Article

Structure of a B_{12}-dependent radical SAM enzyme in carbapenem biosynthesis

https://doi.org/10.1038/s41586-021-04392-4

Received: 10 July 2021

Accepted: 22 December 2021

Published online: 2 February 2022

Check for updates

Hayley L. Knox[1,5], Erica K. Sinner[2,5], Craig A. Townsend[2]✉, Amie K. Boal[1,3]✉ & Squire J. Booker[1,3,4]✉

Carbapenems are antibiotics of last resort in the clinic. Owing to their potency and broad-spectrum activity, they are an important part of the antibiotic arsenal. The vital role of carbapenems is exemplified by the approval acquired by Merck from the US Food and Drug Administration (FDA) for the use of an imipenem combination therapy to treat the increased levels of hospital-acquired and ventilator-associated bacterial pneumonia that have occurred during the COVID-19 pandemic[1]. The C6 hydroxyethyl side chain distinguishes the clinically used carbapenems from the other classes of β-lactam antibiotics and is responsible for their low susceptibility to inactivation by occluding water from the β-lactamase active site[2]. The construction of the C6 hydroxyethyl side chain is mediated by cobalamin- or B_{12}-dependent radical *S*-adenosylmethionine (SAM) enzymes[3]. These radical SAM methylases (RSMTs) assemble the alkyl backbone by sequential methylation reactions, and thereby underlie the therapeutic usefulness of clinically used carbapenems. Here we present X-ray crystal structures of TokK, a B_{12}-dependent RSMT that catalyses three-sequential methylations during the biosynthesis of asparenomycin A. These structures, which contain the two metallocofactors of the enzyme and were determined in the presence and absence of a carbapenam substrate, provide a visualization of a B_{12}-dependent RSMT that uses the radical mechanism that is shared by most of these enzymes. The structures provide insight into the stereochemistry of initial C6 methylation and suggest that substrate positioning governs the rate of each methylation event.

During the biosynthesis of all known complex carbapenem natural products, the assembly of the C6 alkyl side chain (Fig. 1a) is accomplished by a cobalamin (Cbl or B_{12})-dependent radical SAM enzyme[3–5]. These catalysts can perform serial methyl transfers with control of stereochemical outcome for each reaction. The Cbl-dependent RSMT ThnK performs two sequential methyl transfers to its (2*R*)-pantetheinylated carbapenam substrate **1** during the biosynthesis of the paradigm carbapenem antibiotic thienamycin (**2**) (ref. [6]). An orthologue of ThnK, TokK from *Streptomyces tokunonensis* (ATCC 31569), constructs the C6 isopropyl chain of the carbapenem, asparenomycin A (**3**), by deploying three sequential methylations of **1** (Fig. 1b). This biosynthetic approach allows the producing organism to make a small 'library' of alkylated analogues, which may deter the development of resistance in susceptible bacteria. This strategy may also be used in the biosynthesis of cystobactamids, in which a similar Cbl-dependent radical SAM (RS) enzyme, CysS, performs successive methyl transfers[7] (Extended Data Fig. 1). Notably, despite low sequence identity (around 29%) between CysS and TokK or ThnK, all three proteins are located within the same cluster of a sequence similarity network (SSN) composed of approximately 11,000 Cbl-binding RS enzymes obtained from networks provided by radicalSAM.org (https://radicalsam.org/) and the USCF structure–function linkage database (SFLD) (Fig. 1c, Supplementary Fig. 1). It is tempting to speculate that this colocalization might be driven by mechanistic similarities.

Carbapenem C6 alkyl chain construction requires stereoselective formation of carbon–carbon bonds between unactivated sp^3-hybridized carbons. Cbl-dependent RSMTs are the only known biological catalysts capable of such transformations. The Cbl-containing subfamily, depicted as an SSN in Fig. 1c, is also one of the largest in the RS superfamily, a diverse group that functions in the biosynthesis of chlorophyll, lipids and natural products with antiproliferative biological activity[8–11]. Although most Cbl-dependent RS enzymes have unknown functions, those that have been characterized are generally—but not exclusively—methylases that act on carbon or phosphorus centres by using methylcobalamin (MeCbl) as an intermediate methyl donor. All RS enzymes, with a single known exception[12–14], reductively cleave SAM to generate methionine (Met) and a 5′-deoxyadenosyl 5′-radical (5′-dA•) (Fig. 1b). The latter reactive intermediate typically initiates catalysis with a target substrate by

[1]Department of Chemistry, Pennsylvania State University, University Park, PA, USA. [2]Department of Chemistry, Johns Hopkins University, Baltimore, MD, USA. [3]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA. [4]The Howard Hughes Medical Institute, Pennsylvania State University, University Park, PA, USA. [5]These authors contributed equally: Hayley L. Knox, Erica K. Sinner. ✉e-mail: ctownsend@jhu.edu; akb20@psu.edu; squire@psu.edu

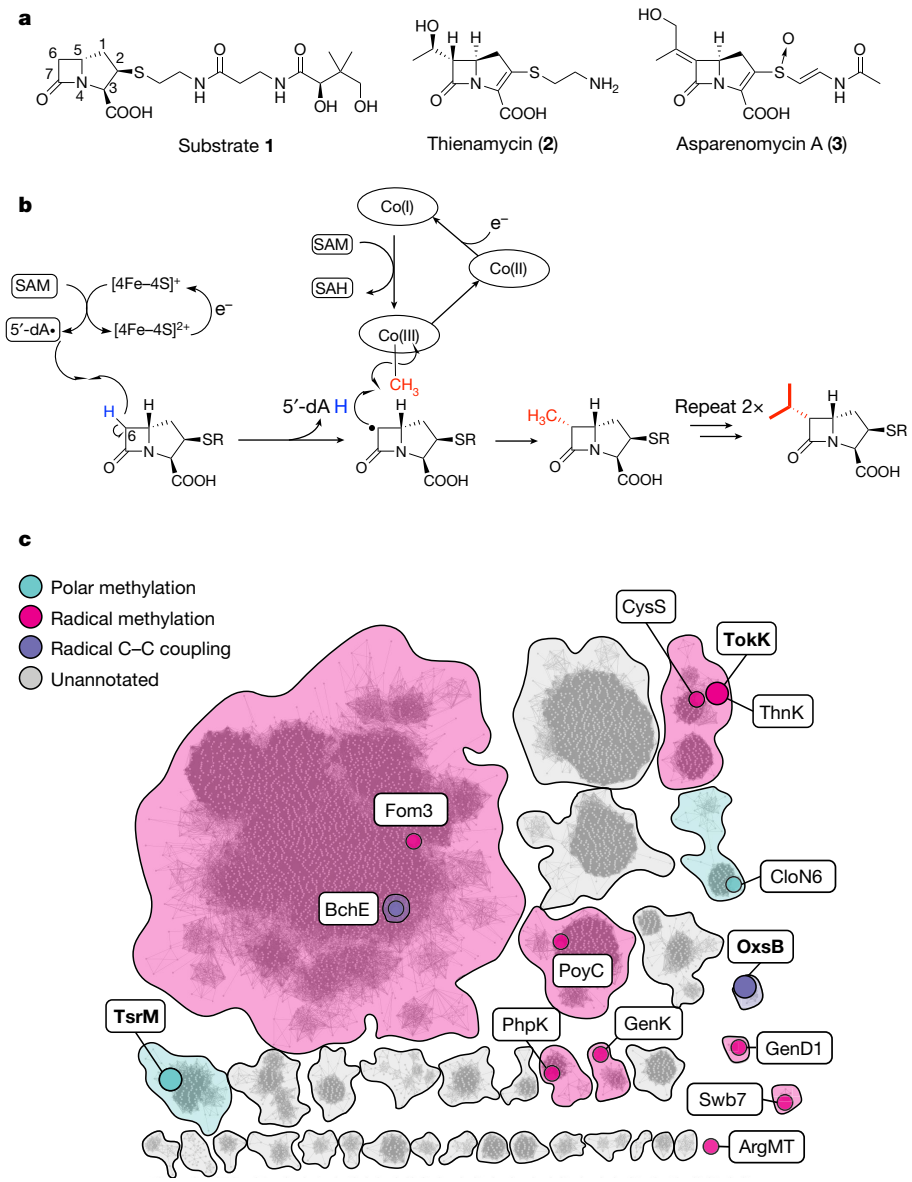


Fig 1 | Cbl-dependent radical-mediated methylations in carbapenem biosynthesis. a, Structure of (2*R*)-pantetheinylated carbapenam precursor substrate **1** modified by ThnK and TokK. Related carbapenem natural products containing C6-alkyl substituents include thienamycin and asparenomycin A. **b**, Proposed mechanism describing the three sequential methylations catalysed by TokK. **c**, An abbreviated SSN of Cbl-binding RS enzymes, highlighting selected sequence clusters and nodes. The full network (Supplementary Fig. 1) was generated from around 11,000 annotated Cbl-dependent RS enzymes with an alignment score of 65. Each node represents a single sequence or a set of sequences with more than 40% sequence identity. Sequence clusters and nodes are coloured by predicted reaction mechanism (see key in the figure). Nodes containing functionally annotated sequences are indicated in colour. Structurally characterized enzymes are represented by enlarged nodes and labeled in boldface.

abstracting a hydrogen atom. In B_{12}-dependent RSMTs, the substrate radical attacks the methyl group of MeCbl, inducing homolytic cleavage of the cobalt–carbon bond to yield cob(II)alamin and the methylated product (Fig. 1b). After dissociation of the methylated product, Met and 5′-deoxyadenosine (5′-dAH), and rebinding of another molecule of SAM, cob(II)alamin is reduced to cob(I)alamin. Co(I) is a supernucleophile, which acquires a methyl group from SAM to regenerate MeCbl (refs. [10,11]) (Fig. 1b).

Two Cbl-dependent RS enzymes have been structurally characterized, TsrM and OxsB (refs. [12,15]) (Fig. 1c, Supplementary Fig. 1), which are involved in the biosynthesis of the antibiotics thiostrepton A and oxetanocin A, respectively. Both enzymes are mechanistic outliers among Cbl-dependent RS enzymes and are found in SSN clusters distinct from each other and from TokK (Fig. 1c). OxsB uses Cbl in an unknown manner to catalyse a complex ring contraction of 2′-deoxyadenosine monophosphate (dAMP)[15] (Extended Data Fig. 2a). TsrM methylates an sp^2-hybridized carbon, C2, of L-tryptophan (Trp) by a polar mechanism[12] (Extended Data Fig. 2b). TsrM is distinctive among all RS enzymes because it does not catalyse the formation of 5′-dA• during catalysis[12]. Instead, TsrM uses SAM's carboxylate moiety as an acceptor of the N1 proton of Trp during C2 electrophilic substitution by MeCbl[12,16]. In addition, the structures of TsrM and OxsB have limitations that prevent full understanding of Cbl-dependent RS catalysis. The structure of OxsB lacks the dAMP substrate[15]. TsrM has been co-crystallized with aza-SAM (a SAM analogue) and Trp, but the Trp substrate is bound in an unproductive conformation, requiring computational docking to understand the structural basis for the reaction outcome.

TokK was crystallized under anoxic conditions in the presence of 5′-dAH and Met, the products of reductive SAM cleavage

(Supplementary Fig. 2). Structures of this complex were solved in the absence and presence of substrate **1** to resolutions of 1.79 Å and 1.94 Å, respectively (Extended Data Table 1). TokK shares in common with TsrM and OxsB an N-terminal Cbl-binding domain and a central RS domain containing a [4Fe–4S] cluster (Extended Data Figs. 3, 4). In both TokK structures, Met binds to the unique iron of the [4Fe–4S] cluster, and the position of the 5′-carbon of 5′-dAH suggests that the binding of SAM in TokK is almost identical to that in OxsB, but quite different from that observed in the TsrM from *Kitasatospora setae* (*Ks*TsrM) (Extended Data Fig. 4d). A third, C-terminal domain is distinct to TokK[12,15] (Fig 2a, Extended Data Fig. 5). Although as-isolated TokK contains MeCbl, hydroxycobalamin (OHCbl) and adenosylcobalamin (AdoCbl), only OHCbl is observed bound to the N-terminal domain in the X-ray crystal structures. This assignment was confirmed by high-resolution mass spectrometry of dissolved crystals (Supplementary Fig. 3). In the structure of the TokK–OHCbl–5′-dAH–Met–substrate complex, which mimics the complex immediately before reaction with substrate, we observed clear $F_o - F_c$ electron density consistent with the shape and size of **1** in one of two monomers in the asymmetric unit (Fig. 2b). In the second monomer, this electron density was also present, but it was of insufficient intensity for substrate modelling (Supplementary Fig. 4). In the chain with substrate bound, the pantetheine tail of **1** occupies a channel that leads from the surface of the protein into the active site (Fig. 2a, Supplementary Fig. 5). This cavity is formed at the interface of all three domains of TokK. The N-terminal Cbl-binding domain and unique C-terminal domain contribute most interactions with the pantetheine unit (Fig. 2a, Extended Data Fig. 6). These include hydrophobic contacts, water-mediated H-bonding interactions and direct polar contacts. For example, Asn515 in the C-terminal domain H-bonds to the terminal -OH of the pantetheine moiety, suggesting that this domain has a key role in substrate recognition. In the N-terminal domain, the Cbl cofactor itself participates in a water-mediated contact to an amide carbonyl of **1**. This network involves one of the Cbl propionamide substituents and Tyr410 of the RS domain. The β-lactam ring of **1** is buried deep within the RS domain and anchored by direct and water-mediated contacts to the C7 carbonyl and C3 carboxylate substituents. This mode of β-lactam interaction resembles that of carbapenem synthase, the enzyme responsible for inversion of stereochemistry at C5 in simple carbapenems[17] (Extended Data Fig. 7). Both enzymes share the use of H-atom abstraction chemistry to selectively target an unactivated C–H bond within the bicyclic β-lactam core, consistent with their conserved substrate-anchoring strategies. The C3 carboxylate of **1** H-bonds to Arg280 in the RS domain and Tyr652 in the C-terminal domain. These side chains move considerably from their positions in the structure without substrate bound (Fig. 2a), and substitution of Arg280 with Gln results in near complete loss of activity (Fig. 2c), confirming the importance of these side chains in substrate binding.

The interactions between TokK and substrate **1** position the β-lactam appropriately both for activation of C6 by 5′-dA• and for subsequent methyl addition by the Cbl cofactor[18] (Fig. 2d). C6 of **1** is located directly in front of the 5′-carbon of 5′-dAH, 3.7 Å away, like other RS enzyme–substrate complexes that initiate reactions by H-atom abstraction. The orientation of these two groups in the structure does not reveal whether the pro-*R* or pro-*S* H-atom is removed from C6 of **1** by 5′-dA•, as these substituents project equally above and below the 5′-carbon of 5′-dAH (Fig. 2d). However, the structure does provide insight into the trajectory of methyl addition. C6 of **1** is located 4.2 Å above the axial ligand of Cbl (Fig. 2d) at an angle of around 85° relative to the 5′-carbon of 5′-dAH. This arrangement suggests that the methyl group adds to the bottom face of the β-lactam ring, consistent with the absolute configurations observed in the TokK products and thienamycin[18,19]. The distance and orientation of reactant functional groups in TokK also compares favourably to other enzymes that catalyse radical-mediated activation and

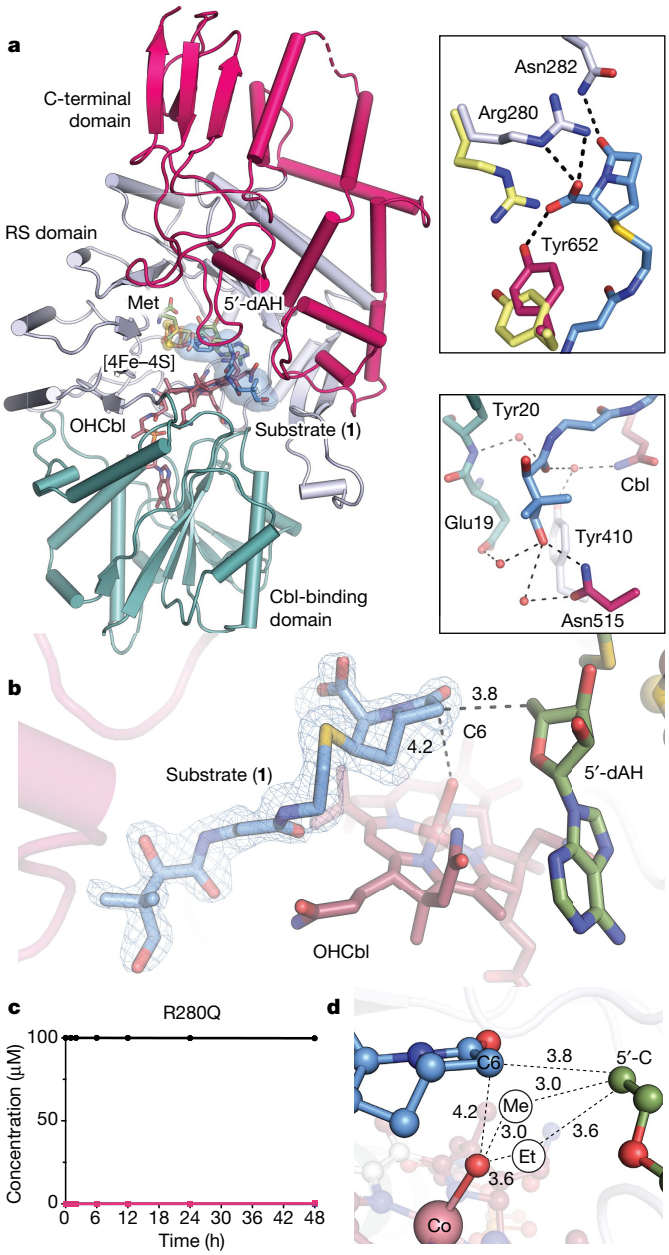


Fig. 2 | TokK binds its carbapenam substrate at the interface of three domains. a, The overall structure of TokK (chain A) is illustrated as a ribbon diagram and coloured by domain. The Cbl-binding Rossmann fold is shown in teal with the Cbl cofactor in stick format and coloured by atom type. The RS domain is shown in light blue. A [4Fe–4S] cluster is shown in orange and yellow spheres. 5′-dAH and Met coproducts are shown in stick format. The C-terminal domain is shown in pink. Carbapenam substrate (**1**) is shown in light blue sticks, coloured by atom type. **b**, An $F_o - F_c$ omit electron density map is shown for **1** (blue mesh, contoured at 3.0σ). Substrates, cofactors and coproducts are shown in stick format. Distances between reactive groups are given in units of Å. **c**, Product formation for the R280Q TokK variant performed in triplicate (each replicate shown as a symbol). Substrate is shown in black spheres and the first methylated product is shown in pink squares. All activity assays were conducted with substrate **1**, shown in Fig. 1a. **d**, Projections of the additional methyl groups added to C6 of **1** and their respective distances (in Å) from the 5′ carbon of 5′-dAH and the hydroxyl moiety of OHCbl, the latter of which serves as a surrogate of the active MeCbl cofactor. Spheres labelled Me and Et represent the suggested positions of the newly installed carbon atoms in the mono-methylated and dimethylated products.

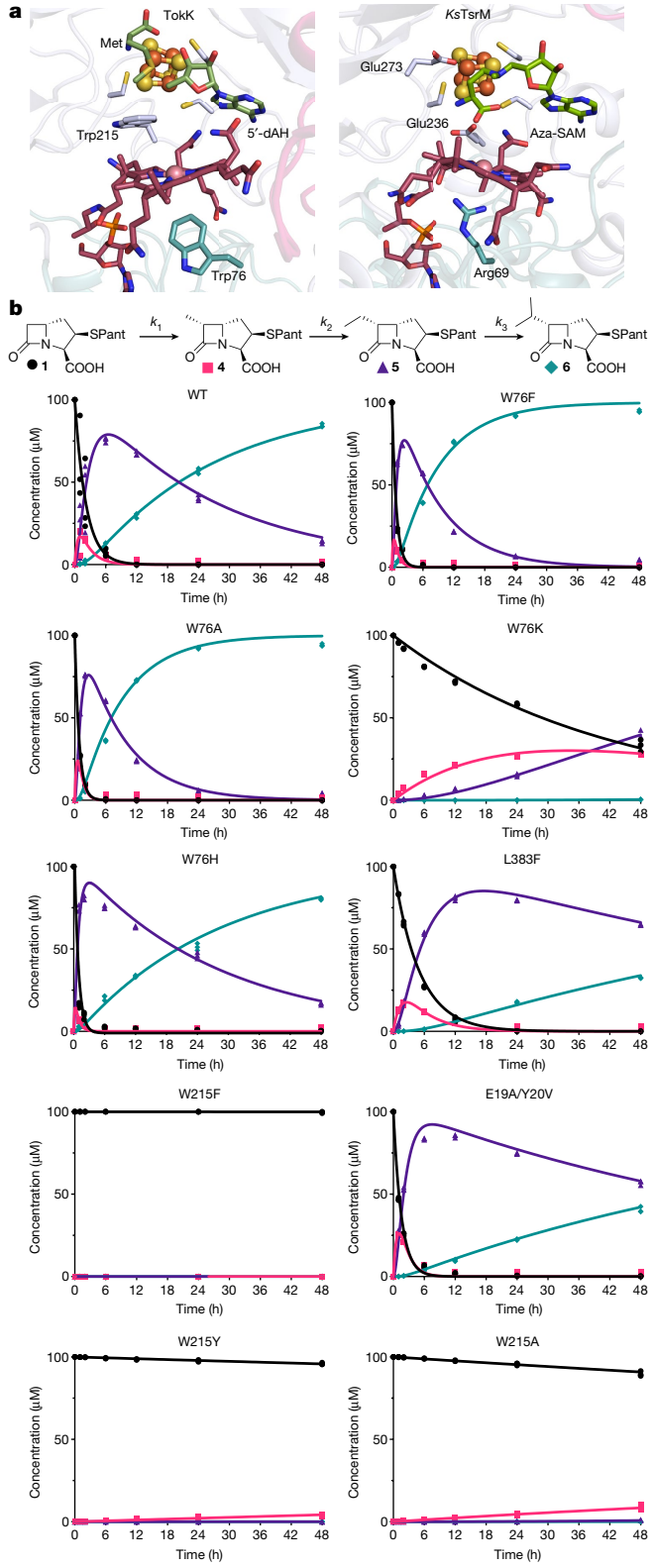


Fig. 3 | The Cbl- and substrate-binding sites influence overall activity and the relative rates of each TokK methylation step. a, Comparative analysis of the side chains proximal to the Co ion in two Cbl-binding RS enzymes, TokK and *Ks*TsrM (Protein Data Bank (PDB) ID: 6WTF). Selected amino acid side chains are shown in stick format and the Co ion is shown as a pink sphere. **b**, Top, schematic of the three sequential methylations performed by TokK (SPant, pantetheine (Fig. 1a). Forty-eight-hour time-course experiments performed in triplicate (each replicate represented as a symbol), tracking the product formation of substrate (black spheres), methyl (pink squares), ethyl (purple triangles) and isopropyl (blue diamonds). Bottom, each product was estimated using COPASI (irreversible mass action model using the reaction scheme shown above) and simulated using Virtual Cell (shown as lines) for wild type (WT), W76F, W76A, W76K, W76H, L383F, W215F, E19A/Y20V, W215Y and W215A.

functionalization of a substrate C–H bond, such as iron-dependent hydroxylases in the cytochrome P450 and iron(II)-oxo-glutarate-(Fe-2OG)-dependent superfamilies (Extended Data Fig. 7). These systems orient their reactive groups similarly, but over a slightly shorter distance range[20,21]. This comparison highlights an important distinction between RSMTs and other radical functionalization enzymes. In P450s and Fe-2OG enzymes, a single reactive entity—a high-valent iron(IV)-oxo or iron(III)-hydroxo group—must both activate substrate and functionalize it. This strategy is inherently limiting because the enzyme can only activate and functionalize substrate from the same side. The Cbl-dependent RS radical functionalization platform is more versatile because the radical activation step is separated from methylation, which allows for more diverse stereochemical outcomes.

Notably, the structure of TokK in complex with the substrate has marked structural similarities to another well-characterized RS methylase that does not rely on MeCbl, RlmN (ref. [22]) (Extended Data Fig. 8). RlmN uses an *S*-methyl cysteinyl (methylCys) residue as an intermediate methyl carrier during the methylation of the sp^2-hybridized C2 atoms of adenosine 2503 in ribosomal RNA and adenosine 37 in several transfer RNAs (tRNAs). When the structure of RlmN crosslinked to an *Escherichia coli* $tRNA^{Glu}$ substrate is compared to that of substrate-bound TokK, the methylCys residue in the RlmN structure is in a position similar to that of the hydroxyl group of OHCbl in the TokK structure. Moreover, their respective substrates occupy similar positions in the active site (Extended Data Fig. 8). Although the catalytic mechanisms of these two enzymes are distinct, both obey a ping-pong kinetic model, in which one SAM molecule is used to methylate the intermediate methyl carrier, while a second SAM molecule is used to generate a 5′-dA•.

Cbl is multifunctional in TokK, mediating both the polar methylation of Co(I) by SAM and the transfer of a methyl radical to C6 of the substrate. It is bound at the interface of the Cbl and RS domains with its dimethylbenzimidazole base tucked into the Rossmann fold of the N-terminal domain, a conformation termed 'base-off' (Fig. 3a, Extended Data Fig. 5). OxsB and TsrM use a similar base-off approach to interact with their Cbl cofactors (Fig. 3a, Extended Data Fig. 4), a binding mode that allows for extensive modulation of the reactivity of the Co(III) ion of MeCbl by the local protein environment[23]. In TsrM, this structural feature is essential for the atypical polar methylation of its substrate, Trp, which requires heterolytic cleavage of the Co(III)-carbon bond of MeCbl. The bottom face of the Co ion in TsrM is adjacent to Arg69 but not directly coordinated, which is likely to promote nucleophilic attack of MeCbl by Trp by blocking coordination of a sixth ligand and destabilizing the Co(III)–C bond owing to charge–charge repulsion[24–26] (Fig. 3a). In TokK, a different side chain, Trp76, occupies the lower axial face of the Cbl, residing 3.8 Å from the metal ion (Fig. 3a). Neither Arg69 nor Trp76 lie in a canonical DX**H**XXG motif exemplified by methionine synthase, wherein the His residue in the motif ligates to the Co, although both are found in the loop following β3 in the Rossmann fold. To investigate the role of this residue, Trp76 was substituted with Phe and Ala. Rates of methylation slightly increased for both site-specific substitutions, and analysis by electron paramagnetic resonance (EPR) spectroscopy suggested that both variants exhibited the same four-coordinate geometry as wild-type TokK (Supplementary Fig. 6). These data suggest that even when the

steric bulk of Trp76 is reduced, water does not coordinate the Cbl (Extended Data Fig. 9). To perturb the local environment of the Cbl cofactor further, Trp76 was also substituted with His and Lys. The activity of Trp76His TokK resembles the activities of the Phe and Ala replacements, but the activity of Trp76Lys TokK was reduced by a factor of around 50 for all three methylation steps (Fig. 3b). The substitution tolerance of Trp76 in TokK contrasts with that of Arg69 of TsrM, which when substituted with Lys was unable to transfer a methyl group from MeCbl to substrate[12]. Although Trp76 is not widely conserved among other well-characterized Cbl-dependent RS methylases, it is found in the same sequence context in the CysS primary structure (Extended Data Fig. 1).

The structure of **1** bound to TokK also rationalizes established differences in rate constants for each of the three methyl transfers catalysed by this enzyme (Fig. 3b). The second methylation to form the ethyl-containing carbapenam product **5** proceeds at least threefold faster than the formation of **4**—a pattern that runs counter to known differences in the reactivity of secondary and primary C–H bonds. Although we do not report a structure containing the singly methylated intermediate **4**, if we presume that **4** remains anchored to Arg280, the methyl group at C6 would be positioned closer to the Cbl axial ligand and potentially at a more optimal angle than the original C6 C–H target. A third methylation to form the isopropyl carbapenem product **6** requires hydrogen atom abstraction from the same carbon, but the newly added ethyl carbon restricts the population of conformers accessible to 5′-dA•. This steric demand could help to explain why the estimated first-order rate constant for the third methylation, k_3, is slower than the first two methyl transfers, k_1 and k_2. The buried location of **1**, 5′-dAH, Met and the Cbl cofactor suggests that dissociation of the methylated carbapenam products must occur before dissociation of the SAM cleavage products. This arrangement is consistent with the non-processive kinetic model used to fit the time-course kinetics of each TokK methylation[18]. A similar mechanism was proposed for CysS (ref. [7]) (Extended Data Fig. 1). Although CysS is only 29% identical both to TokK and to ThnK, all three proteins potentially contain a Trp side chain adjacent both to the Cbl and to the substrate (Extended Data Fig. 1). Substitution of Trp215 with Phe, Ala, or Tyr markedly slows substrate methylation by TokK, which suggests that it could have a role in catalysis (Fig. 3b).

ThnK and TokK share 79.3% sequence identity and act on the same substrate, (2*R*)-pantetheinylated carbapenam (Supplementary Fig. 7), but ThnK performs two sequential methylations whereas TokK catalyses three[18,27]. Nearly all residues in proximity to the active site are identical in the two orthologues. However, three non-conserved amino acids near the active site were examined to determine their role in controlling the extent of methylation. Leu383 is positioned deep in the active site and near 5′-dAH (Supplementary Fig. 7). When this residue is substituted with Phe, which is found at the same position in the primary structure of ThnK (Supplementary Fig. 7), the rate constants for all three methyl transfers are reduced (2.3-, 2.5- and 4.5-fold for k_1, k_2 and k_3, respectively) compared to those of wild-type TokK (Fig. 3b). Two adjacent residues at the entrance to the pantetheine-binding tunnel, Glu19 and Tyr20, (Fig. 2a, Supplementary Fig. 7) were replaced with the cognate residues in ThnK to generate an E19A/Y20V double substitution. In this variant, the rate constant for the first methylation is increased 1.4-fold compared to that of the wild type, and the rate constants for the second and third methylations are decreased 1.4- and 3.4-fold, respectively, therefore shifting the kinetic profile closer to the pattern observed with ThnK ($k_1 > k_2$, $k_3 = 0$) (ref. [18]) (Fig. 3b).

The structure of TokK solved in the absence of substrate reveals very few differences in overall fold or domain organization relative to the TokK–**1** complex (root mean square deviation (RMSD) of 0.53 Å over 603 residues by Cα atoms). The substrate-binding channel, located at the interface of the three domains of TokK, remains intact without substrate with only modest alterations in size caused by the aforementioned conformational changes in the side chains of Arg280 and Tyr652 (Fig. 2a, Supplementary Fig. 5). The preformed nature of the substrate-binding tunnel in TokK contrasts with observations from structures of TsrM in the absence and presence of its substrate, in which a loop from the C terminus moves to cap the active site in the presence of Trp[12]. Although the C-terminal domain of TokK is considerably larger than that of TsrM, these domains appear to share a common role in substrate interaction (Extended Data Fig. 4).

The overwhelming majority of known Cbl-dependent RSMTs operate by the radical mechanism used by TokK, producing an equivalent of 5′-dAH and SAH for each methylated product molecule. The structure of the TokK active site reveals a scaffold for positioning the cofactors responsible for substrate activation and methyl transfer and is consistent with the non-processive mechanism of sequential methylations observed for the enzyme, which requires the release of each partially alkylated intermediate and both SAM coproducts before reloading the active site for subsequent methylation. Moreover, the structure reveals that there is little active participation in catalysis from other amino acids in the active site, apart from substrate or cofactor binding. This scaffolding approach, which is underscored by the notable absence of conformational changes after substrate binding, may be shared by the only other Cbl-dependent RS enzyme that has, to our knowledge, been structurally characterized in complex with its substrate, TsrM. Although the electrophilic substitution mechanism used by TsrM requires a general base to accept the N1 proton of the indole ring of Trp during catalysis, biochemical studies and models of the functional enzyme–substrate complex suggest that the carboxylate group of a cosubstrate, SAM, functions in this capacity instead of an active-site amino acid. Elucidation of additional structures and mechanisms for Cbl-dependent RS enzymes will reveal how a seemingly common approach for catalysis may be further elaborated within this large and diverse group of enzymes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04392-4.

1. US Food and Drug Administration. Supplemental New Drug Approval: RECARBRIO (imipenem, cilastatin, and relebactam) for injection, https://www.fda.gov/RegulatoryInformation/Guidances/default.htm (2020).
2. Fisher, J. F. & Mobashery, S. Three decades of the class A β-lactamase acyl-enzyme. *Curr. Protein Pept. Sci.* **10**, 401–407 (2009).
3. Sinner, E. K., Marous, D. R. & Townsend, C. A. Evolution of methods for the study of cobalamin-dependent radical SAM enzymes. *ACS Bio. Med. Chem. Au* https://doi.org/10.1021/acsbiomedchemau.1c00032 (2021).
4. Li, R., Lloyd, E. P., Moshos, K. A. & Townsend, C. A. Identification and characterization of the carbapenem MM 4550 and its gene cluster in *Streptomyces argenteolus* ATCC 11009. *ChemBioChem* **15**, 320–331 (2014).
5. Nunez, L. E., Mendez, C., Brana, A. F., Blanco, G. & Salas, J. A. The biosynthetic gene cluster for the β-lactam carbapenem thienamycin in *Streptomyces cattleya*. *Chem. Biol.* **10**, 301–311 (2003).
6. Coulthurst, S. J., Barnard, A. M. & Salmond, G. P. Regulation and biosynthesis of carbapenem antibiotics in bacteria. *Nat. Rev. Microbiol.* **3**, 295–306 (2005).
7. Wang, Y., Schnell, B., Baumann, S., Muller, R. & Begley, T. P. Biosynthesis of branched alkoxy groups: iterative methyl group alkylation by a cobalamin-dependent radical SAM enzyme. *J. Am. Chem. Soc.* **139**, 1742–1745 (2017).
8. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical *S*-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
9. Holliday, G. L. et al. Atlas of the radical SAM superfamily: divergent evolution of function using a "plug and play" domain. *Methods Enzymol.* **606**, 1–71 (2018).
10. Wang, S. C. Cobalamin-dependent radical *S*-adenosyl-l-methionine enzymes in natural product biosynthesis. *Nat. Prod. Rep.* **35**, 707–720 (2018).
11. Miller, D. V. et al. in *Comprehensive Natural Products III: Chemistry and Biology* vol. 5 (eds Liu, H.-W. & Begley, T.) 24–69 (Elsevier, 2020).
12. Knox, H. L. et al. Structural basis for non-radical catalysis by TsrM, a radical SAM methylase. *Nat. Chem. Biol.* **17**, 485–491 (2021).

13. Pierre, S. et al. Thiostrepton tryptophan methyltransferase expands the chemistry of radical SAM enzymes. *Nat. Chem. Biol.* **8**, 957–959 (2012).
14. Blaszczyk, A. J. et al. Spectroscopic and electrochemical characterization of the iron-sulfur and cobalamin cofactors of TsrM, an unusual radical *S*-adenosylmethionine methylase. *J. Am. Chem. Soc.* **138**, 3416–3426 (2016).
15. Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L. & Liu, H. W. A B12-dependent radical SAM enzyme involved in oxetanocin A biosynthesis. *Nature* **544**, 322–326 (2017).
16. Blaszczyk, A. J., Wang, B., Silakov, A., Ho, J. V. & Booker, S. J. Efficient methylation of C2 in L-tryptophan by the cobalamin-dependent radical *S*-adenosylmethionine methylase TsrM requires an unmodified N1 amine. *J. Biol. Chem.* **292**, 15456–15467 (2017).
17. Chang, W.-C. et al. Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science* **343**, 1140–1144 (2014).
18. Sinner, E. K., Lichstrahl, M. S., Li, R., Marous, D. R. & Townsend, C. A. Methylations in complex carbapenem biosynthesis are catalyzed by a single cobalamin-dependent radical *S*-adenosylmethionine enzyme. *Chem. Commun.* **55**, 14934–14937 (2019).
19. Albers-Schoenberg, G. A. et al. Structure and absolute configuration of thienamycin. *J. Am. Chem. Soc.* **100**, 6491–6499 (1978).
20. Mitchell, A. J. et al. Visualizing the reaction cycle in an iron(II)- and 2-(oxo)-glutarate-dependent hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
21. Poulos, T. L., Finzel, B. C. & Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **195**, 687–700 (1987).
22. Schwalm, E. L., Grove, T. L., Booker, S. J. & Boal, A. K. Crystallographic capture of a radical *S*-adenosylmethionine enzyme in the act of modifying tRNA. *Science* **352**, 309–312 (2016).
23. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. How a protein binds B-12—a 3.0-Å X-ray structure of B-12-binding domains of methionine synthase. *Science* **266**, 1669–1674 (1994).
24. Kräutler, B. Thermodynamic *trans*-effects of the nucleotide base in the B12 coenzymes. *Helv. Chim. Acta* **70**, 1268–1278 (1987).
25. Krautler, B. Chemistry of methylcorrinoids related to their roles in bacterial C1 metabolism. *FEMS Microbiol. Rev.* **7**, 349–354 (1990).
26. Milton, P. A. & Brown, T. L. The kinetics of benzimidazole dissociation in methylcobalamin. *J. Am. Chem. Soc.* **99**, 1390–1396 (1977).
27. Marous, D. R. et al. Consecutive radical *S*-adenosylmethionine methylations form the ethyl side chain in thienamycin biosynthesis. *Proc. Natl Acad. Sci USA* **112**, 10354–10358 (2015).

Methods

Overexpression and purification of TokK from *Streptomyces tokunonesis*

TokK (Uniprot ID: A0A6B9HEI0) was produced heterologously in *Escherichia coli* BL21 (DE3) by overexpression from a pET29b:*tokK*Tev construct[18]. To facilitate production of soluble TokK protein with maximal occupancy of [4Fe–4S] cluster and Cbl cofactors, the strain was transformed with two additional plasmids, pDB1282 and pBAD42-BtuCEDFB[18,28–30]. A 100-ml LB starter culture with 50 μg ml^{-1} kanamycin (pET29b) containing *tokK*, 50 μg ml^{-1} spectinomycin (pBAD42-BtuCEDFB) and 100 μg ml^{-1} ampicillin (pDB1282) was inoculated from a single colony and incubated for 18 h at 37 °C while shaking at 250 rpm. A 12-ml aliquot of the starter culture was used to inoculate a 4-l culture of LB medium supplemented with OHCbl (1.3 μM) and allowed to shake at 180 rpm at 37 °C. Four of these cultures were grown to an optical density at 600 nm ($OD_{600\,nm}$) of 0.3, at which point induction of genes on pDB1282 and pBAD42-BtuCEDFB was initiated with the addition of arabinose to a final concentration of 0.2%. To facilitate iron–sulfur cluster incorporation, 25 μM $FeCl_3$ and 150 μM cysteine were added at the initiation of induction with arabinose. The cultures were then grown to an $OD_{600\,nm}$ of 0.6 followed by incubation in an ice-water bath for 1 h. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cultures were incubated at 18 °C for an additional 18 h before the cells were collected by centrifugation at 6,000*g*. The resulting cell paste (around 50 g) was flash-frozen in liquid N_2 and stored in liquid N_2 before protein purification.

All purification steps and subsequent manipulations of TokK were performed in a Coy Laboratory Products vinyl anaerobic chamber. Cell paste was resuspended in 100 ml lysis buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 10% glycerol, 5 mM imidazole and 10 mM β-mercaptoethanol (BME)). The cell suspension was incubated with lysozyme (1 mg ml^{-1}), DNase (0.1 mg ml^{-1}) and phenylmethylsulfonyl fluoride (PMSF) (0.18 mg ml^{-1}) for 30 min at room temperature and then cooled to 4 °C (refs. [18,30]). Cells were then lysed by sonic disruption (70% amplitude, 45 s on, 59 s off, around 15 min) and then centrifuged at 50,000*g* for 1 h to separate insoluble material. The resulting supernatant was loaded onto a pre-equilibrated column of Ni-NTA resin and purified by immobilized-metal affinity chromatography. The protein-loaded resin was washed with 50 ml lysis buffer before the addition of elution buffer (lysis buffer supplemented with 500 mM imidazole). Dark-coloured elution fractions were pooled and concentrated to 1 ml in an Amicon 10 kDa MWCO ultrafiltration device (EMD Millipore). The protein fractions were then exchanged into TEV protease cleavage buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 15% glycerol and 10 mM BME) using a PD-10 pre-poured gel-filtration column from GE Biosciences. The exchanged protein sample was allowed to react with 10 units of TEV protease (around 2 mg ml^{-1}) in 25 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM TCEP and 50% glycerol (Millipore Sigma) for two days on ice to generate TokK containing only six additional amino acids (ENLYFQ) on its C terminus. On the second day, the [4Fe–4S] cluster and Cbl cofactors were reconstituted as previously described[30]. The reaction mixture was then reapplied to the Ni-NTA resin to capture any remaining His-tagged protein, and purified TokK-ENLYFQ was collected in the flow-through fraction. TokK was concentrated and exchanged into gel-filtration buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 1 mM DTT and 15% glycerol) for size-exclusion chromatography. In this step, the protein was applied to a HiPrep 16/60 S200 column using an ÄKTA fast protein liquid chromatography (FPLC) system (GE Biosciences) housed in the anaerobic chamber. TokK elutes as an apparent monomer. Fractions were pooled on the basis of UV-vis absorption at 280 and 410 nm and concentrated to 38 mg ml^{-1}.

Synthesis of substrate

Synthesis of the TokK substrate, (2*R*,3*R*,5*R*)-3-((2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)thio)-7-oxo-1-azabicyclo[3.2.0]heptane-2-carboxylic acid (**1**), was carried out as previously described[27]. Characterization matched that previously reported.

Determination of the X-ray crystal structure of TokK

General crystallographic methods. X-ray diffraction datasets were collected at the General Medical Sciences and Cancer Institutes Collaborative Access Team (GM/CA-CAT) and at the Advanced Photon Source, Argonne National Laboratory and Berkeley Center for Structural Biology (BCSB) beamlines at the Advanced Light Source at Lawrence Berkeley National Laboratory. All datasets were processed using the HKL2000 or HKL3000 package, and structures were determined by single anomalous dispersion (SAD) phasing using Autosol/HySS or by molecular replacement using the program PHASER[31–34]. Model building and refinement were performed with Coot and phenix.refine[31,35]. Figures were prepared using PyMOL[36,37]. Substrate channel figures were prepared using Hollow[38]. Ligplot was used to visualize the binding of substrate **1** (ref. [39]).

Crystallization and structure solution of 5′-dAH–Met–TokK. Purified TokK protein aliquots were diluted to 8 mg ml^{-1} in 34 mM HEPES, pH 7.5. Then 5′-dAH and Met (Millipore Sigma) were added to the resulting solution to final concentrations of 2 mM each. The mixture was incubated for 30 min at room temperature. In hanging-drop vapour-diffusion trials with 100 mM magnesium chloride, 100 mM calcium chloride, 20% PEG 8000 and 10% 1,6-hexanediol as the precipitating reagent, brown plate-shaped crystals appeared within three days. Trials were initiated by adding 1 μl of protein with 1 μl precipitating solution, followed by equilibration against 500 μl of a 0.5 M LiCl well solution at room temperature. Crystals were prepared for data collection by mounting on rayon loops followed by a brief soak in cryoprotectant solution (50% (v/v) precipitating reagent and 50% (v/v) ethylene glycol) and flash-freezing in liquid nitrogen.

Diffraction datasets for single-wavelength anomalous diffraction phasing were collected at the iron K-edge X-ray absorption peak (λ = 1.72194 Å) with 360° of data measured using a 0.5° oscillation range to 2.52 Å resolution. In addition, a 1.79 Å-resolution native dataset was collected at λ = 1.03313 Å (Extended Data Table 1). Heavy-atom sites were identified using HySS implemented within Phenix Autosol[31]. The initial overall figure-of-merit (FOM) was 0.265 and the Bayes CC was 18.2 (ref. [31]). Phenix Autobuild was used to generate an initial model of 541 residues out of 687 in chain A and 569 residues out of 687 in chain B with R_{work}/R_{free} of 0.25/0.31. Iterative manual model building and refinement were performed in Coot and Phenix[35]. This model was used to obtain phase information for the 1.79 Å-resolution native dataset by molecular replacement using Phenix Phaser-MR[31]. Geometric restraints for 5′-dAH and Cbl were obtained from the Grade Web Server (Global Phasing). R_{free} flags were maintained so that the same 5% of the reflections were used as the test set. The final model consists of residues 9–412, 416–672, one [4Fe–4S] cluster, one OHCbl cofactor, one Met and one 5′-dAH in chain A; residues 7–413, 416–672, one [4Fe–4S] cluster, one OHCbl cofactor, one Met and one 5′-dAH in chain B. The final model also contains 2 chloride ions, 2 potassium ions, 21 molecules of ethylene glycol and 887 waters. The Ramachandran plot shows that 97.4% residues are in favoured regions with the remaining 2.6% in allowed regions[36]. Data collection and refinement statistics are shown in Extended Data Table 1.

Crystallization and structure solution of substrate-bound 5′-dAH–Met–TokK. Purified TokK was diluted to 8 mg ml^{-1} TokK in 34 mM HEPES, pH 7.5. Then 5′-dAH, Met and substrate were added to final concentrations of 2 mM each. The solution was incubated for 15 min at room temperature. In hanging-drop vapour diffusion crystallization trials with 0.2 M lithium sulfate, 0.1 M Tris-HCl, pH 8.5 and 18% PEG 8000 as the precipitant, brown plate-shaped crystals appeared within two weeks. Trials were initiated by mixing 1 μl of protein and 1 μl of precipitant

followed by equilibration at room temperature against a 500-µl reservoir of the precipitating solution. Before looping, the concentration of substrate **1** was increased to 2.7 mM. Crystals were prepared for data collection by mounting on rayon loops followed by a brief soak in perfluoropolyether oil (Hampton Research) for cryoprotection and flash-freezing in liquid nitrogen.

The structure containing substrate was solved by molecular replacement (Phaser-MR in Phenix) using the coordinates of the 5′-dAH–Met–TokK as the search model. Iterative manual model building and refinement were performed in Coot and Phenix[31,35]. Initial geometric restraints for (2*R*)-pantetheinylated carbapenem substrate were generated by eLBOW in Phenix[31]. Final geometric restraints for the carbapenem substrate were created by using the PRODRG 2 server[40]. The final model consists of residues 8–566, 571–672, one [4Fe–4S] cluster, one OHCbl cofactor, one Met, one 5′-deoxyadenosine (5′-dAH) and one (2*R*)-pantetheinylated carbapenem substrate in chain A; residues 9–672, one [4Fe–4S] cluster, one OHCbl factor, one Met and one 5′-dAH in chain B. The final structure also contains 2 glycerol molecules, 4 potassium ions and 1,193 waters. The Ramachandran plot showed that 97.7% of residues are in favoured regions with the remaining 2.3% in allowed regions[36]. Data collection and refinement statistics are shown in Extended Data Table 1.

Cbl ligand assignment in protein crystals

The protocol was adapted from previously published studies[28]. Approximately 15 TokK + 5′-dAH + Met crystals were looped from the crystallization drop into 40 µl of mother solution in a darkly coloured Eppendorf tube. In the dark, 50 mM H_2SO_4 was added to the tube. The tube was vortexed and then centrifuged for 20 min. Five microlitres of this solution was injected into a Thermo Fisher Scientific UHPLC/QExactive HF-X mass spectrometer equipped with a C18 column (2.1 × 100 mm) equilibrated in 5% solvent A (0.1% formic acid) and 95% solvent B (0.1% formic acid in acetonitrile). The solvent B composition was increased to 98% from 1 to 7 min. Cbl forms were detected by ESI^+, scanning from *m/z* 150 to 1,700 with a resolution of 120,000. A calibration curve (0.1 µM–5 µM) of Cbl standards was run concurrently to quantify the Cbl forms in the sample.

Generation of SSNs

A pool of annotated Cbl-dependent RS enzymes was generated by merging representative annotated B_{12}-dependent RS enzyme sequences from radicalsam.org (megacluster 2-1, subgroup 5; https://radicalsam.org/explore.php?id=cluster-2-1&v=3.0, accessed March 2021) with those from the USCF structure-function linkage database (SFLD) (http://sfld.rbvi.ucsf.edu/archive/django/index.html, accessed April 2021)[41,42]. Additional sequences were included for the following functionally characterized enzymes (with Uniprot accession codes): *Sl*TsrM (C0JRZ9), *Ks*TsrM (E4N8S5), GenK (Q70KE5), GenD1 (Q2MG55), PhpK (A0A0M3N271), Fom3 (Q56184), CysS (A0A0H4NV78), OxsB (O24770), PoyC (J9ZXD6), swb7 (D2KTX6), ArgMT (Q8THG6), ThnK (F8JND9), TokK (A0A6B9HEI0), BchE (Q7X2C7), CouN6 (A0A1H2F7M3) and CloN6 (Q9F8U1)[7,12,14,15,18,43–51]. The pool of B_{12}-dependent RS enzyme sequences from radicalsam.org contains 9,724 representative entries that reflect 53,470 unique sequences. The class B and B_{12}-binding RS enzyme sequence groups from the SFLD contain 1,525 and 5,920 representatives, respectively. Each group reflects a total pool of 4,232 and 15,983 unique sequences, respectively[41]. The final representative sequence set used for SSN generation contained approximately 11,000 sequences after removal of duplicates. This dataset represents a larger sequence pool of more than 50,000 unique entries.

The Enzyme Function Initiative enzyme similarity tool (EFI-EST) (https://efi.igb.illinois.edu) was used to perform an all-by-all BLAST analysis of the representative sequence dataset described above to create an initial SSN[42,52–54] with an alignment score threshold of 65. To eliminate protein fragments, sequence length was restricted to greater than 300 amino acids. The final SSN (Fig. 1c, Supplementary Fig. 1) is depicted as a representative node network in which each node reflects sequences with more than 40% identity. All networks were visualized with the Organic layout in Cytoscape[55]. The SSN in Fig. 1c represents 36 sequence clusters extracted from the full SSN shown in Supplementary Fig. 1. The clusters in Fig. 1c were selected on the basis of the number of nodes or the presence of functionally annotated or structurally characterized sequences.

Preparation of EPR samples

TokK was diluted to a final concentration of approximately 250 µM for each EPR sample. The samples were photolysed on ice for 45 min to generate the cob(II)alamin state. All samples were flash-frozen in cryogenic liquid isopentane in an anaerobic chamber. The resulting samples were stored in liquid nitrogen before analysis. EPR measurements were taken on a Magnettech 5000 x-band ESR spectrometer equipped with an ER 4102ST resonator. Temperature was controlled by an ER 4112-HV Oxford Instruments variable-temperature helium-flow cryostat. All measurements were taken at 70 K, with 1 mT modulation amplitude and 1 mW power.

Construction of TokK variants

TokK variants were generated by overlap extension PCR using pET29b:*tokK*Tev as a template and the primers described in Supplementary Table 1. TokK_F_s was used as the forward primer for all constructs except the E19A/Y20V double variant, for which TokK_F_EYmut was used instead. After amplification, PCR products were digested with NdeI and XhoI and ligated into a similarly digested pET29b vector. Sequence-verified constructs were used to transform *E. coli* BL21 (DE3) along with helper plasmids pDB1282 and pBAD42-BtuCEDFB as described above for overexpression.

Methylation assays of TokK variants

Expression and purification of TokK variants was carried out as previously described for wild-type TokK[18], except reconstitution[30] was done concurrently with overnight TEV protease cleavage, and an additional buffer exchange was done using an Econo-Pac 10DG column (Bio-Rad) to remove excess reconstitution reagents. Methylation assays were carried out in triplicate and contained 100 mM HEPES, pH 7.5, 200 mM KCl, 1 mM SAM, 1 mM methyl viologen, 2 mM NADPH, 0.5 mM MeCbl, 100 µM substrate and 100 µM enzyme. At each time point, an aliquot of the reaction mixture was diluted 5×, filtered through a 10-kDa Amicon ultrafiltration device and analysed for product formation using ultraperformance liquid chromatography-high resolution mass spectrometry (UPLC-HRMS) as previously described[18]. First-order rate constants were determined using the COPASI parameter estimation tool[56], and curves were simulated using Vcell[57].

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Atomic coordinates and structure factors for the reported crystal structures in this work have been deposited to the Protein Data Bank (PDB) under accession numbers 7KDX (structure with 5′-dAH + Met) and 7KDY (structure with 5′-dAH + Met + carbapenam substrate).

28. Lanz, N. D. et al. Enhanced solubilization of class B radical *S*-adenosylmethionine methylases by improved cobalamin uptake in *Escherichia coli*. *Biochemistry* **57**, 1475–1490 (2018).
29. Lanz, N. D. et al. RlmN and AtsB as models for the overproduction and characterization radical SAM Proteins. *Methods Enzymol.* **516**, 125–152 (2012).
30. Blaszczyk, A. J., Wang, R. X. & Booker, S. J. TsrM as a model for purifying and characterizing cobalamin-dependent radical *S*-adenosylmethionine methylases. *Methods Enzymol.* **595**, 303–329 (2017).

31. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
32. Bunkoczi, G. et al. Phaser.MRage: automated molecular replacement. *Acta Crystallogr. D* **69**, 2276–2286 (2013).
33. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
36. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
37. The PyMOL Molecular Graphics Systems v.2.0 (Schrödinger).
38. Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).
39. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
40. Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).
41. Akiva, E. et al. The Structure–Function Linkage Database. *Nucleic Acids Res.* **42**, D521–D530 (2014).
42. Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* **56**, 4293–4308 (2017).
43. Kim, H. J. et al. GenK-catalyzed C-6′ methylation in the biosynthesis of gentamicin: isolation and characterization of a cobalamin-dependent radical SAM enzyme. *J. Am. Chem. Soc.* **135**, 8093–8096 (2013).
44. Huang, C. et al. Delineating the biosynthesis of gentamicin x2, the common precursor of the gentamicin C antibiotic complex. *Chem. Biol.* **22**, 251–261 (2015).
45. Werner, W. J. et al. In vitro phosphinate methylation by PhpK from *Kitasatospora phosalacinea*. *Biochemistry* **50**, 8986–8988 (2011).
46. Wang, B. et al. Stereochemical and mechanistic investigation of the reaction catalyzed by Fom3 from *Streptomyces fradiae*, a cobalamin-dependent radical *S*-adenosylmethionine methylase. *Biochemistry* **57**, 4972–4984 (2018).
47. Parent, A. et al. The B12-radical SAM enzyme PoyC catalyzes valine Cβ-methylation during polytheonamide biosynthesis. *J. Am. Chem. Soc.* **138**, 15515–15518 (2016).
48. Watanabe, K. et al. *Escherichia coli* allows efficient modular incorporation of newly isolated quinomycin biosynthetic enzyme into echinomycin biosynthetic pathway for rational design and synthesis of potent antibiotic unnatural natural product. *J. Am. Chem. Soc.* **131**, 9347–9353 (2009).
49. Suzuki, J. Y., Bollivar, D. W. & Bauer, C. E. Genetic analysis of chlorophyll biosynthesis. *Annu. Rev. Genet.* **31**, 61–89 (1997).
50. Westrich, L., Heide, L. & Li, S. M. CloN6, a novel methyltransferase catalysing the methylation of the pyrrole-2-carboxyl moiety of clorobiocin. *ChemBioChem* **4**, 768–773 (2003).
51. Radle, M. I., Miller, D. V., Laremore, T. N. & Booker, S. J. Methanogenesis marker protein 10 (Mmp10) from *Methanosarcina acetivorans* is a radical *S*-adenosylmethionine methylase that unexpectedly requires cobalamin. *J. Biol. Chem.* **294**, 11712–11725 (2019).
52. Gerlt, J. A. et al. Enzyme Function Initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **1854**, 1019–1037 (2015).
53. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182 (2019).
54. Zallot, R., Oberg, N. O. & Gerlt, J. A. 'Democratized' genomic enzymology web tools for functional assignment. *Curr. Opin. Chem. Biol.* **47**, 77–85 (2018).
55. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
56. Hoops, S. et al. COPASI-a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067–3074 (2006).
57. Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H. & Loew, L. M. A general computational framework for modeling cellular structure and function. *Biophys. J.* **73**, 1135–1146 (1997).

Acknowledgements This work was supported by the National Institutes of Health (NIH) (GM122595 to S.J.B., GM119707 to A.K.B., AI121072 to C.A.T. and GM080189 to E.K.S.) and the Eberly Family Distinguished Chair in Science (S.J.B.). S.J.B. is an investigator of the Howard Hughes Medical Institute. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of GM/CA@APS has been funded in whole or in part with federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). This research also used the resources of the Berkeley Center for Structural Biology supported in part by the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under contract no. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported in part by the NIH, National Institute of General Medical Sciences, grant P30 GM124169. E.K.S. and C.A.T. thank M. S. Lichstrahl for providing a synthetic intermediate, and I. P. Mortimer and J. Tang for their help with ESI-MS and NMR experiments, respectively. We thank D. Iwig for his help in the mass spectrometry analysis of the TokK crystals.

Author contributions H.L.K., E.K.S., C.A.T. and S.J.B. developed the research plan and experimental strategy. H.L.K. isolated and crystallized proteins and collected crystallographic data. H.L.K. and E.K.S. performed biochemical experiments. H.L.K., E.K.S., C.A.T., A.K.B. and S.J.B. analysed and interpreted crystallographic data. H.L.K., E.K.S., C.A.T., A.K.B. and S.J.B. wrote the manuscript.

Competing interests The authors declare no competing interests.

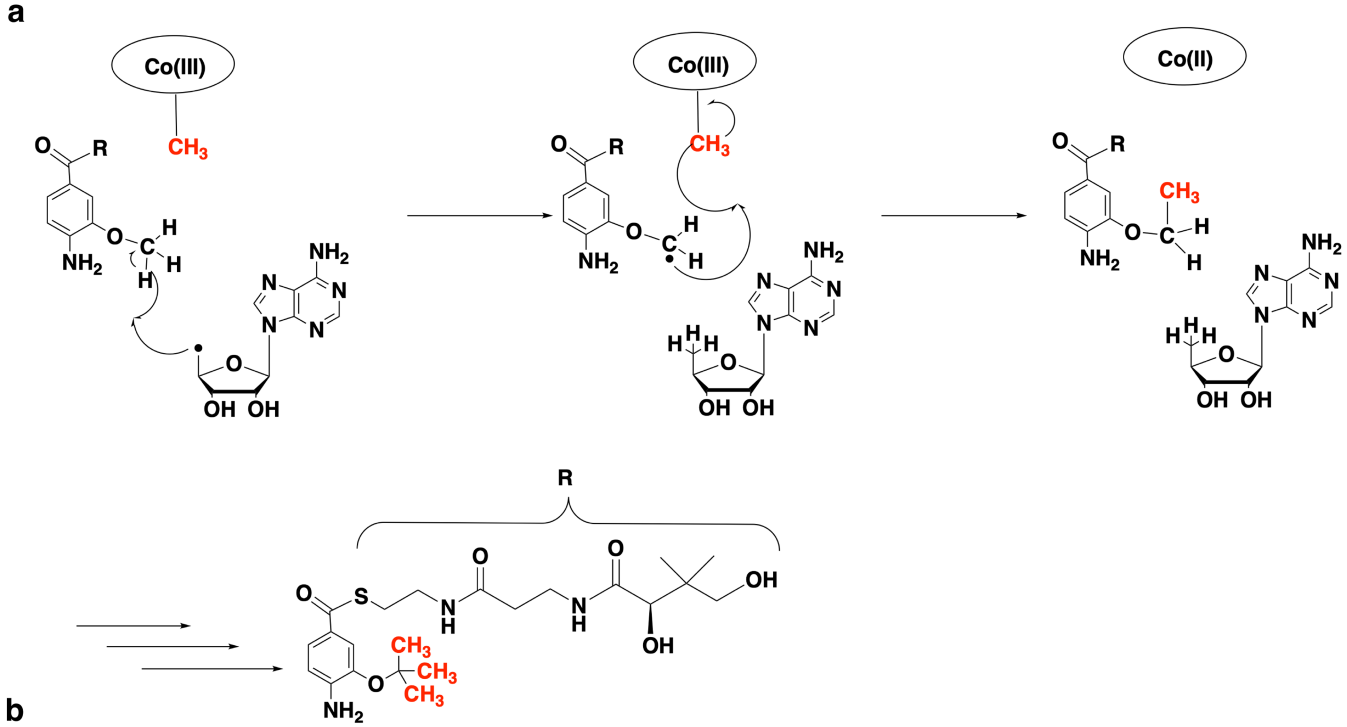
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04392-4.

Correspondence and requests for materials should be addressed to Craig A. Townsend, Amie K. Boal or Squire J. Booker.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

a

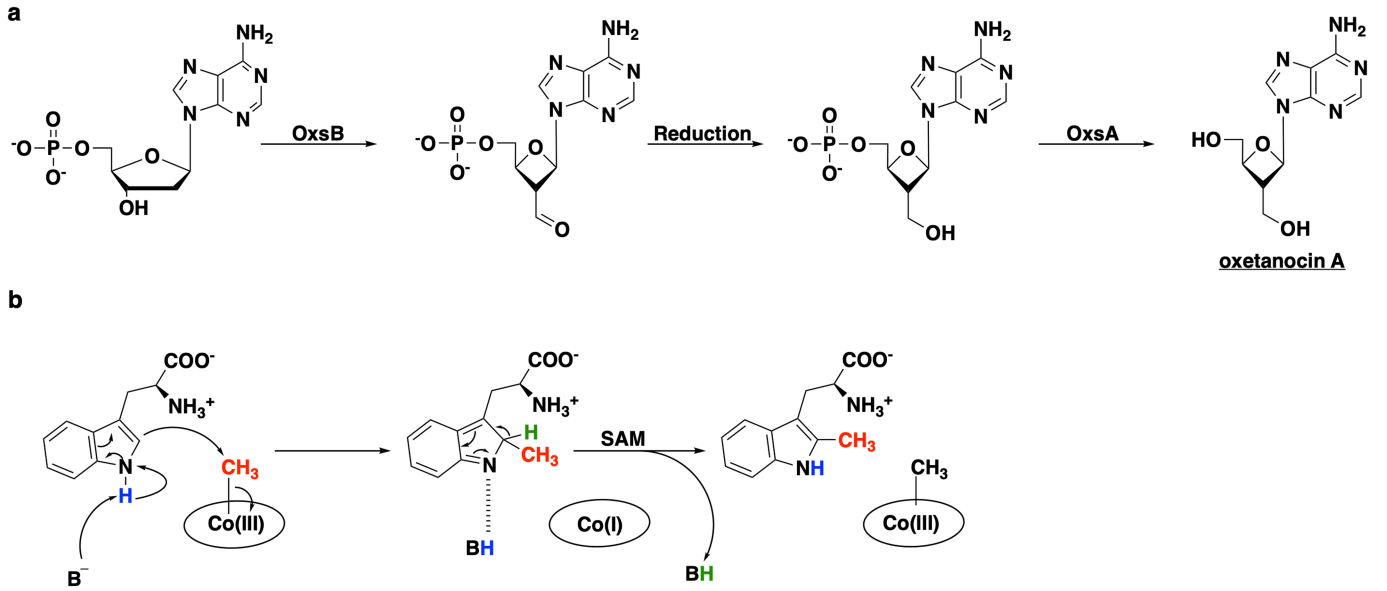


b

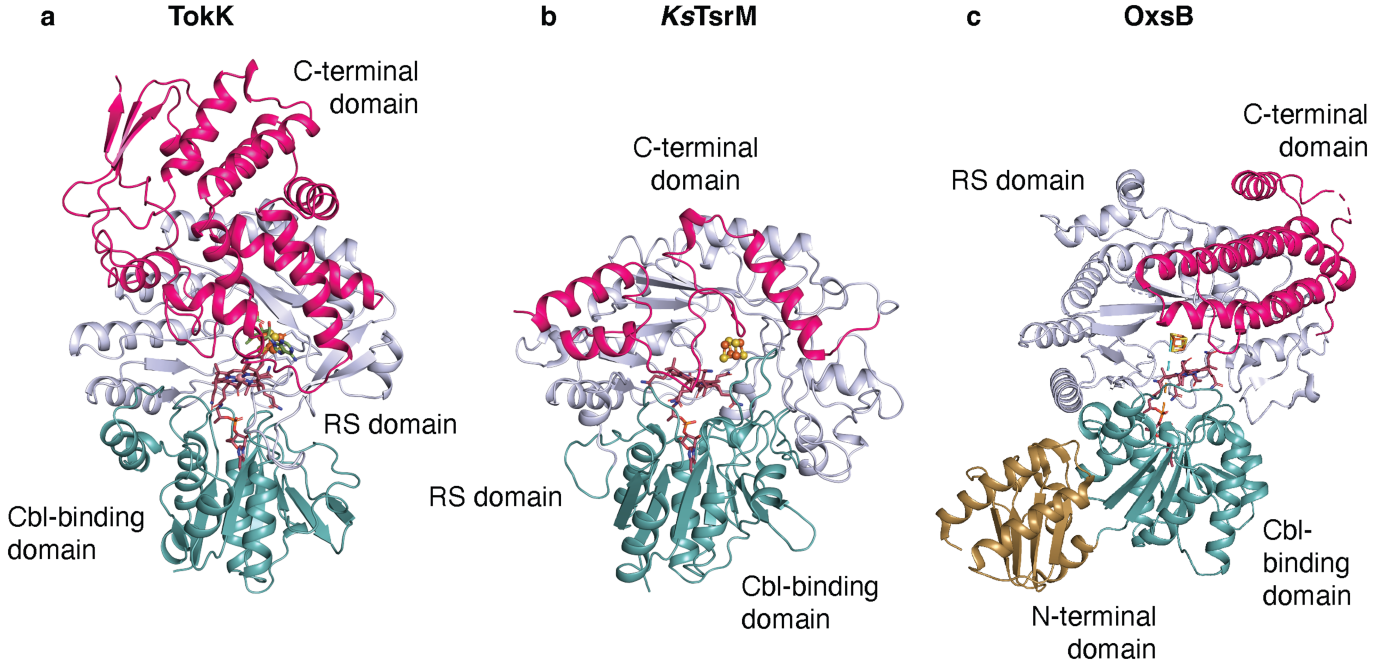
```
TokK    58 EMADGIVEPDVLALSCYVWNF---------RRQM-KVAKLVKERYPNVLVVAGGPHVPDR
ThnK    58 EMAAGITDPDVLALSCYVWNF---------RRQM-KVARLVKERHPGMLVVAGGPHVPDR
CysS    54 EDKLNKIDADVYAISCYVWNM-----G----FVK-RWLPTLTARKPNAHIILGGPQVMNH
TsrM    55 RDYFRAERPLLVGVTCRNTDTVYALEQRPFVDGYKAVIDEV-RRLTAAPVVAGGVGFSTM
Fom3   109 VKVMKEFGPDVVGISSIFSNQ-----A---DNVH-HLLKLADLVTPEAVTAIGGAHARYF
PhpK   107 AKLL-ADGPASVAITTTFYMM--------PAPVI-EIIRFVRQHNATVPIIVGGPLIDNH
GenK   102 EAALAASEADVVGISCMFTPY--------YESAY-ELARMAKRVLPNAKVIVGGQHGTVA

TokK   197    ----YALWETNRGCPYSCSFCDWGSATMSTLRKFEDERLQDEIEWF-ARHDVEDLFICDA
ThnK   197    ----YALWETNRGCPYSCAFCDWGSATMSALRLFDAERLQEEIEWF-AEHDVEDLFVCDA
CysS   185    ----WAPLETNRGCPYQCTYCFWGAATNSRVFKSDMDRVKAEITWL-SQHRAFYIFITDA
TsrM   241    -EGGLGSILTKNGCVYKCSFCVEPDAKGTQFARRGITAVVDEMEAL-TAQGIHDLHTTDS
Fom3   269    VGSRVGQLYTSRGCTAHCTFCTTTHFWGQKLRRRSVQDVVDEVLRLRDEYGIDEFHIQDD
PhpK   240    CLGRTVSTRTARSCAFNCAFCDYPERAGA-LTLADLSTVERELEEL-AERGVKRVAFIDD
GenK   240    ---NAVTLITSRGCPFSCSFCTVHATVGKQFRARDPQRVVDEIEHYVNVHGVRRFLVEDD
```

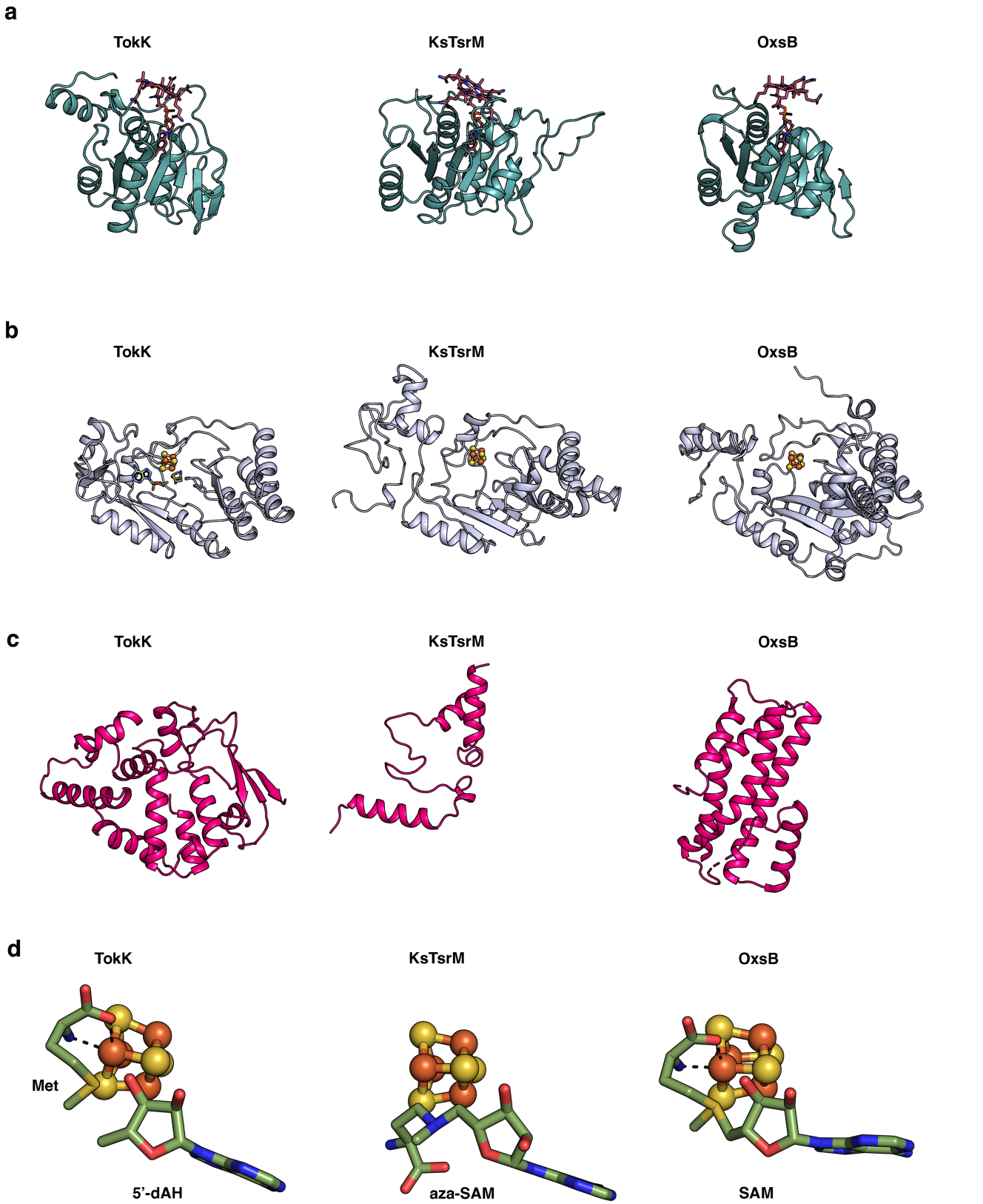
Extended Data Fig. 1 | Comparison to CysS, another class B sequential methylase. a, CysS performs sequential radical methylations like TokK and ThnK. **b**, Partial alignment of TokK (Uniprot ID: A0A6B9HEI0), ThnK (Uniprot ID: F8JND9), CysS (Uniprot ID: A0A0H4NV78), TsrM (Uniprot ID: C0JRZ9), Fom3 (Uniprot ID: Q56184), PhpK (Uniprot ID: A0A0M3N271), and GenK (Uniprot ID: Q70KE5). Cysteines that coordinate the iron-sulfur cluster are highlighted in blue. Trp76 (highlighted in red) and Trp215 (highlighted in yellow) in TokK is conserved in ThnK and CysS, but not in other known Cbl-binding RS methylases. Areas of conservation for TokK, ThnK, and CysS around Trp76 are highlighted in grey. The bottom axial amino acid residue for TsrM (Arg69) is highlighted in red. Completely conserved residues are bolded.



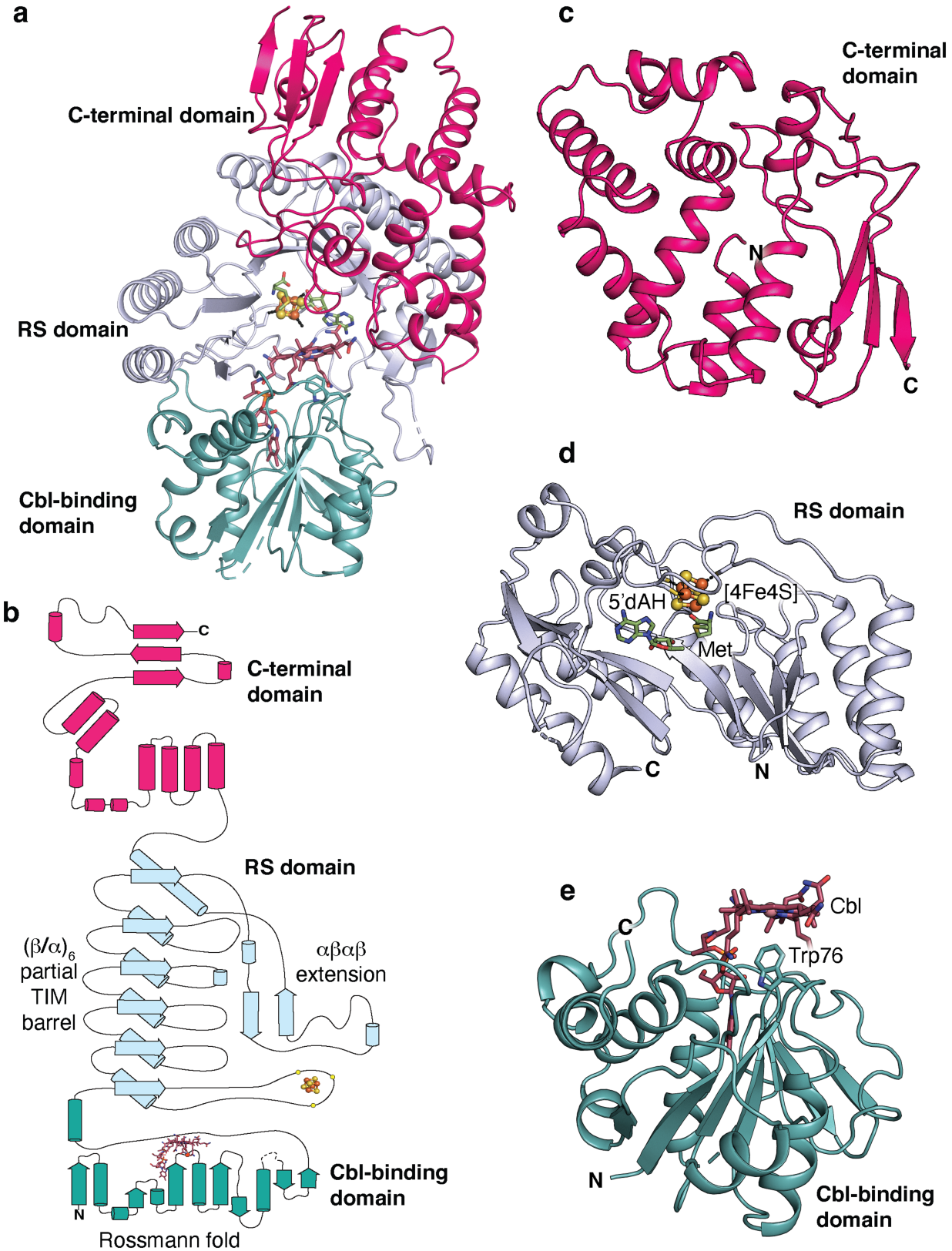
Extended Data Fig. 2 | Reactions performed by OxsB and TsrM, notable Cbl-binding radical SAM enzymes. a, Proposed pathway for the biosynthesis of oxetanocin A by OxsB and OxsA. Currently the aldehyde reduction step is unknown. **b**, Proposed non-radical reaction performed by TsrM. The carboxylate in SAM is implicated in playing a dual role in catalysis, both as the source of the methyl donor and as the base to prime the substrate for nucleophilic attack.



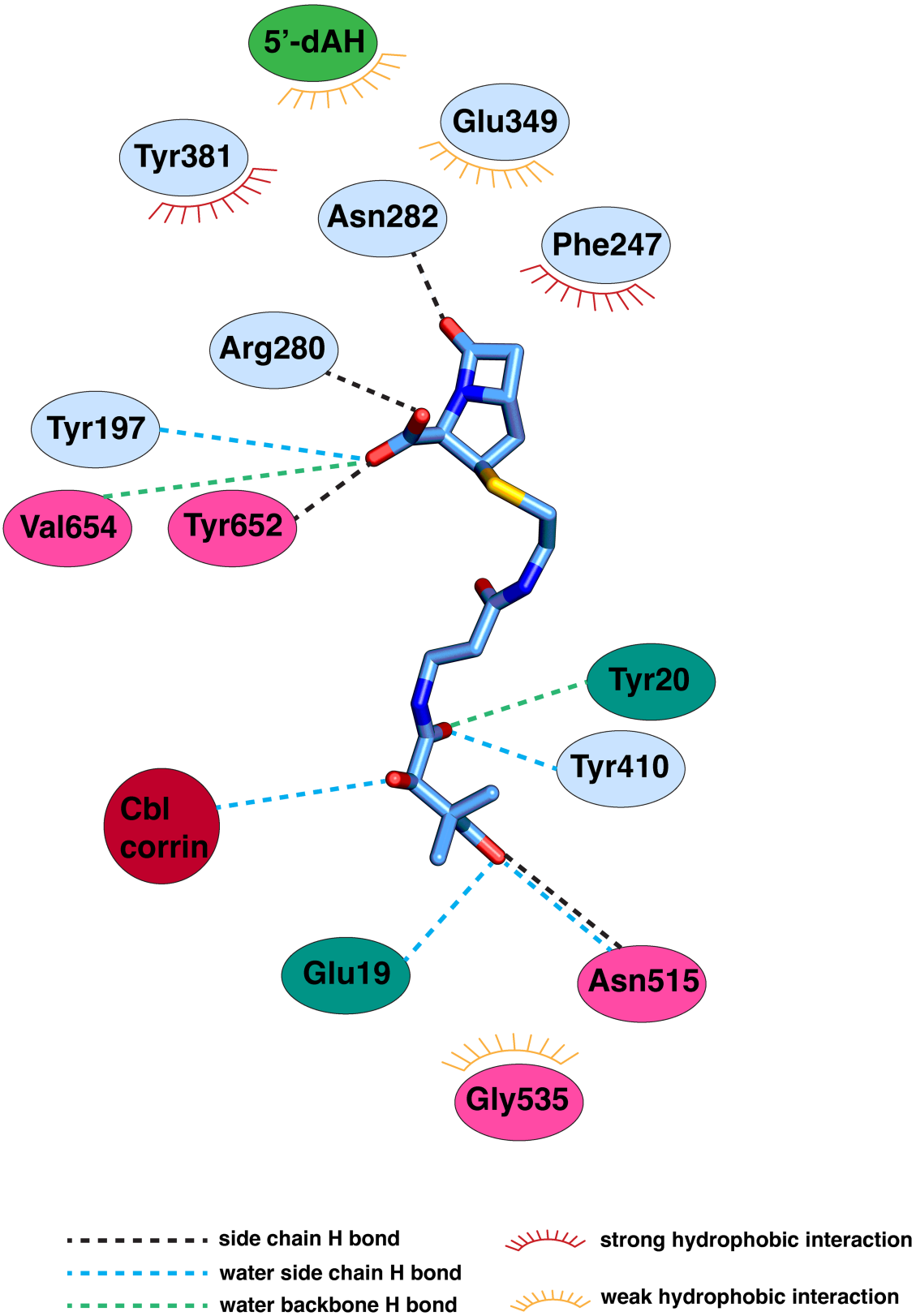
Extended Data Fig. 3 | Structural comparisons of Cbl-dependent RS enzymes. a–**c**, The overall structures of TokK (**a**), *Ks*TsrM (PDB ID: 6WTE) (**b**) and OxsB (PDB ID: 5UL3) (**c**) are shown as ribbon diagrams and coloured by domain (Cbl-binding domain, teal; RS domain, light blue; and C-terminal domain, pink). OxsB has a fourth domain, an N-terminal domain of an unknown function, shown in gold. All three enzymes share very similar Cbl-binding domains with a characteristic Rossmann fold. However, as shown in Extended Data Fig. 4, the RS domain and C-terminal domain differ in each system.



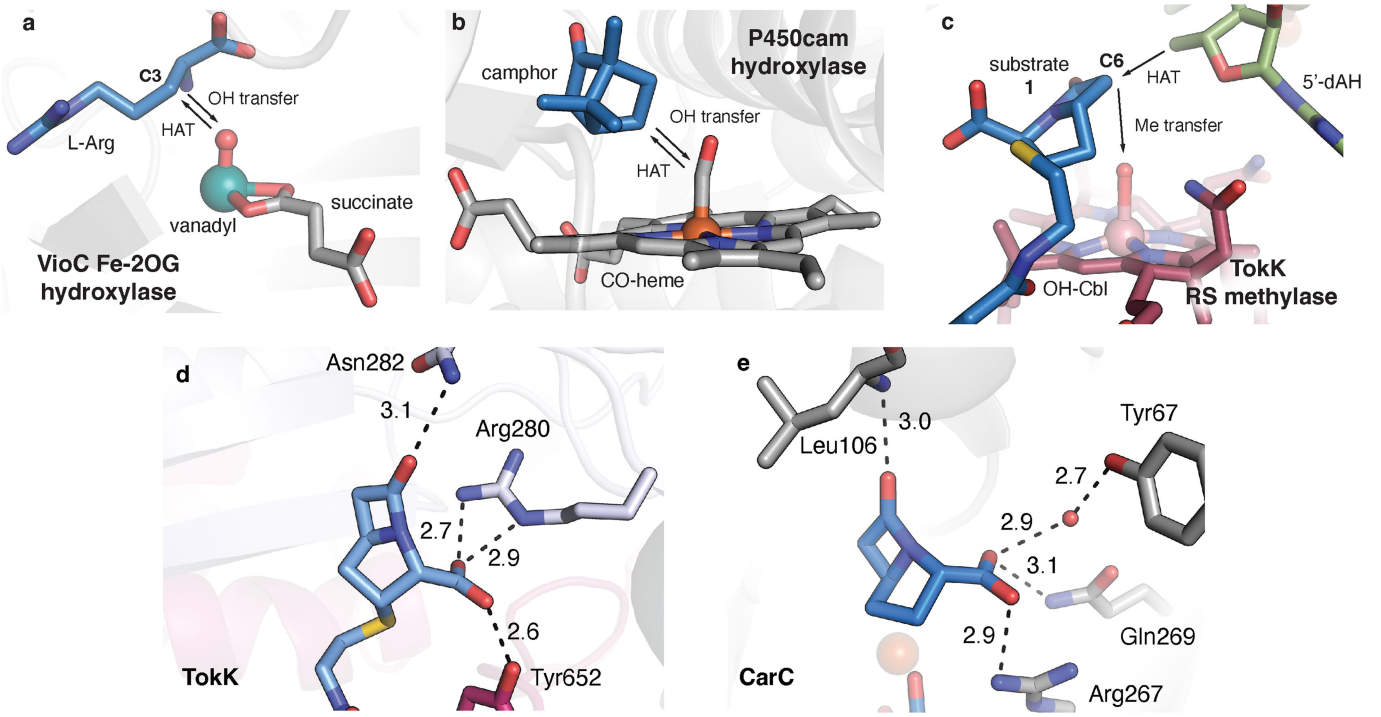
Extended Data Fig. 4 | Comparison of the Cbl-binding, RS, and C-terminal domains of TokK, *Ks*TsrM, and OxsB. The domains are coloured as in Extended Data Fig. 3. **a**, The Rossmann fold, in teal, is highly similar among TokK (PDB ID: 7KDY), *Ks*TsrM (PDB ID: 6WTF) and OxsB (PDB ID: 5UL4). **b**, The core of each of the RS domains is a $(\beta/\alpha)_6$ motif; however, there are distinct differences. The RS domain of OxsB is more compact than those of TokK or *Ks*TsrM, and all three have unique extra secondary structure features. **c**, The C-terminal domains for TokK, *Ks*TsrM, and OxsB are vastly different in architecture. **d**, Comparison of the binding of Met and 5′-dAH, aza-SAM, and SAM for TokK, *Ks*TsrM, and OxsB structures, respectively. Only the relevant binding of SAM to the cluster is shown for OxsB. OxsB has two binding positions of SAM, one to the cluster and one in what is proposed to be an intermediate state towards methylating the Cbl.



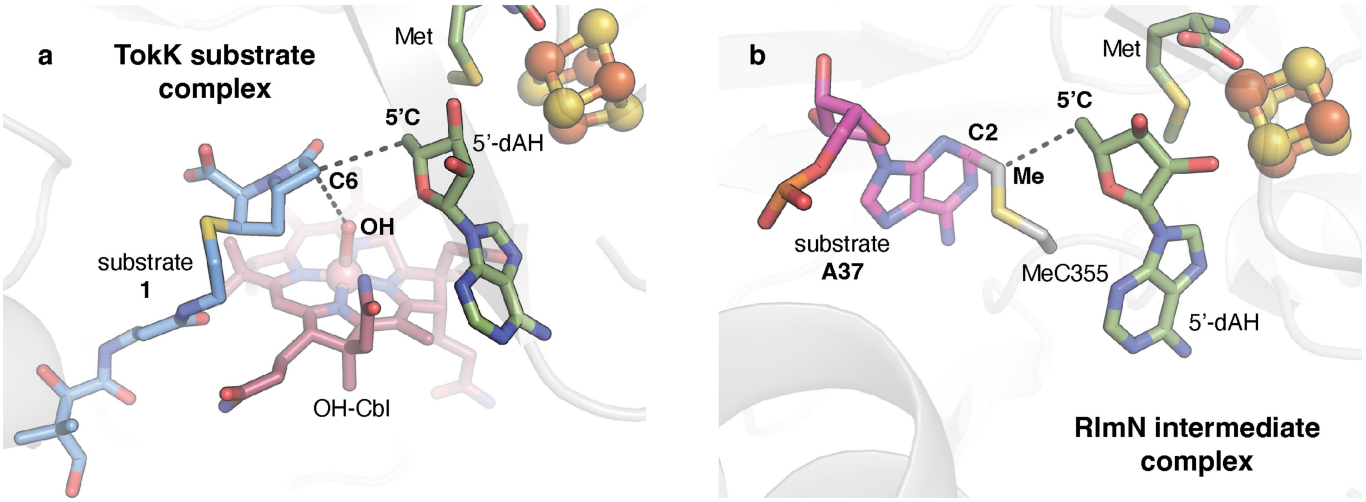
Extended Data Fig. 5 | Domain architecture of TokK. a, The three domains of TokK are portrayed in a ribbon diagram. The N-terminal Cbl-binding domain is shown in teal, the RS domain is shown in light blue, and the C-terminal domain is shown in pink. **b**, A topology diagram of TokK with domains coloured as in **a**. The Cbl is portrayed in sticks, and the lower axial Trp side chain is shown as a red dot. The iron-sulfur cluster is shown in orange and yellow spheres, and the ligating Cys residues are shown as small yellow spheres. **c**–**e**, Zoomed in views of the C-terminal domain (**c**), RS domain (**d**) and Cbl-binding domain (**e**) are shown as ribbon diagrams.



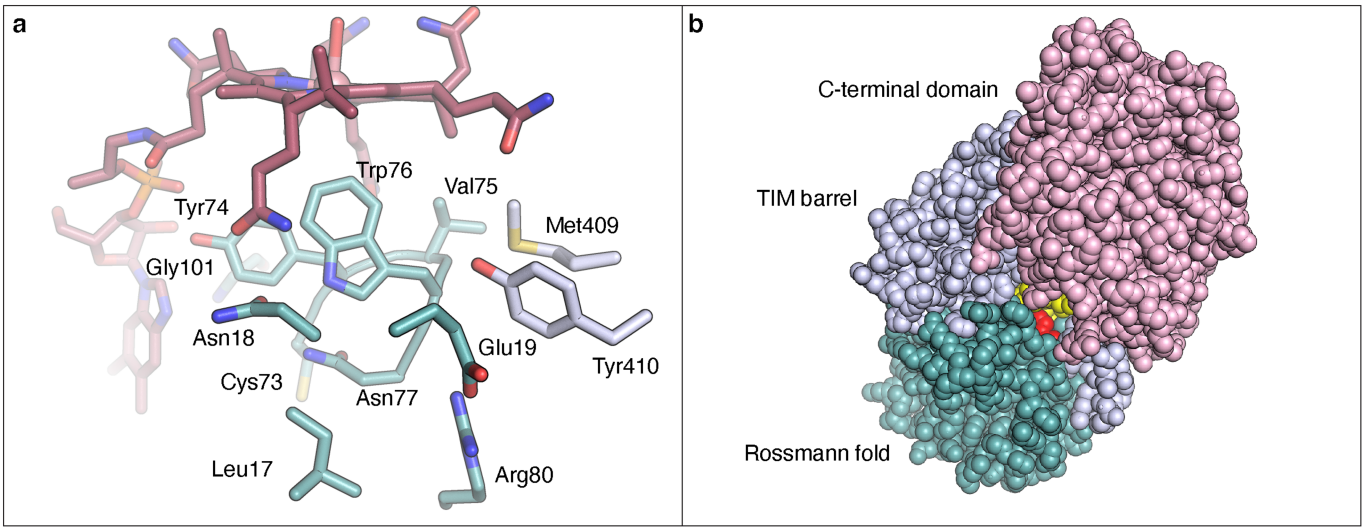
Extended Data Fig. 6 | Diagram of carbapenam interactions in the TokK complex with substrate. Carbapenam substrate **1** shown in blue sticks and coloured by atom type. Polar and hydrophobic interactions were mapped with the LigPlot program and indicated with dashed lines or a starburst symbol.



Extended Data Fig. 7 | Comparison of TokK structure with those of enzymes using alternative radical-generating mechanisms. a–**e**, Residues that interact with the polar substituents in the β-lactam ring are shown in stick format for TokK (**a**) and the epimerase CarC (PDB ID: 4OJ8) (**b**). Although the two enzymes are structurally distinct, they use a similar number and type of functional groups to anchor the β-lactam by using direct and water-mediated contacts to the C7 carbonyl and C3 carboxylate substituents. In addition, comparison of the active site with those of hydroxylases reveals differences in the orientation of hydrogen atom transfer (HAT) intermediates and -OH or -CH_3 functionalization moieties. Fe(II)- and 2-oxo-glutarate-dependent (Fe-2OG) oxygenases use a ferryl [Fe(IV)-oxo)] intermediate to abstract an H-atom from an unactivated substrate carbon. The resulting Fe(III)-OH complex then couples with the substrate radical to yield a hydroxylated product. A vanadyl [V(IV)-oxo] mimic of the ferryl intermediate in L-Arg C3 hydroxylase, VioC, reveals that HAT and -OH transfer must occur from the same side of the substrate target carbon (**c**). The distance between the reactive oxo group and the substrate target carbon (indicated by the arrows) is 3.1 Å in VioC. A structure of the haem-dependent hydroxylase, P450cam, shows a similar phenomenon. A CO-bound mimic of the Fe(IV)-porphyrin radical intermediate, compound I, demands the same arrangement of HAT and -OH transfer components relative to the hydroxylation target on the camphor substrate (**d**). The distance between a mimic of the reactive group and the substrate target carbon (indicated by the arrows) is 3.1 Å in P450cam. By contrast, RS methylases use separate HAT reagents (5′-dA•) and methyl group donors (Me-Cbl), allowing for more diverse stereochemical outcomes in C-H functionalization reactions. (**e**) The structure of TokK in complex with carbapenam substrate, **1**, shows a ~120° angle between the HAT acceptor (5′C of 5′-dAH), the substrate target carbon (C6), and the Cbl top ligand (-OH of OHCbl, a surrogate for Me-Cbl).



Extended Data Fig. 8 | A comparison of the substrate-binding sites in two structurally characterized RS methylases. a, **b**, The substrate complexes of TokK (**a**) and RlmN (PDB ID: 5HR7) (**b**) are shown. RlmN is an RS methylase that uses a radical-based mechanism to methylate an sp^2-hybridized carbon (C2) of an adenine base in transfer or ribosomal RNA. By contrast to TokK and other Cbl-dependent RS enzymes, RlmN uses a 5′-dA• to activate a post-translationally modified methyl-Cys residue on a C-terminal loop in the active site to modify its aromatic substrate via radical addition. Comparison of a structure of RlmN in a cross-linked methylCys-tRNA intermediate state to the TokK substrate complex shows that, despite the differences in reaction mechanism, these systems use a similar orientation of radical initiator (5′-dA•), substrate target carbon (C6, C2), and methyl donor (OH, Me). In the TokK substrate complex, the -OH ligand of the Cbl cofactor serves as a surrogate for the position of the methyl donor.



Extended Data Fig. 9 | Solvent accessibility of the bottom axial face of the Cbl in TokK. a, Nearest residues (~6 Å) to the bottom face of the Cbl. **b**, View of TokK using space-filling model to show a channel to the active site. As can be seen, only a very small portion of the Cbl (coloured yellow) is solvent accessible, with most of the Cbl being buried within the Rossmann fold. Trp76 is coloured red.

Extended Data Table 1 | X-ray crystallographic data collection and refinement statistics

	5'-dAH•Met•TokK Fe anomalous (peak)	5'-dAH•Met•TokK native	Substrate-bound 5'-dAH•Met•TokK native
Data collection			
Space group	$P\,2_1$	$P\,2_1$	$P\,2_1\,2_1\,2_1$
Wavelength (Å)	1.72194	1.03313	1.03320
Cell dimensions			
a, *b*, *c* (Å)	78.18, 138.58, 78.18	68.64, 137.74, 77.91	84.50, 132.82, 155.51
α, β, γ (°)	90.73	90.47	90
Resolution (Å)	50.0-2.52 (2.56-2.52)	50.0-1.79 (1.82-1.79)	50.0-1.94 (2.01-1.94)
No. of unique reflections	50,189	132,404	129,775
R_{sym} or R_{merge}	0.08 (0.42)	0.08 (0.84)	0.11 (0.59)
R_{pim}	0.03 (0.18)	0.03 (0.34)	0.04 (0.22)
$I / \sigma I$	67.5 (5.8)	56.9 (4.9)	58.5 (11.8)
$CC_{1/2}$	(0.89)	(0.78)	(0.99)
Completeness (%)	99.8 (96.2)	97.8 (96.4)	99.9 (99.4)
Redundancy	6.4 (5.0)	6.7 (6.0)	7.3 (7.1)
Refinement			
Resolution (Å)		1.79	1.94
No. reflections		130,677	129,773
R_{work} / R_{free}		0.17/0.19	0.16/0.19
No. atoms		11,947	12,159
Protein		10,738	10,687
Fe/S cluster		16	16
OH-Cobalamin		183	183
5'-dAH		218	218
Met		236	236
Ethylene glycol		84	-
Chlorine		2	-
Potassium		2	4
Substrate		-	29
Water		887	1,192
B-factors Å^2			
Protein		25.5	24.3
Fe/S cluster		12.6	12.4
Cobalamin		15.5	16.0
5'-dAH		12.8	14.6
Met		11.6	12.4
Ethylene glycol		36.7	-
Chlorine		47.5	-
Potassium		22.0	25.1
Substrate		-	33.7
Water		34.8	31.8
R.m.s. deviations			
Bond lengths (Å)		0.006	0.007
Bond angles (°)		1.11	1.12
Clashscore		4.49	3.06
Rotamer outliers (%)		0.27	0.09
Ramachandran			
Most favored (%)		97.41	97.73
Allowed (%)		2.59	2.27
Outliers (%)		0.00	0.00
Number of TLS groups		1	1
PDB accession code		7KDX	7KDY

Values in parentheses represent the highest-resolution shell. All structures result from one single crystal.

nature portfolio

Corresponding author(s): Squire J. Booker

Last updated by author(s): November 11, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	HKL2000/3000 packages were used for data processing of TokK crystals. Phaser 2.8.3 and Autobuild 2.15 were used for SAD phasing.
Data analysis	The Phenix suite (including phenix.refine, Autosol, Autobuild, etc) v. 1.12-2829 was used for crystallography data refinement. Coot 0.8.9.2. was used for crystallography modeling. PyMOL 2.5.0 was used to make crystallography figures. HOLLOW v. 1.2 was used for visualization of protein cavities. LigPLOT v. 4.5.3 was used to visualize substrate binding. Cytoscape 3.8.2 was used for analysis of the SSN. First order rate constants were determined using the COPASI 4.26 parameter estimation tool and curves were simulated using Vcell 7.4.0.00.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Atomic coordinates and structure factors for the reported crystal structures in this work have been deposited to the Protein Data Bank (PDB) under the accession numbers 7KDX (TokK with 5'-dAH and Met bound) and 7KDY (substrate-bound TokK).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Structures from individual protein crystals were solved by SAD phasing. Sample size calculation is not applicable to this study, because single protein crystals were solved.
Data exclusions	Anomalous and native TokK datasets were integrated and scaled using the HKL2000 or HKL3000 package. No data exclusions were applied.
Replication	The assays of the variant and wild-type TokK protein were performed in triplicate and noted. All attempts at replication were successful.
Randomization	Randomization is not applicable to this study, because single protein crystal structures were solved.
Blinding	No crystal structure of this enzyme exists and as such, no expectations were made on how it would appear. Thus, blinding was not necessary for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging