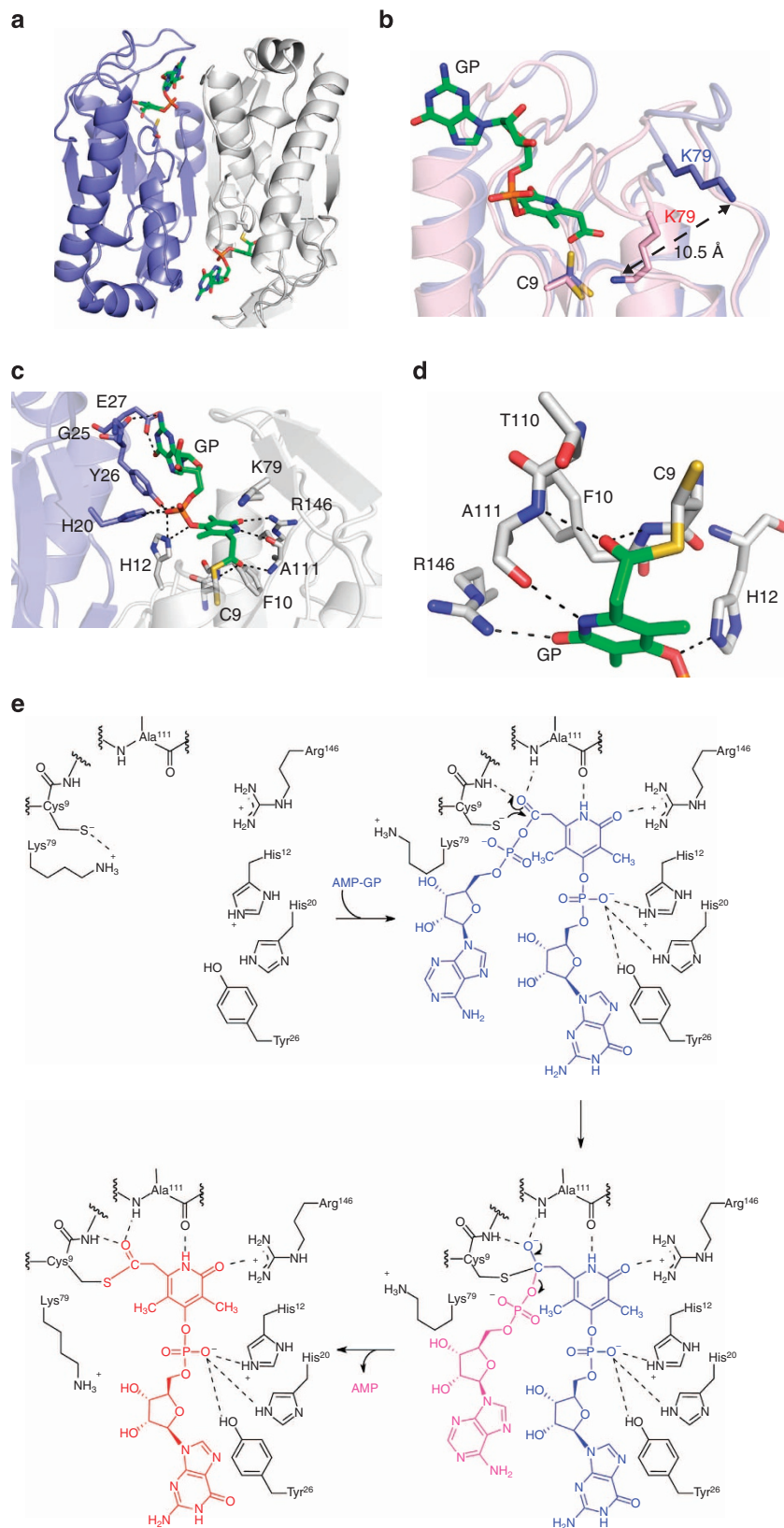


**Figure 2 | Crystal structure and mechanism of HcgE.** (a) Structure of HcgE (grey) in complex with ATP (carbon in cyan) and GP (carbon in green). (b) Structure of MoeB (yellow) with ATP (carbon in cyan) and MoaD (green)<sup>22</sup>. The ATP-binding site of HcgE and MoeB are conserved, whereas their GP- and MoaD-binding sites but not the position of their corresponding carboxy group markedly vary. (c) The ATP- and GP-binding sites. ATP is present in a hairpin-like and GP in a slightly bent conformation. Polar interactions are indicated as black dashed lines. (d) Transition state-like arrangement of GP and ATP. The carboxy group of GP points to the  $\alpha$ -phosphate of ATP. The distance between the oxygen atom of the carboxy group of GP and the phosphorus atom of the  $\alpha$ -phosphate group of ATP is shown as a red dashed line. (e) Mechanism proposed for the HcgE-catalysed adenylylation of GP with ATP.

followed Michaelis–Menten kinetics with an apparent  $K_m$  for ATP of  $27.0 \pm 3.5 \mu\text{M}$  and a  $k_{\text{cat}}$  of  $7.2 \pm 0.2 \text{ min}^{-1}$  based on determination of the pyrophosphate concentration (Supplementary Fig. 6). Thus, structural, mass spectrometry and kinetic data identified HcgE as an adenylyltransferase that forms AMP-GP from ATP and GP.

**Catalytic mechanism of HcgE.** The ternary structure of the HcgE–ATP–GP complex revealed that the carboxy group of GP points towards the  $\alpha$ -phosphate group of ATP (Fig. 2c,d). This

transition state-like arrangement not only strongly argues for an adenylylation reaction catalysed by HcgE but also provides clues to its catalytic mechanism (Fig. 2e). In this proposed mechanism, the carboxy group of GP attacks the  $\alpha$ -phosphate of ATP via an  $S_N2$  reaction thereby adenylylating GP via a pentavalent intermediate. The negative charge of the carboxy oxygen of GP is compensated by a hydrogen bond to Leu24-NH located at the N-terminal side of helix 24:34. The guanidinium group of Arg23 and the imidazole group of His36 of the partner subunit of the trimer compensate the negative charge of the  $\alpha$ -phosphate oxygen of ATP (Fig. 2c,d). The negative charge of the released



**Figure 3 | Crystal structure and catalytic mechanism of HcgF.** (a) Structure of HcgF in complex with GP (green). In the two active sites of the homodimer, the Cys9 sulfur and the GP acyl carbon either contact each other or partially form a covalent bond. (b) Superimposed structures of GP-free (light magenta) and non-covalently GP-bound HcgF (blue). Cys9 is presumably kept in the more reactive thiolate form by Lys79 in the resting state and attacks activated GP (green) after its binding. (c) The GP-binding site with GP covalently bound to Cys9. (d) The thioester bond between GP and Cys9. The extra S (yellow) indicates the second conformation of the Cys9 side chain. (e) Proposed catalytic mechanism of the thioester formation of AMP-GP with Cys9 of HcgF.

pyrophosphate is delocalized by interactions with  $Mg^{2+}$  (not visible in the electron density map), which probably interacts with its  $\beta$ - and  $\gamma$ -phosphates and with several polypeptide side chains (Fig. 2e).

**Possible function of HcgF provided by protein structure.** We determined the function of HcgF using the structure-to-function strategy used for HcgE. HcgF does not show sequence similarities to any enzyme of known function (Supplementary Figs 7 and 8), but the crystal structure determined for homodimeric HcgF from *Methanocaldococcus jannaschii* is strongly related to that of the nicotinamide mononucleotide (NMN) deaminase PncC<sup>25</sup> (Supplementary Fig. 9). HcgF does not exhibit NMN hydrolase activity, but the profiles of the proposed NMN-binding cavity of PncC and the equivalent site of HcgF are compatible with the demands for accommodating an NMN-like compound, that is, GP (Supplementary Fig. 9). Indeed, the X-ray crystal structure of HcgF at 2.0 Å resolution co-crystallized with GP revealed bound GP (Fig. 3a–d). Interestingly, GP in one HcgF monomer of the asymmetric unit partially forms a covalent thioester bond between its carboxy carbon and the thiolate of Cys9 (Fig. 3c,d), while in the other monomer, Cys9 and GP are in van der Waals contact (Fig. 3b; Supplementary Fig. 10). In the substrate-free HcgF, the strictly conserved Cys9 is activated by the neighbouring Lys79-N. Upon binding of GP, conformational changes of the segments after strands 46:53 and 125:131 up to 6 Å enlarge the substrate-binding site and, concomitantly, induce a movement of the side chain of Lys79 by more than 10 Å from the largely buried Cys9 towards the protein surface (Fig. 3b,c). The key residue Cys9 positioned at the C-terminal end of the shortened strand 4:9 is strictly conserved in HcgF.

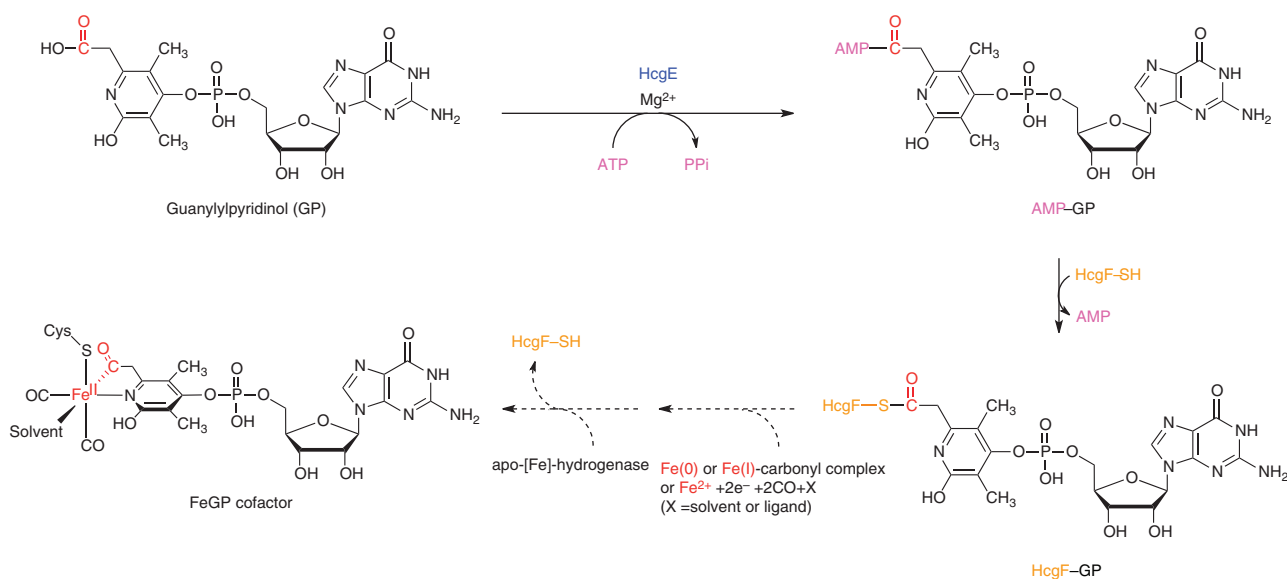
**AMP-GP is the substrate of HcgF.** Consideration of the obtained structural data and the established reaction of HcgE led to the proposal that the same two-step carboxylate activation strategy reported for ubiquitin activation is used in the biosynthesis of the FeGP cofactor, namely adenylation of the carboxy group of GP by HcgE, followed by thioester formation by HcgF (Fig. 4). This

hypothesis implies that not GP but rather AMP-GP formed by the HcgE reaction is the natural substrate of HcgF. Unfortunately, the instability of AMP-GP, precludes its use in crystallographic studies. Docking simulation of AMP on GP covalently bound to HcgF, however, indicated that it is spatially possible for AMP-GP to be accommodated in the binding site adjacent to Cys9 (Supplementary Fig. 9c). To demonstrate the transesterification reaction from AMP-GP to Cys9 (HcgF)-GP in solution, we added HcgF to the HcgE reaction assay described above. The production of AMP confirmed the proposed function of HcgF (Supplementary Fig. 11).

**Catalytic mechanism of HcgF.** A proposed reaction mechanism for the thioester bond formation between Cys9 of HcgF using AMP-GP as a substrate is shown in Fig. 3e. In this mechanism, the thiolate of Cys9 attacks the adenylated carboxy group of AMP-GP in a  $S_N2$  reaction, resulting in the formation of a thioester bond and the release of AMP. The oxygen of the thioester group interacts with Phe10-NH and Ala111-NH (see also Fig. 3d), which most likely also stabilizes the anionic tetrahedral transition state. Notably, the GP-HcgF adduct is the first example of a ubiquitin activation system for which the thioester product has been structurally characterized.

## Discussion

The identified GP carboxylate activation involved in Fe-acyl formation corresponds to the mechanism established for ubiquitin activation and thereby expands this versatile strategy from (poly)peptides to other carboxylated substrates. Activation via a thioester bond suggests a nucleophilic substitution reaction for Fe-acyl formation to allow the release of Cys9 (HcgF) as thiolate, which is a favourable leaving group (Fig. 4). An iron-carbonyl precursor, so far uncharacterized in the biosynthesis of the [Fe]-hydrogenase cofactor but experimentally identified in that of [NiFe]- and [FeFe]-hydrogenase<sup>11,26,27</sup>, can in principle, provide a nucleophilic iron as central atom. An Fe(0)-carbonyl complex as the precursor of acyl-iron ligand biosynthesis appears to be suitable and has been used for chemical synthesis of the



**Figure 4 | A proposed biosynthetic pathway to the FeGP cofactor.** The carboxy group of GP is adenylated by HcgE using ATP and  $Mg^{2+}$  to form AMP-GP; released AMP-GP is bound by HcgF and forms a covalent HcgF-GP thioester adduct with the Cys9 thiolate; and the reactive thioester between GP and HcgF is converted to an acyl-Fe bond perhaps by a nucleophilic substitution of an electron-rich Fe from an unknown iron-carbonyl precursor. The thiolate of Cys9 of HcgF is an excellent leaving group, especially when its negative charge is compensated by protonation or hydrogen-bond interactions.

Fe-acyl bond of FeGP cofactor model compounds<sup>28</sup>. However, the formation of an Fe(0)–carbonyl complex precursor requires a chemistry challenging in a water-rich biological environment, such as that of the novel iron(II)-complex reduction system in methanogenic archaea (Fig. 4). Alternatively, a one-electron reduction system might formally generate a Fe(I)–carbonyl complex for nucleophilic attack, followed by a second one-electron reduction step such as proposed for the catalytic mechanism of acetyl-coenzyme A synthase<sup>8</sup>. The enzymatic system involved in the Fe-acyl formation is unknown. HcgF could also function as a scaffold protein on which biosynthesis of the FeGP cofactor is completed after an electron-rich Fe–carbonyl complex is accepted. The intact cofactor biosynthesized on HcgF could be transferred to the [Fe]-hydrogenase apoenzyme to form its holoenzyme. However, a scaffold function of the other enzymes cannot be excluded<sup>17,29</sup>. Nevertheless, the established thioester activation offers first insights into the chemistry of the subsequent Fe-acyl bond formation and places this reaction in the pathway of Fe centre biosynthesis after iron–carbonyl formation.

## Methods

**Chemicals.** All chemicals were used without further purification. ATP, AMP, ammonium acetate, pentaerythritol ethoxylate 270 (3/4 EO/OH), (±)-2-methyl-2,4-pentandiol, nickel sulphate, potassium hydroxide, lithium sulphate, thrombin from bovine plasma and inorganic pyrophosphatase from *Saccharomyces cerevisiae* were purchased from Sigma-Aldrich. L-(+)-Selenomethionine (SeMet) was purchased from Acros Organics. Magnesium chloride, PEG6000, PEG3000 and thiamine were purchased from Merck. The following reagents were obtained from Roth: acetonitrile, kanamycin sulphate, 3-(*N*-morpholino)propanesulfonic acid (MOPS), 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 2-(*N*-morpholino)ethanesulfonic acid (MES), imidazole, ammonium sulphate, ammonium chloride, sodium acetate, sodium chloride, sodium citrate, hydrochloric acid, potassium chloride, potassium dihydrogen phosphate, dipotassium hydrogen phosphate, dithiothreitol (DTT), D-(+)-glucose, FeCl<sub>3</sub> · 6H<sub>2</sub>O. Isopropyl β-D-thiogalactopyranoside (IPTG) was purchased from Fermentas and Pi ColorLock Gold for quantification of inorganic monophosphate from Innova Biosciences. The FeGP cofactor was extracted from [Fe]-hydrogenase from *M. marburgensis* in a solution containing 60% methanol, 1 mM 2-mercaptoethanol and 1% ammonia and then purified with Q-Sepharose and G25 columns<sup>29,30</sup>. GP was generated by irradiation of the extracted FeGP cofactor with room light under N<sub>2</sub> and then purified using HPLC with 6 ml Resource Q column (GE Healthcare) and further purified with Synergi 4 μ Polar RP 80 A column (250 × 4.6 mm, Phenomenex)<sup>30</sup>.

The *M. marburgensis* *hcgE* gene and the *M. jannaschii* *hcgF* gene with codon optimization for *Escherichia coli* host cells were purchased from GenScript USA Inc as *hcgE*-pET24b (Novagen) and *hcgF*-pET28b (Novagen) plasmids, respectively. The synthesized *M. marburgensis* *hcgE* gene encoded a thrombin cleavage site and a His<sub>6</sub>-tag at its 3' end.

**Heterologous production and purification of HcgE.** The synthesized *M. marburgensis* *hcgE* gene (Supplementary Fig. 12) inserted into the *Nde*I and *Sal*I sites of pET24b plasmid was introduced into *E. coli* BL21 (DE3) (Agilent Technologies) by transformation. The transformed cells were incubated in LB medium containing 50 μg ml<sup>-1</sup> kanamycin at 37 °C to an optical density (OD)<sub>600</sub> of 1.0. Gene expression was induced by addition of 0.5 mM IPTG. After further incubation for 4–6 h, the cells were harvested by centrifugation and stored at –80 °C until use.

The frozen cells were resuspended in a solution of 50 mM potassium phosphate, pH 7.0, 0.5 M KCl and 20 mM imidazole (buffer A) and were disrupted by sonication. The supernatant was collected after centrifugation and loaded onto a Ni<sup>2+</sup>-charged HiTrap chelating column (GE Healthcare) equilibrated with buffer A. The column was washed with 15 column volumes of buffer A and eluted with a linear gradient of imidazole from 20 to 500 mM in 50 mM potassium phosphate, pH 7.0, with 0.5 M KCl. After the fractions containing C-terminal His<sub>6</sub>-tagged HcgE protein were pooled and concentrated, the protein solution was loaded onto a HiPrep 26/10-desalting column (GE Healthcare) equilibrated with 50 mM potassium phosphate, pH 7.0, containing 0.3 M KCl. After desalting, the protein fraction was incubated with thrombin (10 units per mg HcgE) at 4 °C overnight to cleave the His<sub>6</sub>-tag. The His<sub>6</sub>-tag and thrombin were then removed on a Ni<sup>2+</sup>-charged HiTrap chelating column and a HiTrap Benzamide FF column (GE Healthcare) equilibrated with buffer A, and eluted with a linear gradient of KCl from 0.5 to 2.0 M in 50 mM KPi buffer (pH 7.0) containing 20 mM imidazole. After fractions containing the HcgE protein were pooled and concentrated, the protein solution was loaded onto a Sephacryl S-200 column (GE Healthcare) equilibrated with 50 mM potassium phosphate, pH 7.0, with 0.3 M KCl and eluted with the same buffer as used for the equilibration.

**Heterologous production and purification of HcgF.** The synthesized *M. jannaschii* *hcgF* gene (Supplementary Fig. 13) inserted into the *Nde*I and *Hind*III sites of pET28b was introduced into *E. coli* BL21 (DE3) Star (Invitrogen) by transformation. The transformed cells were grown in LB medium containing 50 μg ml<sup>-1</sup> kanamycin at 37 °C to an OD<sub>600</sub> of 1.0. Protein expression was induced by addition of 1.0 mM IPTG and further incubation for 4–6 h.

The N-terminal His<sub>6</sub>-tagged HcgF protein was purified, and the His<sub>6</sub>-tag was cleaved as described for the His<sub>6</sub>-tagged HcgE protein. To remove the His<sub>6</sub>-tag and thrombin, the reaction mixture of HcgF and thrombin was loaded onto a Ni<sup>2+</sup>-charged HiTrap chelating column and a HiTrap Benzamide FF column (GE Healthcare) equilibrated with buffer A. After the fractions containing the HcgF protein were isocratically eluted and concentrated, the protein solution was loaded onto a Sephacryl S-200 column (GE Healthcare) equilibrated with 50 mM potassium phosphate buffer (pH 7.0) containing 0.3 M KCl and eluted with the same buffer as used for the equilibration. For the SeMet-labelled HcgF preparation, *E. coli* B834 (DE3) (Novagen) cells transformed with pET28b carrying the synthesized *M. jannaschii* *hcgF* gene inserted into its *Nde*I and *Hind*III sites were incubated in M9 medium supplemented with 2.5 mM MgSO<sub>4</sub>, 2% (w/v) D-(+)-glucose, 0.01% (w/v) thiamine, 0.025 mM FeCl<sub>3</sub>, 50 μg ml<sup>-1</sup> L-selenomethionine and 50 μg ml<sup>-1</sup> kanamycin at 37 °C to an OD<sub>600</sub> of 0.5. SeMet-HcgF was purified as described for HcgE, except that all buffers contained 1 mM DTT.

**Crystallization of HcgE.** A 1-μl aliquot of HcgE (12 mg ml<sup>-1</sup>) in 10 mM MOPS/KOH, pH 7.0, was mixed with 1 μl of a reservoir solution composed of 15% (w/v) PEG6000 and 100 mM sodium citrate (pH 5.0). Crystals of HcgE were obtained by sitting-drop vapour diffusion at room temperature within 2 weeks. For co-crystallization of HcgE with ATP, an aliquot of 1 μl of HcgE (19 mg ml<sup>-1</sup>) in 10 mM MOPS/KOH, pH 7.0, was mixed with 0.2 μl of 20 mM ATP and 1.2 μl of a reservoir solution composed of 35% (w/v) pentaerythritol ethoxylate 270 (3/4 EO/OH), 200 mM ammonium sulphate and 100 mM sodium acetate, pH 4.6. Co-crystals of HcgE with ATP were obtained by sitting-drop vapour diffusion at room temperature within 1 week. For co-crystallization of HcgE with ATP and GP, a 0.5-μl aliquot of HcgE (30 mg ml<sup>-1</sup>) in 10 mM MOPS/KOH, pH 7.0, was mixed with 0.5 μl of 2.5 mM GP, 0.5 μl of 2.5 mM ATP and 1.5 μl of a reservoir solution composed of 35% (w/v) pentaerythritol ethoxylate 270 (3/4 EO/OH), 200 mM ammonium sulphate and 100 mM sodium acetate, pH 4.6. Crystals of HcgE incubated with ATP and GP were obtained by sitting-drop vapour diffusion at room temperature within 1 week.

**Crystallization of HcgF.** A 1-μl aliquot of HcgF (15 mg ml<sup>-1</sup>) in 50 mM MES/KOH, pH 7.0, was mixed with 1 μl of a reservoir solution composed of 20% (w/v) PEG3000, 200 mM NaCl and 100 mM HEPES/KOH, pH 7.5. Crystals of HcgF were obtained by sitting-drop vapour diffusion at room temperature within 2 weeks. For crystallization of SeMet-labelled HcgF, a 1 μl aliquot of SeMet-HcgF (19 mg ml<sup>-1</sup>) in 50 mM MES/KOH, pH 7.0, containing 1 mM DTT was mixed with 1 μl of a reservoir solution composed of 20% (w/v) PEG3000, 200 mM NaCl and 100 mM HEPES-NaOH, pH 7.5. Crystals of SeMet-HcgF were obtained by sitting-drop vapour diffusion at room temperature within 2 weeks. For co-crystallization of HcgF with GP, a 0.5 μl aliquot of HcgF (8.6 mg ml<sup>-1</sup>) in 10 mM MOPS/KOH, pH 7.0, containing 1 mM DTT was mixed with 0.5 μl of 2.5 mM GP and 1 μl of a reservoir solution composed of 40% (v/v) (±)-2-methyl-2,4-pentandiol and 200 mM Li<sub>2</sub>SO<sub>4</sub>. Crystals of HcgF incubated with GP were obtained by sitting-drop vapour diffusion at room temperature within 3 days.

**X-ray data collection and refinement.** The crystal of HcgE was immersed into a reservoir solution containing 20% (v/v) ethylene glycol for cryoprotection before freezing the crystal under a cryo-stream of N<sub>2</sub> at 100 K. The crystals of ATP-bound HcgE, ATP-GP-bound HcgE, HcgF, SeMet-labelled HcgF and GP-bound HcgF were frozen under a cryo-stream of N<sub>2</sub> at 100 K without adding a cryoprotectant. Diffraction data were collected on the beamline X10SA equipped with a PILATUS 6 M detector at the Swiss-Light-Source (Villigen, Switzerland) at 100 K. Data were processed using X-ray Detector Software (XDS)<sup>31</sup>. The HcgE structure was solved by molecular replacement with Molrep<sup>32</sup> using a ThiF (PDB code: 1ZUD)-based model created by the Robetta server<sup>33</sup>. Automated model building was then performed by ARP/wARP<sup>34</sup> to model amino-acid fragments and loops of HcgE. ATP-bound and ATP-GP-bound HcgE structures were solved by molecular replacement with Molrep using the solved HcgE structure. The HcgE models after molecular replacement and automated model building were completed and refined using COOT<sup>35</sup>, REFMAC5 (ref. 36) and PHENIX<sup>37</sup>. For the HcgE structure, one out of three monomers in the asymmetric unit showed electron density at the equivalent position of ATP in ATP-bound HcgE, although no ligands were added during the crystallization. On the basis of the shape of the density, ADP and not ATP could be fitted. An explanation is that ADP or ADP-like compounds may slightly co-purify with HcgE from *E. coli* cells. Phenix Xtriage<sup>37</sup> indicated that the data of HcgE co-crystallized with ATP or with both ATP and GP were merohedral twinned with twin fractions of 0.43 and 0.44 and with twin operators of *k, h, -l* and *-h, -k, l, k, h, -l, -k, -h, -l*, respectively. Twin refinement was performed with Refmac5 in the amplitude-based mode.

To determine the crystal structure of HcgF, multiple anomalous dispersion data at the selenium edge of the SeMet-HcgF crystal was measured. Selenium atom sites

were detected with SHELX C/D<sup>38</sup>. The selenium sites were refined and the phase was determined using the program SHARP and improved by the solvent-flattening procedure of SOLOMON implemented in SHARP<sup>39</sup>. The electron density after phasing and solvent flattening was shown as a blue mesh contoured at 1 $\sigma$  (Supplementary Fig. 14). Automatic model building was performed using ARP/wARP<sup>34</sup>. GP-HcgF was solved by molecular replacement with Molrep using the solved HcgF structure as a search model. Further modelling and refinement of HcgF and GP-HcgF were performed using COOT, REFMAC5 and PHENIX. The thioesterified GP model was generated by the Dundee Prodr server<sup>40</sup>.

TLS refinement was performed in the final stage of refinement<sup>41</sup>. Structures were validated using MolProbity<sup>42</sup>. Simulated annealing refined  $2F_o - F_c$  omit maps for ATP-bound HcgE, ATP-GP-bound HcgE and GP-bound HcgF were calculated using PHENIX (Supplementary Fig. 15). The resulting structures were established at 1.6 Å (HcgE, PDB: 3WV7), 1.8 Å (ATP-bound HcgE, PDB: 3WV8), 2.8 Å (ATP-GP-bound HcgE, PDB: 3WV9), 1.7 Å (HcgF, PDB: 3WVB), 1.4 Å (SeMet-labelled HcgF, PDB: 3WVA) and 2.0 Å (GP-bound HcgF, PDB: 3WVC). Data collection and refinement statistics are summarized in Supplementary Table 1. All the protein figures were generated using PyMOL (version 1.3r1, Schrödinger, LLC).

**Search for structural homologues of HcgE and HcgF.** Structural homologues of HcgE and HcgF were identified with the Dali server<sup>43</sup> using the crystal structure of HcgE and HcgF. Crystal structures of HcgE and HcgF were superimposed on their homologues using the SUPERPOSE program<sup>44</sup> in the CCP4 suite<sup>45</sup>.

**Sequence comparison of HcgE and HcgF.** The amino-acid sequences of HcgE and HcgF from *M. jannaschii*, *Methanococcus maripaludis*, *Methanobrevibacter smithii*, *M. marburgensis* and *Methanopyrus kandleri* were aligned using Clustal W2 (ref. 46). The amino-acid sequence alignments of HcgE with its structural homologues, *Schizosaccharomyces pombe* E1 ubiquitin-activating enzyme<sup>47</sup>, *E. coli* MoaB<sup>22</sup>, *E. coli* ThiF<sup>48</sup> and *E. coli* MccB<sup>24</sup> were also performed using the same method. HcgF structural homologues of *Agrobacterium fabrum* CinA (PDB code: 2A9S) and NMN deamidases, that is, *E. coli* YgaD, *Shewanella oneidensis* PncC<sup>25</sup> and *Oceanobacillus ihayensis* PncC<sup>49</sup>, were also aligned with HcgF from *M. jannaschii* using Clustal W2. The figures of the alignment were generated using ESPript3 (ref. 50).

**Docking simulation.** Molecular docking of AMP to the GP-thioester-binding site of HcgF was performed with AutoDock Vina<sup>51</sup>. For docking calculations, an AMP molecule was placed in a pocket adjacent to the GP-binding site. The calculation converged in an orientation of AMP in which its phosphate is close to the carboxy group of GP. In the case of NMN docking to CinA, the NMN molecule was placed close to a NMN-binding site proposed previously<sup>25</sup>. The calculation has converged, and the most probable binding mode of NMN to CinA was selected.

**Assay for adenylation of GP catalysed by HcgE.** The HcgE-catalysed adenylation reaction of GP was carried out in 1  $\mu$ M HcgE, 1  $\mu$ M inorganic pyrophosphatase, 0.25 mM GP, 0–0.25 mM ATP, 1 mM MgCl<sub>2</sub> and 10 mM MOPS/KOH, pH 7.0, at 30 °C for 10 min. The activity of HcgE was determined by analysing the amount of inorganic pyrophosphate released from the HcgE-catalysed reaction; the pyrophosphate was converted to monophosphate by inorganic pyrophosphatase from *S. cerevisiae* (Sigma-Aldrich Co. LCC.), and the monophosphate was analysed using Pi ColorLock Gold (Innova Biosciences). The AMP-GP product in the HcgE-catalysed reaction was analysed by matrix-assisted laser desorption time-of-flight MS and MS/MS.

**Assay for GP-HcgF thioester bond formation.** The GP-HcgF thioester bond formation was assayed by monitoring the amount of AMP resulting from a plausible reaction of HcgF with AMP-GP that was produced by HcgE-catalysed adenylation of GP with ATP. The reaction conditions were as follows: 5 mM MgCl<sub>2</sub>, 5 mM ATP, 0.25 mM GP, 27  $\mu$ M HcgE and 0–49  $\mu$ M HcgF in 11 mM potassium phosphate, pH 7.0, containing 63 mM KCl and 0.05 mM DTT at 65 °C for 10 min. The AMP amount was analysed with a HPLC system (JASCO Co.) equipped with a Synergi 4  $\mu$  fusion-RP column (4.6  $\times$  250 mm; Phenomenex), which was equilibrated with a 5 mM ammonium acetate buffer pH 4.5 to acetonitrile ratio of 97:3 (v/v). HPLC peaks were monitored at 252 nm. The amount of AMP was estimated by calculating the peak area of AMP with a standard calibration curve, which was made by HPLC analysis of the peak area of different concentrations of AMP.

## References

- Fontecilla-Camps, J. C., Amara, P., Cavazza, C., Nicolet, Y. & Volbeda, A. Structure-function relationships of anaerobic gas-processing metalloenzymes. *Nature* **460**, 814–822 (2009).
- Fontecilla-Camps, J. C., Volbeda, A., Cavazza, C. & Nicolet, Y. Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chem. Rev.* **107**, 4273–4303 (2007).
- Lubitz, W., Ogata, H., Rudiger, O. & Reijerse, E. Hydrogenases. *Chem. Rev.* **114**, 4081–4148 (2014).

- Shima, S. & Ermler, U. Structure and function of [Fe]-hydrogenase and its iron-guanilylpyridinol (FeGP) cofactor. *Eur. J. Inorg. Chem.* **2011**, 963–972 (2011).
- Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).
- Hiromoto, T. *et al.* The crystal structure of C176A mutated [Fe]-hydrogenase suggests an acyl-iron ligation in the active site iron complex. *FEBS Lett.* **583**, 585–590 (2009).
- Shima, S. *et al.* The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. *Science* **321**, 572–575 (2008).
- Ragsdale, S. W. Nickel and the carbon cycle. *J. Inorg. Biochem.* **101**, 1657–1666 (2007).
- Tan, X. S., Surovtsev, I. V. & Lindahl, P. A. Kinetics of CO insertion and acetyl group transfer steps, and a model of the acetyl-CoA synthase catalytic mechanism. *J. Am. Chem. Soc.* **128**, 12331–12338 (2006).
- Böck, A., King, P. W., Blokesch, M. & Posewitz, M. C. Maturation of hydrogenases. *Adv. Microb. Physiol.* **51**, 1–71 (2006).
- Kuchenreuther, J. M. *et al.* The HydG enzyme generates an Fe(CO)<sub>2</sub>(CN) synthon in assembly of the FeFe hydrogenase H-cluster. *Science* **343**, 424–427 (2014).
- Shepard, E. M. *et al.* [FeFe]-hydrogenase maturation. *Biochemistry* **53**, 4090–4104 (2014).
- Karyakin, A. A. *et al.* Hydrogenase electrodes for fuel cells. *Biochem. Soc. Trans.* **33**, 73–75 (2005).
- Turner, J. A. Sustainable hydrogen production. *Science* **305**, 972–974 (2004).
- Lie, T. J., Costa, K. C., Pak, D., Sakesan, V. & Leigh, J. A. Phenotypic evidence that the function of the [Fe]-hydrogenase Hmd in *Methanococcus maripaludis* requires seven hcg (hmd co-occurring genes) but not hmdII. *FEMS Microbiol. Lett.* **343**, 156–160 (2013).
- Thauer, R. K. *et al.* Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and H<sub>2</sub> storage. *Annu. Rev. Biochem.* **79**, 507–536 (2010).
- Jenney, F. E. & Adams, M. W. W. The impact of extremophiles on structural genomics (and vice versa). *Extremophiles* **12**, 39–50 (2008).
- Zhang, C. & Kim, S. H. Overview of structural genomics: from structure to function. *Curr. Opin. Chem. Biol.* **7**, 28–32 (2003).
- Fujishiro, T. *et al.* Identification of the HcgB enzyme in [Fe]-hydrogenase-cofactor biosynthesis. *Angew. Chem. Int. Ed.* **52**, 12555–12558 (2013).
- Schulman, B. A. & Harper, J. W. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat. Rev. Mol. Cell Biol.* **10**, 319–331 (2009).
- Burroughs, A. M., Iyer, L. M. & Aravind, L. Natural history of the E1-like superfamily: implication for adenylation, sulfur transfer, and ubiquitin conjugation. *Proteins* **75**, 895–910 (2009).
- Lake, M. W., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature* **414**, 325–329 (2001).
- Settembre, E., Begley, T. P. & Ealick, S. E. Structural biology of enzymes of the thiamin biosynthesis pathway. *Curr. Opin. Struct. Biol.* **13**, 739–747 (2003).
- Regni, C. A. *et al.* How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. *EMBO J.* **28**, 1953–1964 (2009).
- Galeazzi, L. *et al.* Identification of nicotinamide mononucleotide deamidase of the bacterial pyridine nucleotide cycle reveals a novel broadly conserved amidohydrolyase family. *J. Biol. Chem.* **286**, 40365–40375 (2011).
- Burstel, I. *et al.* A universal scaffold for synthesis of the Fe(CN)<sub>2</sub>(CO) moiety of [NiFe] hydrogenase. *J. Biol. Chem.* **287**, 38845–38853 (2012).
- Soboh, B. *et al.* [NiFe]-hydrogenase maturation: isolation of a HypC-HypD complex carrying diatomic CO and CN-ligands. *FEBS Lett.* **586**, 3882–3887 (2012).
- Royer, A. M., Salomone-Stagni, M., Rauchfuss, T. B. & Meyer-Klaucke, W. Iron acyl thiolato carbonyls: structural models for the active site of the [Fe]-hydrogenase (Hmd). *J. Am. Chem. Soc.* **132**, 16997–17003 (2010).
- Shima, S. & Thauer, R. K. A third type of hydrogenase catalyzing H<sub>2</sub> activation. *Chem. Rev.* **7**, 37–46 (2007).
- Shima, S. *et al.* The cofactor of the iron-sulfur cluster free hydrogenase Hmd: structure of the light-inactivation product. *Angew. Chem. Int. Ed.* **43**, 2547–2551 (2004).
- Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
- Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
- Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531 (2004).
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).

37. Afonine, P. V. *et al.* *phenix.model\_vs\_data*: a high-level tool for the calculation of crystallographic model and data statistics. *J. Appl. Crystallogr.* **43**, 669–676 (2010).
38. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
39. de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
40. Schüttelkopf, A. W. & van Aalten, D. M. F. *PRODRG*: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
41. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122–133 (2001).
42. Chen, V. B. *et al.* *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
43. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
44. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
45. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
46. Larkin, M. A. *et al.* Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
47. Olsen, S. K. & Lima, C. D. Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Mol. Cell* **49**, 884–896 (2013).
48. Lehmann, C., Begley, T. P. & Ealick, S. E. Structure of the *Escherichia coli* ThiS-ThIF complex, a key component of the sulfur transfer system in thiamin biosynthesis. *Biochemistry* **45**, 11–19 (2006).
49. Sanchez-Carron, G. *et al.* New insights into the phylogeny and molecular classification of nicotinamide mononucleotide deamidases. *PLoS ONE* **8**, e82705 (2013).
50. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323 (2003).
51. Trott, O. & Olson, A. J. Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

### Acknowledgements

We thank Professor Dr Rolf Thauer for discussions and helpful suggestions, Professor Dr Hartmut Michel for continuous support and the staff of the PXII beamline at the Swiss-Light-Source (Villigen) and Ulrike Demmer for help during data collection and X-ray structure analysis. This work was supported by a grant of the Max Planck Society to R. Thauer and a grant of the PRESTO program from the Japan Science and Technology Agency (S.S.).

### Author contributions

S.S. directed the research. T.F. and S.S. designed the research. T.F. prepared, crystallized and characterized proteins. J.K. performed mass spectroscopic analysis. T.F. and U.E. performed X-ray crystallographic analysis. T.F., U.E. and S.S. interpreted the data and wrote the manuscript.

### Additional information

**Accession codes:** Atomic coordinates and structure factors have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID codes: 3WV7, 3WV8, 3WV9, 3WVA, 3WVB and 3WVC).

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Fujishiro, T. *et al.* Protein-pyridinol thioester precursor for biosynthesis of the organometallic acyl-iron ligand in [Fe]-hydrogenase cofactor. *Nat. Commun.* 6:6895 doi: 10.1038/ncomms7895 (2015).