COMMUNICATION

Structure of the ForT/PRPP complex uncovers the mechanism of C-C bond formation in C-nucleotide antibiotic biosynthesis

Sisi Gao,[a,b] Ashish Radadiya,[c] Wenbo Li,[d] Huanting Liu,[b] Wen Zhu,[e] Valérie de Crécy-Lagard,[f] Nigel G. J. Richards*[c,g] and James H. Naismith*[d,h]

[a]   S. Gao[+]
        Research Complex at Harwell
        Didcot, OX11 0FA, United Kingdom
[b]   S. Gao[+], Dr. H. Liu
        BSRC
        University of St. Andrews
        St. Andrews, KY16 9ST, United Kingdom
[c]   Dr. A. Radadiya[+], Prof. Dr. N. G. J. Richards
        School of Chemistry
        Cardiff University
        Cardiff, CF10 3AT, United Kingdom
        E-mail: RichardsN14@cardiff.ac.uk
[d]   W. Li[+], Prof. Dr. J. H. Naismith
        Division of Structural Biology
        University of Oxford
        Oxford, OX3 7BN, United Kingdom
        E-mail: naismith@strubi.ox.ac.uk
[e]   Dr. W. Zhu
        Department of Chemistry and California Institute for Quantitative Biosciences
        University of California
        Berkeley, CA 94720, USA
[f]   Prof. Dr. V. de Crécy-Lagard
        Department of Microbiology
        University of Florida
        Gainesville, FL 32611, USA
[g]   Prof. Dr. N. G. J. Richards
        Foundation for Applied Molecular Evolution
        Alachua, FL 32415, USA
[h]   Prof. Dr. J. H. Naismith
        The Rosalind Franklin Institute
        Didcot, OX11 0FA, United Kingdom

[+]   These authors contributed equally to this work.

        Supporting information for this article is given via a link at the end of the document.

Abstract: C-C bond formation is at the heart of anabolism and organic chemistry, but relatively few enzymatic strategies for catalyzing this reaction are known. The enzyme ForT catalyzes C-C bond formation between 5'-phosphoribosyl-1'-pyrophosphate (PRPP) and 4-amino-*1H*-pyrazole-3,5-dicarboxylate to make a key intermediate in the biosynthesis of the C-nucleotide formycin A 5'-phosphate; we now report the 2.5 Å resolution structure of the ForT/PRPP complex and thus locate the active site. Site-directed mutagenesis has identified those residues critical for PRPP recognition and catalysis. Structural conservation with GHMP kinases suggests that stabilization of the negatively charged pyrophosphate leaving group is crucial for catalysis in ForT. A mechanism for this new class of C-C bond forming enzymes is proposed.

There is renewed interest in the synthesis and characterization of C-nucleosides[1] driven by the potency of GS-5734 against ebola,[2] and coronaviruses, including those that cause SARS[3] and MERS.[4] In C-nucleosides, the nucleobase is connected to C-1' of the sugar ring by a C-C bond rather than a C-N bond, thereby rendering the molecule stable and altering its stereoelectronic properties.[5,6] With the exception of pseudouridine synthase,[7,8] relatively little is known about the
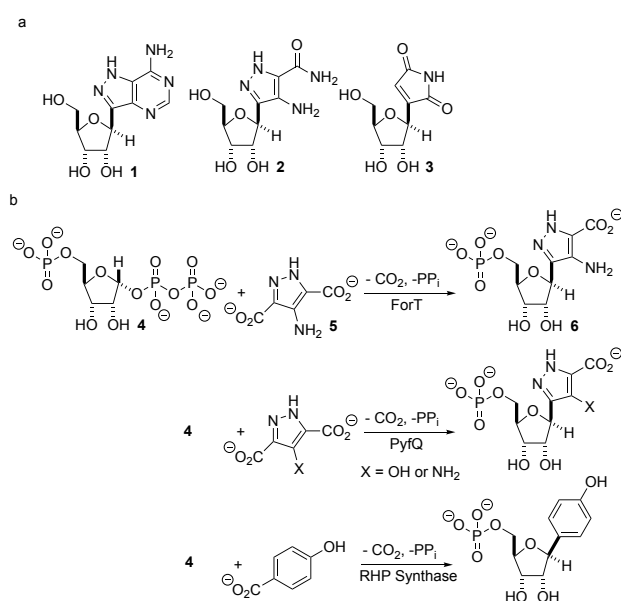
Figure 1. C-nucleotides. (a) Formycin **1**, pyrazomycin **2** and showdomycin **3**. (b) A new class of C-C bond forming enzymes that utilise 5'-phosphoribosyl-1'-pyrophosphate **4** and aromatic carboxylic acids to make C-nucleotides.

COMMUNICATION

structures and catalytic mechanisms of enzymes that form C-ribosides and C-glycosides.[9,10] Recent work identifying the biosynthetic gene clusters for formycin A **1**,[11-13] pyrazomycin **2** (also known as pyrazofurin)[14,15] and showdomycin **3**[16] (Figure 1a) has laid the foundation for obtaining an enhanced understanding of two new enzymes (ForT and PyfQ) that catalyse C-C bond-forming steps in C-nucleoside biosynthesis. Both ForT and PyfQ were originally functionally assigned based on homology to (4-(β-D-ribofuranosyl)hydroxybenzene (RHP) synthase, which mediates a key step in methanopterin biosynthesis.[17-19] These enzymes all utilize 5'-phosphoribosyl-1'-pyrophosphate (PRPP) and an aromatic carboxylic acid to make the new C-C bond (Figure 1b). The liberation of inorganic pyrophosphate and irreversible CO_2 release provide the driving force for these reactions. Sequence alignments show that these three enzymes are homologous with, and thus likely are evolutionarily related to, homoserine kinase,[20] a member of the GHMP kinase superfamily (Figure S1, Supporting Information).[21]

Two previous studies[12,14] have shown that the substrate of ForT is 4-amino-*1H*-pyrazole-3,5-dicarboxylate (ADPA) **5** (Figure 1b) and that the enzyme-catalyzed reaction yields C-nucleotide **6** (Figure 1b), an intermediate in formycin biosynthesis. Nothing is known, however, about the structure of ForT or the active site residues that play a role in substrate binding and/or catalysis. We therefore overexpressed and purified recombinant, wild type (WT) ForT in *Escherichia coli*, and grew crystals of the enzyme in the presence of 10 mM PRPP **4**. The resulting crystals were then soaked in 200 mM of PRPP **4** before data collection at the Diamond Light Source, which allowed us to solve the structure of the ForT/PRPP complex to 2.5 Å resolution (Figure 2a and Table S1, Supporting Information). The crystal asymmetric unit contains a ForT monomer with residues 11-171, 180-205 and 209-341 (C-terminus) experimentally located in electron density; we assume that the missing residues are located in conformationally disordered regions. The monomer has three antiparallel β-sheets which form the core of the structure. Two of the sheets share the same elongated strand (residues 11 to 17 for sheet 1; 18 to 24 for sheet 2) and sit end to end (Figure 2a). The third sheet sits opposite, and partially stacks against, sheet 2. There are two α-helices packed against one face of sheet 1 with a third packing against the other two. Sheet 2 has a very small α-helix packed against one of its faces whilst sheet 3 has two helices attached. A bundle of four helices is packed against the ends of sheet 2 and sheet 3. Gel filtration and multi-angle light scattering suggests the enzyme to be a dimer in solution (Figure S2, Supporting Information), and the 2-fold rotation axis in the crystal does result in one (and only one) plausible dimer that relies on contacts between the helical bundles of each monomer (Figure 2b). Interestingly, analysis of this dimeric arrangement with the PISA server[22] assigns the dimer with low confidence (0.3 on a 0 to 1 scale) due to relatively few interactions between the monomers and the limited amount of buried surface area.

The sequence of ForT is distantly related to homoserine kinase (< 20% identity), a member of the GHMP kinase superfamily, although there are regions of strong sequence conservation. The ForT structure reveals that the closest structural analogue is indeed homoserine kinase from *Methanocaldococcus jannaschii* (1fwl)[20] (root mean square deviation (rmsd) of 2.6 Å over 266 Cα atoms). Superposition reveals the conservation of the three antiparallel β-sheets and the helices (Figure S3a, Supporting Information). There is a notable

difference between the structures centred on the bundle of four helices, ForT has two longer loops (residues 25 to 35 and residues 157 to 181). Homoserine kinase is a dimer and like ForT, the dimer interface uses the same bundle of helices. However, the longer loops in ForT preclude exactly the same dimeric arrangement seen for homoserine kinase; the second monomer in the ForT dimer is rotated by approximately 60º relative to second monomer in the homoserine kinase dimer (Figure S3b, Supporting Information). ForT is also structurally related to mevalonate kinase[23] (PDB 6mde; rmsd 2.6 Å over 262 Cα atoms). The mevalonate kinase has a completely different dimeric arrangement, however, even though it too uses the same helical bundle to form the dimer (Figure S3c, Supporting Information).
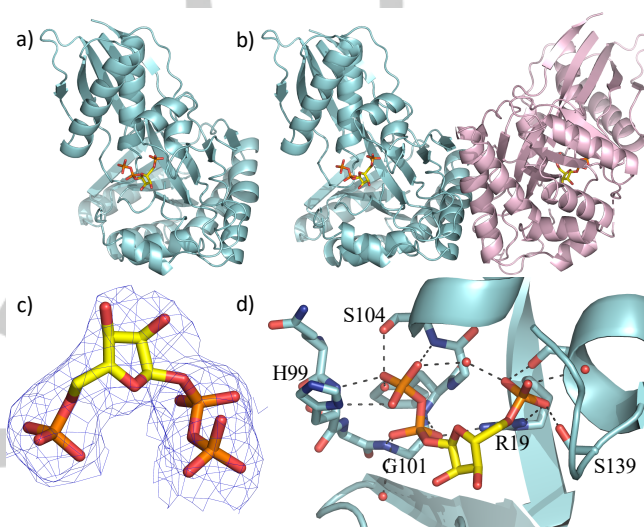
Figure 2. X-ray crystal structure of the WT ForT/PRPP complex. (a) The ForT monomer shown as a cyan cartoon, PRPP, shown in sticks (carbon yellow, phosphorous orange, oxygen red) is at the centre of the monomer. (b) The ForT dimer is generated by crystal symmetry, the second monomer is pale pink cartoon, PRPP is remote from the dimer interface. (c) The final 2Fo-Fc map contoured at 1.2σ for PRPP (shown as Figure 2a) (the original Fo-Fc map is shown in Figure S3, Supporting Information). (d) PRPP is bound to the enzyme by an extensive array of hydrogen bonds. The loop (Gln-98 to Ser-108), characteristic of GHMP kinase superfamily, plays a crucial role in substrate binding. Protein carbon atoms are colored in cyan, nitrogen in blue, other atoms as in Figure 2a.

The difference electron density map indicated the presence of bound PRPP in the enzyme (Figure S3d, Supporting Information). Pyrophosphate and phosphate only were therefore added to the model and the structure further refined. The resulting difference density connecting the phosphate and pyrophosphate moieties improved allowing fitting of the ribose ring, thereby locating PRPP in the ForT/PRPP complex (Figure 2c). PRPP is located in the centre of the monomer, remote from the dimer interface (Figure 2b,2d). The 5'-phosphate of PRPP sits in a pocket and forms salt bridges with the side chains of Arg-19 and Arg-135; it also interacts with the N-terminus of short helix in the helical bundle. Hydrogen bonds are also observed between the 5'-phosphate and the side chains of Thr-106, Ser-139, Ser-142 and the backbone NH of Ser-142 (Figure 2d). The pyrophosphate moiety sits in different pocket and also makes a salt bridge with Arg-19. The pyrophosphate is hydrogen bonded to all seven backbone amides in a tight turn comprising residues His-99 to Ser-104 as well as with the side chain of His-99 and two water

COMMUNICATION

molecules (Figure 2d). This (pyro)phosphate binding tight turn is a conserved feature of the GHMP kinases (Figures 2d and S1, Supporting Information). A bound water also appears to bridge the phosphate and pyrophosphate moieties. Our assignment relies on the observed coordination sphere being consistent with water and not of a magnesium ion. The ribose ring only makes one van der Waals contact with the side chain of Val-294. The weaker density of the ribose also suggests some conformational flexibility. We note that two disordered loops (172 to 179 and 210 to 340) are plausibly within reach of this region and it is thus possible additional protein/ribose interactions may exist.

Using isothermal titration calorimetry (ITC) the dissociation constant of the ForT/PRPP complex was determined to be 4.0 µM (Table S2, Supporting Information). We did not observe any binding of ribose 5-phosphate. Informed by the crystal structure, we prepared a series of site-specific ForT variants by mutagenesis (Table S3, Supporting Information) and used these variants in ITC measurements (Figure S4a, Supporting Information), which confirmed that Arg-19, Thr-106 and Arg-135 are indeed essential for PRPP binding (Table S2, Supporting Information). Replacing Gly-134 by proline decreased PRPP binding by almost 100-fold, consistent with the idea that the conformation of this loop segment is important in recognition.

With the crystal structure in hand, standard modeling methods were used to construct the disordered loops in ForT (Supporting Information).[24] During turnover, the active site has to accommodate the atoms of pyrophosphate and the reaction product **6**. We thus constructed two tautomers of a probe molecule (**7** and **8**), which combines the molecular features of both substrate and product (Figure 1). These "hybrid" probes (Figure 3a, 3b) were manually placed in the structure by simple superimposition on the experimentally located PRPP. The ADPA moiety of the probe molecule has extensive clashes with the protein (Figure S3e, Supporting Information), indicating either the PRPP molecule or the protein (or both) adjust their position upon catalysis. A large positively charged pocket is close the probe (Figure 3c) which would be well matched for the negatively charged ADPA molecule. This pocket is lined by several polar residues (Figure 3d).
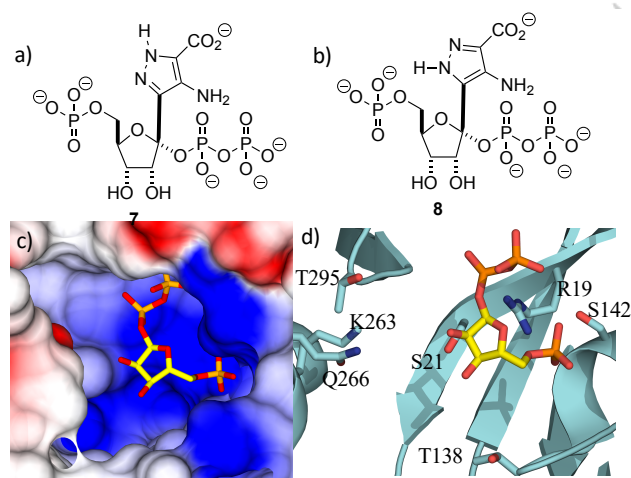
Figure 3. Location of the second substrate (a) Tautomer of a simple model for transition state. (b) Alternative tautomer of model for transition state. (c) PRPP sits a large positively charged pocket that would bind the negatively charged ADPA molecule. The binding pocket is shown as an electrostatic surface. (d)The active site pocket is lined by several polar residues, coloured as in Figure 2d.

We also evaluated the catalytic activity of the recombinant WT ForT used to obtain the X-ray crystal structure. Thus, LCMS confirmed formation of the expected product[12] when the enzyme was incubated with PRPP **4** and ADPA **5**, analysis with LCMS confirmed the production of the expected product[12] (Figure S4b, Supporting Information). The activity of WT ForT and a number of ForT variants were also assayed using membrane-inlet mass spectrometry (MIMS)[25-27] to measure ForT-catalyzed CO_2 production (Table S2 and Figure S4c, Supporting Information). These experiments showed that WT ForT has a specific activity of 0.002 U/mg under the reaction conditions used to identify the substrates for the ForT-catalysed reaction by Liu and co-workers.[12] The addition of EDTA to the reaction mixture abolished activity, suggesting that Mg^{2+} is needed for catalytic activity. The ease of the MIMS-based assay also permitted a preliminary kinetic assessment of selected ForT variants (Table S2, Supporting Information). Unfortunately, many of these ForT variants exhibited poor stability under our standard reaction conditions.
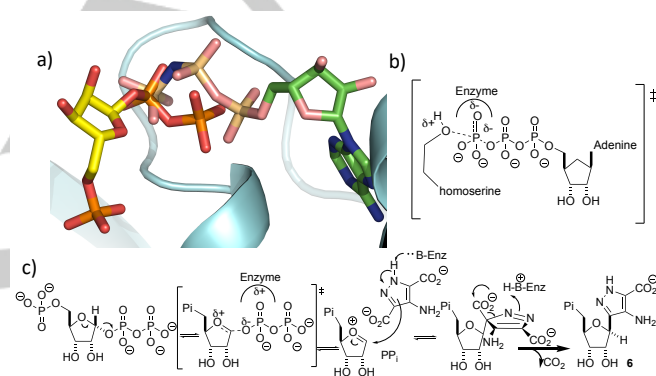
Figure 4. Mechanistic implications of the ForT/PRPP structure. (a) Overlay of PRPP in the ForT/PRPP complex with AMP-PNP homoserine kinase complex. The proximal pyrophosphate of PRPP overlaps with the γ-phosphate of AMP-PNP. (b) The transition state for γ-phosphoryl transfer in the reaction catalysed by homoserine kinase. (c) Proposed mechanism for ForT-catalysed C-C bond formation.

It is perhaps surprising that ForT should share structural homology with kinases because phosphorylation and C-nucleoside bond formation are very different chemical transformations. Superposition of the ForT/PRPP complex with that of homoserine kinase bound to phosphoaminophosphonic acid-adenylate ester (AMP-PNP) (PDB 1h72),[28] however, shows that the γ-phosphate of ATP overlaps with the C-1' phosphate of the pyrophosphate (proximal phosphate) of PRPP (Figure 4a). Similarly, superposing the ForT/PRPP complex with homoserine kinase bound to ADP (PDB 1fwk)[20] shows that the β-phosphate of ADP overlaps with the distal phosphate of the pyrophosphate in PRPP (Figure S3f, Supporting Information). We have previously noted that structural and chemical similarity between an adenylating enzyme and a (non-GHMP superfamily) kinase arose from the shared need to stabilize a negatively charged phosphate transition state[29]. In GHMP kinases, the γ-phosphate undergoes nucleophilic attack and the enzyme, using the GHMP kinase loop and a Mg^{2+} ion[20], stabilizes the developing negative charge (Figure 4b). Both the γ-phosphate in homoserine kinase and the proximal phosphate of PRRP in ForT make very similar interactions with the GHMP kinase loop in their respective

COMMUNICATION

structures. This degree of conservation we take to imply that in both enzymes, stabilization of the negatively charged phosphate is key. In ForT, this need for stabilization implies that the C-1'– pyrophosphate bond undergoes significant dissociation in the rate determining step forming an S_N1-like transition state with an oxocarbenium ion (Figure 4c). The breakage of the pyrophosphate ribose bond would permit the movement of the ribose ring which we noted earlier, is required for formation of the adduct with ADPA. The oxocarbenium ion would be highly reactive to the electron rich aromatic ADPA molecule and the subsequent elimination of CO_2 would act as an irreversible step. Based on the similar to homoserine kinase, we presume the apparent requirement for Mg^{2+} arises from stabilisation of pyrophosphate.

C-nucleotides, such as formycin, are promising medicines for the treatment of viral diseases. As part of our studies of this pathway [30], we now report the structure of ForT, the first for of a novel class of C-C bond forming enzymes. We have identified key residues that mediate substrate binding and catalysis. These findings suggest a mechanism for C-C bond formation that provides new insights into the different ways enzymes can utilize the high energy PRPP molecule.[31,32]

Acknowledgements

Keywords: C-nucleosides • Formycin • X-ray Structure • Docking • Enzyme Mechanism

[1] E. De Clerq, J. Med. Chem. **2016**, *59*, 2301-2311.

[2] T. K. Warren, R. Jordan, M. K. Lo, A. S. Ray, R. L. Mackman, V. Soloveva et al., Nature **2016**, *531*, 381-385.

[3] M. L. Agostini, E. L. Andres, A. C. Sims, R. L. Graham, T. P. Sheahan, X. Lu et al., MBio **2018**, *9*, e00222-18.

[4] E. de Wit, F. Feldmann, J. Cronin, R. Jordan, A. Okumura, T. Thomas, T. Cihlar, H. Feldmann, Proc. Natl. Acad. Sci., USA **2020**, published online Feb. 13, 2020. doi: 10.1073/pnas.1922083117.

[5] T. Hamma, A. R. Ferré-D'Amaré, Chem. Biol. **2006**, *13*, 1125-1135.

[6] X. Gu, Y. Liu, D. V. Santi, Proc. Natl. Acad. Sci., USA **1999**, *96*, 14270-14275.

[7] J. Štambasky, M. Hocek, P. Kocovský, Chem. Rev. **2009**, *109*, 6729-6764.

[8] K. Temburnikar, K. L. Seley-Radtke, Beilstein J. Org. Chem. **2018**, *14*, 772-785.

[9] T. Oja, L. Niiranen, T. Sandalova, K. D. Klika, I. Niemo, P. Mäntsälä, G. Schneider, M. Metsä-Ketelä, Proc. Natl. Acad. Sci., USA **2013**, *110*, 1291-1296.

[10] D. Foshag, C. Campbell, P. D. Pawelek, Biochim. Biophys. Acta: Proteins & Proteomics **2014**, *1844*, 1619-1630.

[11] S.-A. Wang, Y. Ko, J. Zeng, Y. Geng, D. Ren, Y. Ogasawara, Y. Irani, Y. Zhang, H.-W. Liu, J. Am. Chem. Soc. **2019**, *141*, 6127-6131.

[12] D. Ren, S.-A. Wang, Y. Ko, Y. Geng, Y. Ogasawara, H.-W. Liu, Angew. Chem. Int. Edn. **2019**, *58*, 16512-16516.

[13] T. Shiraishi, T. Kuzuyama, J. Antibiot. **2019**, *72*, 913-923.

[14] M. Zhang, P. Zhang, G. Xu, W. Zhou, Y. Gao, R. Gong, Y.-S. Cai, H. Cong, Z. Deng, N. J. P. Price, X. Mao, W. Chen, Appl. Environ. Microbiol. **2020**, *86*, e01971-19.

[15] G. Zhao, S. Yao, K. W. Rothchild, T. Liu, Y. Liu, J. Lian, H.-Y. He, K. S. Ryan, Y.-L. Du, ChemBioChem **2020**, *21*, 644-649.

[17] R. H. White, Biochemistry **2011**, *50*, 6041-6052.

[18] R. V. Dumitru, S. W. Ragsdale, J. Biol. Chem. **2004**, *279*, 39389-39395.

[19] M. E. Rasche, R. H. White, Biochemistry **1998**, *37*, 11343-11351.

[20] T. Zhou, M. Daugherty, N. V. Grishin, A. L. Osterman, H. Zhang, Structure, Fold. Des. **2000**, *8*, 1247-1257.

[21] P. Bork, C. Sander, A. Valencia, Protein Sci. **1993**, *2*, 31-40.

[22] E. Krissinel, K. Henrick, J. Mol. Biol. **2007**, *372*, 774-797.

[23] B. R. Miller, Y. Kung, PLoS One. **2018**, *13*, e0208419.

[24] N. Eswar, D. Eramian, B. Webb, M. Y. Shen, A. Sali, Methods Mol. Biol. **2008**, *426*, 145-159.

[25] M. E. G. Moral, C. K. Tu, N. G. J. Richards, D. N. Silverman, Analyt. Biochem. **2011**, *418*, 73-77.

[26] X. Sheng, W. Zhu, J. Huddleston, D. F. Xiang, F. M. Raushel, N. G. J. Richards, F. Himo, ACS Catal. **2017**, *7*, 4968-4974.

[27] W. Zhu, L. A. Reinhardt, N. G. J. Richards, Biochemistry **2018**, *57*, 3425-3432.

[28] S. S. Krishna, T. Zhou, M. Daugherty, A. L. Osterman, H. Zhang, Biochemistry **2001**, *40*, 10810-10818.

[29] S. Schmelz, N. Kadi, S. A. McMahon, L. Song, D. Oves-Costales, M. Oke, H. Liu, K. A. Johnson, L. G. Carter, C. H. Botting, M. F. White, G. L. Challis, J. H. Naismith, Nature Chem. Biol. **2009**, *5*, 174-182.

[30] S. Gao, H. Liu, V.de Crécy-Lagard, W. Zhu, N. G. J. Richards, J. H. Naismith, Chem. Commun. **2019**, *55*, 14502-14505.

[31] M. S. Alphey, G. Fisher, Y. Ge, E. R. Gould, T. F. G. Machado, H. Liu, G. J. Florence, J. H. Naismith, R. G. da Silva, ACS Catal. **2018**, *8*, 5601-5610.

[32] B. Hove-Jensen, K. R. Andersen, M. Kilstrup, J. Martinussen, R. L. Switzer, M. Willemoës, Microbiol. Mol. Biol. Rev. **2017**, *81*, e00040-16.

WILEY-VCH

Entry for the Table of Contents
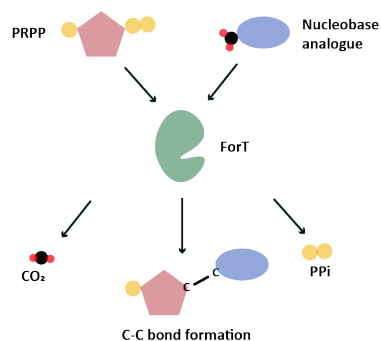
A new class of enzymes catalyse C-C bond formation by irreversible CO_2 and pyrophosphate production.