


 Cite this: *Chem. Commun.*, 2020, 56, 7617

 Received 19th April 2020,  
 Accepted 14th May 2020

DOI: 10.1039/d0cc02834g

rsc.li/chemcomm

# Uncovering the chemistry of C–C bond formation in C-nucleoside biosynthesis: crystal structure of a C-glycoside synthase/PRPP complex†

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

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 formycin A 5'-phosphate by *Streptomyces kaniharzensis*. We report the 2.5 Å resolution structure of the ForT/PRPP complex and locate active site residues critical for PRPP recognition and catalysis.

There is renewed interest in the synthesis and characterization of C-nucleosides<sup>1</sup> driven by the potency of GS-5734 against ebola,<sup>2</sup> and coronaviruses, including those that cause SARS<sup>3</sup> and MERS.<sup>4</sup> 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 improving hydrolytic stability and altering its stereo-electronic properties.<sup>5,6</sup> With the exception of pseudouridine synthase,<sup>7,8</sup> relatively little is known about the structures and catalytic mechanisms of enzymes that form C-ribosides and C-glycosides.<sup>9,10</sup> Recent work identifying the biosynthetic gene clusters for formycin A 1,<sup>11–13</sup> pyrazomycin 2 (also known as pyrazofurin)<sup>14,15</sup> and showdomycin 3<sup>16</sup> (Fig. 1a) has laid the foundation for obtaining an enhanced understanding of two new enzymes (ForT and PyfQ), both of which catalyse C–C bond-forming steps in C-nucleotide biosynthesis. The function of ForT and PyfQ was originally assigned based on their homology to the enzyme (4-(β-D-ribofuranosyl)hydroxybenzene) (RHP) synthase, which mediates a key step in methanopterin

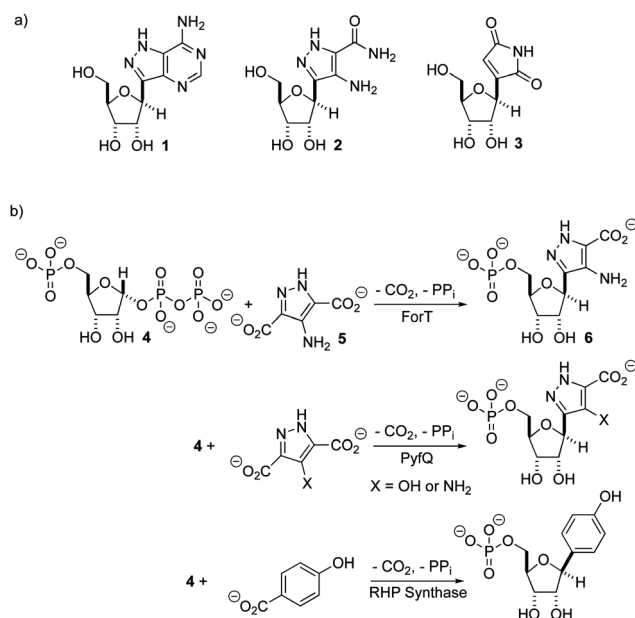


Fig. 1 (a) Formycin A 1, pyrazomycin 2 and showdomycin 3. (b) A new class of C–C bond forming enzymes that utilise 5'-phosphoribosyl-1'-pyrophosphate 4 and aromatic carboxylic acids to make C-nucleotides.

biosynthesis.<sup>17–20</sup> These enzymes all utilize 5'-phosphoribosyl-1'-pyrophosphate (PRPP) and an aromatic carboxylic acid to make the new C–C bond (Fig. 1b). The liberation of inorganic pyrophosphate and irreversible CO<sub>2</sub> release provide the driving force for these reactions. Sequence alignments show that these three enzymes are homologous with, and therefore evolutionarily related to, homoserine kinase,<sup>21</sup> a member of the GHMP kinase superfamily (Fig. S1, ESI†).<sup>22</sup>

Two previous studies<sup>12,14</sup> have shown that the substrate of ForT is 4-amino-1H-pyrazole-3,5-dicarboxylate (ADPA) 5 (Fig. 1b) and that the enzyme-catalyzed reaction yields C-nucleotide 6 (Fig. 1b), an intermediate in formycin biosynthesis. Nothing was reported, however, concerning the structure of ForT or the active site residues that play a role in substrate binding and/or catalysis.

<sup>a</sup> Research Complex at Harwell, Didcot, OX11 0FA, UK

<sup>b</sup> BSRC, University of St Andrews, St Andrews, KY16 9ST, UK

<sup>c</sup> School of Chemistry, Cardiff University, Cardiff, CF10 3AT, UK

<sup>d</sup> Division of Structural Biology, University of Oxford, Oxford, OX3 7BN, UK.

E-mail: naismith@strubi.ox.ac.uk

<sup>e</sup> Department of Chemistry and California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720, USA

<sup>f</sup> Department of Microbiology, University of Florida, Gainesville, FL 32611, USA

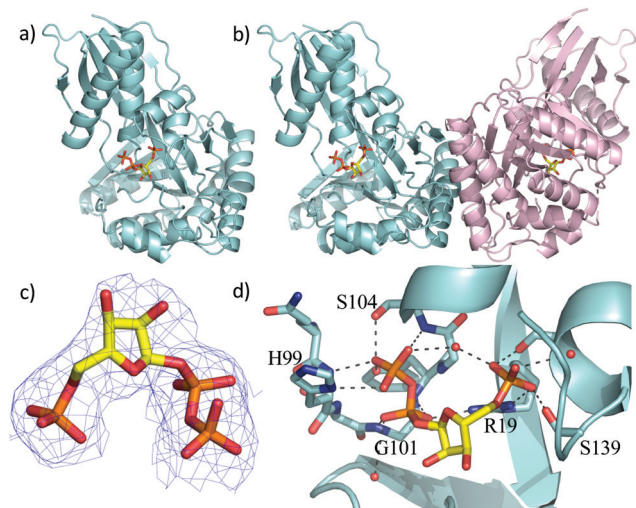
<sup>g</sup> Foundation for Applied Molecular Evolution, Alachua, FL 32615, USA

<sup>h</sup> The Rosalind Franklin Institute, Didcot, OX11 0FA, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cc02834g

‡ These authors contributed equally.





**Fig. 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  $2F_o - F_c$  map contoured at  $1.2\sigma$  for PRPP (shown as in Fig. 2a) (the original  $F_o - F_c$  map is shown in Fig. S3, ESI<sup>†</sup>). (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 coloured in cyan, nitrogen in blue, other atoms as in Fig. 2a.

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 soaked in 200 mM of PRPP **4** before data collection at the Diamond Light Source, allowing us to solve the structure of the ForT/PRPP complex to 2.5 Å resolution (Fig. 2a and Table S1, ESI<sup>†</sup>).§ 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 missing residues are located in conformationally disordered regions. The ForT monomer has three antiparallel  $\beta$ -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 (Fig. 2a). The third sheet sits opposite, and partially stacks against, sheet 2. There are two  $\alpha$ -helices packed against one face of sheet 1 with a third packing against the other two. Sheet 2 has a very small  $\alpha$ -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 (Fig. S2, ESI<sup>†</sup>), 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 (Fig. 2b). Interestingly, analysis of this dimeric arrangement with the PISA server<sup>23</sup> 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)<sup>21</sup> (root mean square deviation (rmsd) of 2.6 Å over 266 C $\alpha$  atoms). Superposition reveals the conservation of the three antiparallel  $\beta$ -sheets and the helices (Fig. S3a, ESI<sup>†</sup>). 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 60° relative to second monomer in the homoserine kinase dimer (Fig. S3b, ESI<sup>†</sup>). ForT is also structurally related to mevalonate kinase<sup>24</sup> (PDB 6mde; rmsd 2.6 Å over 262 C $\alpha$  atoms). The mevalonate kinase has a completely different dimeric arrangement, however, even though it too uses the same helical bundle to form the dimer (Fig. S3c, ESI<sup>†</sup>).

The difference electron density map indicated the presence of bound PRPP in the enzyme (Fig. S3d, ESI<sup>†</sup>). 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 (Fig. 2c). PRPP is located in the centre of the monomer, remote from the dimer interface (Fig. 2b and d). 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 (Fig. 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 molecules (Fig. 2d). This (pyro)phosphate binding tight turn is a conserved feature of the GHMP kinases (Fig. 2d and Fig. S1, ESI<sup>†</sup>). 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. A model of the PRPP complex of RFA-P synthase, another member of this class of C–C bond-forming enzymes (Fig. 1b) and homologous to ForT, has been reported.<sup>20</sup> In the model, R26 of RFA-P synthase,<sup>20</sup> was identified to be adjacent to PRPP and important for binding the second substrate. The equivalent residue, (R33) in our ForT/PRPP complex is around 18 Å distant from PRPP with intervening secondary structure, thus any role in catalysis would require significant structural re-arrangement.



It therefore appears that the model<sup>20</sup> and our crystal structure differ significantly in their placement of PRPP.

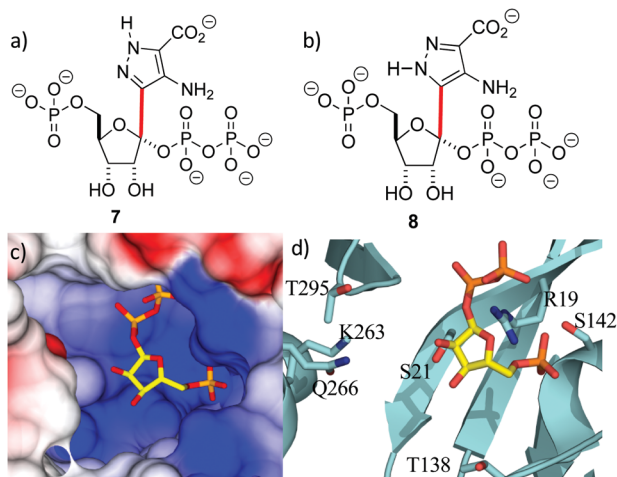
Using isothermal titration calorimetry (ITC) the dissociation constant of the ForT/PRPP complex was determined to be 4.0  $\mu\text{M}$  (Table S2, ESI<sup>†</sup>). We did not observe any binding of ribose 5-phosphate. Informed by the crystal structure, we prepared a series of site-specific ForT variants by mutagenesis (Table S3, ESI<sup>†</sup>) and used these variants in ITC measurements (Fig. S4a, ESI<sup>†</sup>), which confirmed that Arg-19, Thr-106 and Arg-135 are indeed essential for PRPP binding (Table S2, ESI<sup>†</sup>). 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 (ESI<sup>†</sup>)<sup>25</sup> The active site has to accommodate the atoms of both PRPP **4** and the reaction product **6** during catalytic turnover. Making the reasonable assumption that C–C bond formation takes place on the opposite face of the ribose ring to that of the pyrophosphate moiety, we constructed molecular probes **7** and **8** (Fig. 3a and b). The probes represent molecular “hybrids” of PRPP **4** and product **6** and when docked into the PRPP complex crystal structure, the rotamers around the C–C bond (highlighted red in Fig. 3) helped explore the active site. These qualitative studies revealed a large positively charged pocket lined by several polar residues (Fig. 3c and d), which would be well suited to accommodate the negatively charged ADPA substrate **5**. These models also suggest that the protein likely undergoes some conformational rearrangement to accommodate both substrates within the active site (Fig. S3e, ESI<sup>†</sup>).

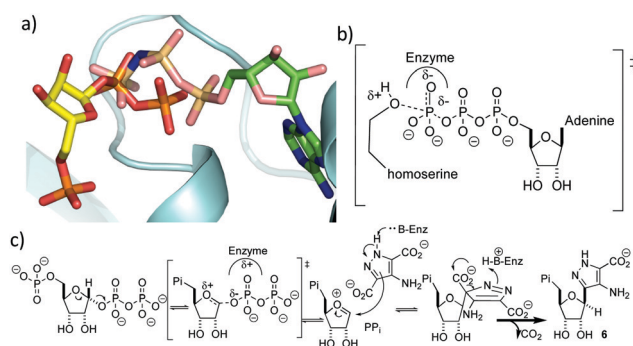
We also evaluated the catalytic activity of the recombinant WT ForT used to obtain the X-ray crystal structure. Thus, LCMS confirmed formation of the expected product<sup>12</sup> when the

enzyme was incubated with PRPP **4** and ADPA **5**, analysis with LCMS confirmed the production of the expected product<sup>12</sup> (Fig. S4b, ESI<sup>†</sup>). The activity of WT ForT and a number of ForT variants were also assayed using membrane-inlet mass spectrometry (MIMS)<sup>26–28</sup> to measure ForT-catalysed CO<sub>2</sub> production (Table S2 and Fig. S4c, ESI<sup>†</sup>). These experiments showed that WT ForT has a specific activity of 0.002 U mg<sup>−1</sup> under the reaction conditions used to identify the substrates for the ForT-catalysed reaction by Liu and co-workers.<sup>12</sup> The addition of EDTA to the reaction mixture abolished activity, suggesting that Mg<sup>2+</sup> is needed for catalytic activity. The ease of the MIMS-based assay also permitted a preliminary kinetic assessment of selected ForT variants (Table S2, ESI<sup>†</sup>). Unfortunately, many of these ForT variants exhibited poor stability under our standard reaction conditions.

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),<sup>29</sup> however, shows that the  $\gamma$ -phosphate of ATP overlaps with the C-1' phosphate of the pyrophosphate (proximal phosphate) of PRPP (Fig. 4a). Similarly, superposing the ForT/PRPP complex with homoserine kinase bound to ADP (PDB 1fwk)<sup>21</sup> shows that the  $\beta$ -phosphate of ADP overlaps with the distal phosphate of the pyrophosphate in PRPP (Fig. S3f, ESI<sup>†</sup>). 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.<sup>30</sup> In GHMP kinases, the  $\gamma$ -phosphate undergoes nucleophilic attack and the enzyme, using the GHMP kinase loop and a Mg<sup>2+</sup> ion,<sup>21</sup> stabilizes the developing negative charge (Fig. 4b). Both the  $\gamma$ -phosphate in homoserine kinase and the proximal phosphate of PRPP in ForT make very similar interactions with the GHMP kinase loop in their respective structures. This degree of conservation we take to imply that in both enzymes, stabilization of the negatively



**Fig. 3** Location of the second substrate (a and b) probes **7** and **8** used in model building studies. Several rotamers about the C-1'–C-4 bond (red) were considered. (c) PRPP sits in a large positively charged pocket that could also 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 Fig. 2d.



**Fig. 4** (a) Overlay of PRPP in the ForT/PRPP complex with AMP-PNP homoserine kinase complex. The proximal pyrophosphate of PRPP overlaps with the  $\gamma$ -phosphate of AMP-PNP. (b) The transition state for  $\gamma$ -phosphoryl transfer in the reaction catalysed by homoserine kinase. (c) Proposed mechanism for ForT-catalysed C–C bond formation. Note that the configuration of the correct epimer at the new stereogenic centre remains to be established.



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<sub>N</sub>1-like transition state with an oxocarbenium ion (Fig. 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<sub>2</sub> would act as an irreversible step. Based on the similarity to homoserine kinase, we presume the apparent requirement for Mg<sup>2+</sup> arises from stabilisation of pyrophosphate.

C-Nucleosides, such as formycin, are promising medicines for the treatment of viral diseases. Understanding their biosynthesis lays a firm foundation for identifying new biocatalytic approaches to novel anti-viral agents.<sup>31</sup> The work reported here, which continues our ongoing effort to obtain structures for all of the enzymes involved in formycin biosynthesis,<sup>32</sup> represents a significant advance by identifying key residues that mediate substrate binding and catalysis, setting the scene for determining detailed functional studies and reengineering efforts. In addition, these findings provide new insights into the different ways enzymes can bind and utilize the high energy PRPP molecule.<sup>33,34</sup>

We thank for the Diamond Light Source for beam time allocation and beam line staff for assistance with data collection. Funding for these studies was provided by BBSRC (BB/T006161/1 & BB/T006188/1 to J. H. N. & N. G. J. R., respectively), and the National Institutes of Health (R01 GM129793 to V. d. C.-L.)

## Conflicts of interest

There are no conflicts to declare.

## Notes and references

§ Coordinates and structure factors for the ForT/PRPP complex have been deposited in the PDB as entry 6YQQ. The accession number for the gene encoding *Streptomyces kaniharaensis* ForT is WP\_153459976.<sup>35</sup>

- 1 E. De Clercq, *J. Med. Chem.*, 2016, **59**, 2301.
- 2 T. K. Warren, R. Jordan, M. K. Lo, A. S. Ray, R. L. Mackman and V. Soloveva, *et al.*, *Nature*, 2016, **531**, 381.
- 3 M. L. Agostini, E. L. Andres, A. C. Sims, R. L. Graham, T. P. Sheahan and 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 and H. Feldmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 6771.
- 5 T. Hamma and A. R. Ferré-D'Amaré, *Chem. Biol.*, 2006, **13**, 1125.

- 6 X. Gu, Y. Liu and D. V. Santi, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14270.
- 7 J. Štambaský, M. Hocek and P. Kocovský, *Chem. Rev.*, 2009, **109**, 6729.
- 8 K. Temburnikar and K. L. Seley-Radtke, *Beilstein J. Org. Chem.*, 2018, **14**, 772.
- 9 T. Oja, L. Niiranen, T. Sandalova, K. D. Klika, I. Niemo, P. Mäntsälä, G. Schneider and M. Metsä-Ketelä, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 1291.
- 10 D. Foshag, C. Campbell and P. D. Pawelek, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 1619.
- 11 S.-A. Wang, Y. Ko, J. Zeng, Y. Geng, D. Ren, Y. Ogasawara, S. Irani, Y. Zhang and H.-W. Liu, *J. Am. Chem. Soc.*, 2019, **141**, 6127.
- 12 D. Ren, S.-A. Wang, Y. Ko, Y. Geng, Y. Ogasawara and H.-W. Liu, *Angew. Chem., Int. Ed.*, 2019, **58**, 16512.
- 13 T. Shiraishi and T. Kuzuyama, *J. Antibiot.*, 2019, **72**, 913.
- 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 and 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 and Y.-L. Du, *ChemBioChem*, 2020, **21**, 644.
- 16 K. Palmu, P. Rosenqvist, K. Thapa, Y. Ilina, V. Siitonen, B. Baral, J. Mäkinen, G. Belogurov, P. Virta, J. Niemi and M. Metsä-Ketelä, *ACS Chem. Biol.*, 2017, **12**, 1472.
- 17 R. H. White, *Biochemistry*, 2011, **50**, 6041.
- 18 R. V. Dumitru and S. W. Ragsdale, *J. Biol. Chem.*, 2004, **279**, 39389.
- 19 M. E. Rasche and R. H. White, *Biochemistry*, 1998, **37**, 11343.
- 20 M. E. Bechard, P. Farahani, D. Greene, A. Pham, A. Orry and M. E. Rasche, *AIMS Microbiol.*, 2019, **5**, 186.
- 21 T. Zhou, M. Daugherty, N. V. Grishin, A. L. Osterman and H. Zhang, *Structure Fold. Des.*, 2000, **8**, 1247.
- 22 P. Bork, C. Sander and A. Valencia, *Protein Sci.*, 1993, **2**, 31.
- 23 E. Krissinel and K. Henrick, *J. Mol. Biol.*, 2007, **372**, 774.
- 24 B. R. Miller and Y. Kung, *PLoS One*, 2018, **13**, e0208419.
- 25 N. Eswar, D. Eramian, B. Webb, M. Y. Shen and A. Sali, *Methods Mol. Biol.*, 2008, **426**, 145.
- 26 M. E. G. Moral, C. K. Tu, N. G. J. Richards and D. N. Silverman, *Anal. Biochem.*, 2011, **418**, 73.
- 27 X. Sheng, W. Zhu, J. Huddleston, D. F. Xiang, F. M. Raushel, N. G. J. Richards and F. Himo, *ACS Catal.*, 2017, **7**, 4968.
- 28 W. Zhu, L. M. Easton, L. A. Reinhardt, C.-K. Tu, D. N. Silverman, K. N. Allen and N. G. J. Richards, *Biochemistry*, 2016, **55**, 2163.
- 29 S. S. Krishna, T. Zhou, M. Daugherty, A. L. Osterman and H. Zhang, *Biochemistry*, 2001, **40**, 10810.
- 30 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 and J. H. Naismith, *Nat. Chem. Biol.*, 2009, **5**, 174–182.
- 31 T. Shirashi and T. Kuzuyama, *J. Antibiot.*, 2019, **72**, 913.
- 32 S. Gao, H. Liu, V. de Crécy-Lagard, W. Zhu, N. G. J. Richards and J. H. Naismith, *Chem. Commun.*, 2019, **55**, 14502.
- 33 M. S. Alphey, G. Fisher, Y. Ge, E. R. Gould, T. F. G. Machado, H. Liu, G. J. Florence, J. H. Naismith and R. G. da Silva, *ACS Catal.*, 2018, **8**, 5601.
- 34 B. Hove-Jensen, K. R. Andersen, M. Kilstруп, J. Martinussen, R. L. Switzer and M. Willemoës, *Microbiol. Mol. Biol. Rev.*, 2017, **81**, e00040-16.
- 35 W. Zhu, X. Liu, M. Hughes, V. de Crécy-Lagard and N. G. J. Richards, *Microbiol. Resour. Announce.*, 2020, **9**, e01434.

