



Published in final edited form as:

*Nature*. 2022 February ; 602(7896): 343–348. doi:10.1038/s41586-021-04392-4.

## Structure of a B<sub>12</sub>-dependent radical SAM enzyme in carbapenem biosynthesis

Hayley L. Knox<sup>1,\*</sup>, Erica K. Sinner<sup>4,\*</sup>, Craig A. Townsend<sup>4,†</sup>, Amie K. Boal<sup>1,2,†</sup>, Squire J. Booker<sup>1,2,3,†</sup>

<sup>1</sup>Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

<sup>2</sup>Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

<sup>3</sup>Howard Hughes Medical Institute, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

<sup>4</sup>Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA.

### Summary Paragraph

Carbapenems are antibiotics of last resort in the clinic. Due to their potency and broad-spectrum activity, they are an important part of the antibiotic arsenal. The vital role of carbapenems is exemplified by Merck's FDA approval for the use of an imipenem combination therapy to treat elevated hospital-acquired and ventilator-associated bacterial pneumonia during the COVID-19 pandemic<sup>1</sup>. The C6 hydroxyethyl side chain distinguishes the clinically-used carbapenems from the other classes of  $\beta$ -lactam antibiotics and is responsible for their low susceptibility to inactivation by occluding water from the  $\beta$ -lactamase active site<sup>2</sup>. The construction of the C6 hydroxyethyl side chain is mediated by cobalamin- or B<sub>12</sub>-dependent radical S-adenosylmethionine (SAM) enzymes<sup>3</sup>. These radical SAM methylases (RSMTs) assemble the alkyl backbone by sequential methylation reactions, thereby underlying the therapeutic usefulness of these essential medicines. Here we present x-ray crystal structures of TokK, a B<sub>12</sub>-dependent RSMT that catalyzes three-sequential methylations during the biosynthesis of asparenomicin A. These structures, containing the enzyme's two metallofactors and determined in the presence and absence of a carbapenam substrate, give the first visualization of a B<sub>12</sub>-dependent RSMT that employs the radical mechanism shared by a vast majority 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.

<sup>†</sup>Correspondence and requests for materials should be addressed to Squire J. Booker (squire@psu.edu), Amie K. Boal (akb20@psu.edu) or Craig A. Townsend (ctownsend@jhu.edu).

**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. analyzed and interpreted crystallographic data. H.L.K., E.K.S. C.A.T. A.K.B. and S.J.B. wrote the manuscript.

\*These authors contributed equally to this work.

The authors declare no competing interests

**Additional Information:** supplementary information is available for this paper (Figs. S1–S7 and Table S1).

## One Sentence Summary:

Structural insight into a cobalamin-dependent radical SAM methylase that performs three sequential radical-mediated methylations to install the C6 side chain of a carbapenem antibiotic.

---

## INTRODUCTION

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<sub>12</sub>)-dependent radical *S*-adenosylmethionine (SAM) enzyme<sup>4-6</sup>. 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**)<sup>7</sup>. An ortholog of ThnK, TokK from *Streptomyces tokunonensis* (ATCC 31569), constructs the C6 isopropyl chain of the carbapenem, asparenomyacin 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 employed in the biosynthesis of cystobactamids, in which a similar Cbl-dependent radical SAM enzyme, CysS, performs successive methyl transfers (Extended Data Fig.1)<sup>8</sup>. Interestingly, despite low sequence identity (~29%) between CysS and TokK or ThnK, all three proteins are located within the same node of a sequence similarity network (SSN) composed of approximately 11,000 cobalamin-binding radical SAM (RS) enzymes obtained from networks provided by [radicalSAM.org](http://radicalSAM.org) and the SFLD (Fig. 1c and S1). 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*<sup>3</sup>-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, novel lipids, and natural products with antiproliferative biological activity<sup>9-12</sup>. While 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 centers by using methylcobalamin (MeCbl) as an intermediate methyl donor. All RS enzymes, with a single known exception<sup>13-15</sup>, 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 abstracting a hydrogen atom. In B<sub>12</sub>-dependent RSMTs, the substrate radical attacks the methyl group of MeCbl, inducing homolytic cleavage of the cobalt-carbon bond to afford cob(II)alamin and the methylated product (Fig. 1b). Upon 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), a supernucleophile, which acquires a methyl group from SAM to regenerate MeCbl (Fig. 1b)<sup>11,12</sup>.

Two Cbl-dependent RS enzymes have been structurally characterized, TsrM and OxsB (Fig. 1c and S1)<sup>13,16</sup>, involved in the biosynthesis of antibiotics thiostrepton A and

oxetanocin A, respectively. Both enzymes are mechanistic outliers among Cbl-dependent RS enzymes and are found in SSN nodes distinct from each other and from TokK (Fig. 1c). OxsB uses cobalamin in an unknown fashion to catalyze a complex ring contraction of 2'-deoxyadenosine monophosphate (dAMP) (Extended Data Fig. 2a)<sup>16</sup>. TsrM methylates an  $sp^2$ -hybridized carbon, C2, of L-tryptophan (Trp) by a polar mechanism (Extended Data Fig. 2b)<sup>13</sup>. TsrM is distinctive among all RS enzymes because it does not catalyze the formation of 5'-dA• during catalysis<sup>13</sup>. Instead, TsrM uses SAM's carboxylate moiety as an acceptor of the N1 proton of Trp during C2 electrophilic substitution by MeCbl<sup>13,17</sup>. Additionally, 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<sup>16</sup>. TsrM has been co-crystallized with aza-SAM (a SAM analog) and Trp, but the Trp substrate is bound in an unproductive conformation, requiring computational docking to understand the structural basis for the reaction outcome.

## RESULTS AND DISCUSSION

TokK was crystallized under anoxic conditions in the presence of 5'-dAH and Met, the products of reductive SAM cleavage (Fig. S2). 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 *KsTsrM* (Extended Data Fig. 4d). A third, C-terminal domain, is distinct to TokK (Fig 2a, Extended Data Fig. 5)<sup>13,16</sup>. While as-isolated TokK contains methylcobalamin (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 (Fig. S3). In the structure of the TokK•OHCbl•5'-dAH•Met•substrate complex, which mimics the complex immediately prior to 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 modeling (Fig. S4). 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 (Figs. 2a and S5). 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 a key role for this domain 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  $\beta$ -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  $\beta$ -lactam interaction resembles that of carbapenem synthase, the enzyme responsible for inversion of stereochemistry at C5 in simple carbapenems (Extended Data Fig. 7)<sup>18</sup>. Both enzymes share use of H-atom abstraction chemistry to selectively target an unactivated C-H bond within the bicyclic  $\beta$ -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 significantly from their positions in the structure without substrate bound (Fig. 2a), and substitution of Arg280 to 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  $\beta$ -lactam appropriately both for activation of C6 by 5'-dA• and subsequent methyl addition by the Cbl cofactor (Fig. 2d)<sup>19</sup>. 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 if 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  $\sim 85^\circ$  relative to the 5'-carbon of 5'-dAH. This arrangement suggests that the methyl group adds to the bottom face of the  $\beta$ -lactam ring, consistent with the absolute configurations observed in the TokK products and thienamycin<sup>19,20</sup>. The distance and orientation of reactant functional groups in TokK also compares favorably to other enzymes that catalyze radical-mediated activation and functionalization of a substrate C-H bond, such as iron-dependent hydroxylases in the cytochrome P450 and 2-oxo-glutarate-(2OG)-dependent superfamilies (Extended Data Fig. 7). These systems orient their reactive groups similarly, but over a slightly shorter distance range<sup>21,22</sup>. 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, allowing for more diverse stereochemical outcomes.

Intriguingly, 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 (Extended Data Fig. 8)<sup>23</sup>. RlmN uses an *S*-methyl cysteinyl (methylCys) residue as an intermediate methyl carrier during the methylation of the *sp*<sup>2</sup>-hybridized C2 atoms of adenosine 2503 in ribosomal RNA and adenosine 37 in several transfer RNAs (tRNA). When the structure of RlmN crosslinked to an *E. coli* tRNA<sup>Glu</sup> 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, wherein 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 (DMB) 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<sup>24</sup>. 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, likely promoting 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 (Fig. 3a)<sup>25–27</sup>. 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 DXHXXG motif exemplified by methionine synthase, although both are found in the loop following  $\beta$ 3 in the Rossmann fold. To investigate the role of this residue, Trp76 was substituted with Phe and Ala. Unexpectedly, rates of methylation slightly increased for both site-specific substitutions, while analysis by electron paramagnetic resonance (EPR) spectroscopy suggested that both variants exhibited the same 4-coordinate geometry as wild-type TokK (Fig. S6). 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 ~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<sup>13</sup>. While 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 catalyzed by this enzyme (Fig. 3b). The second methylation to form the ethyl-containing carbapenam product **5** proceeds at least three-fold faster than the formation of **4**, a pattern that runs counter to known differences in 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 carbapenam 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 explain why  $k_3$  is markedly slower than  $k_1$  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 prior to 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.<sup>19</sup> A similar mechanism was proposed recently for CysS (Extended Data Fig. 1)<sup>8</sup>. While 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 dramatically slows substrate methylation by TokK, suggesting it may play a role in catalysis (Fig. 3b).

ThnK and TokK share 79.3% sequence identity and act on the same substrate, (2*R*)-pantetheinylated carbapenam (Fig. S7), but ThnK performs two sequential methylations while TokK catalyzes three<sup>19,28</sup>. Nearly all residues in proximity to the active site are identical in the two orthologs. However, three non-conserved amino acids near the active site were explored to determine their role in controlling the extent of methylation. Leu383 is positioned deep in the active site and near 5'-dAH (Fig. S7). When this residue is substituted with Phe, which is found at the same position in the primary structure of ThnK (Fig. S7), 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, E19 and Y20, (Fig. 2a, S7) 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 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 ( $k_1 > k_2$ ,  $k_3 = 0$ ) (Fig. 3b)<sup>19</sup>.

The structure of TokK solved in the absence of substrate reveals remarkably few differences in overall fold or domain organization relative to the TokK•1 complex (rmsd of 0.53 Å over 603 residues by C-alphas). 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 (Figs. 2a, S5). 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, where a loop from the C-terminal moves to cap the active site in the presence of Trp<sup>13</sup>. Although the C-terminal domain of TokK is significantly 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 prior to 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. Interestingly, this scaffolding approach, which is underscored by the notable absence of conformational changes upon substrate binding, may be shared by the only other Cbl-dependent RS enzyme 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.

## METHODS

### Overexpression and purification of TokK from *Streptomyces tokunonesis*

TokK (Uniprot ID: A0A6B9HEI0) was produced heterologously in *Escherichia coli* BL21 (DE3) via overexpression from a pET29b:*tokK*Tev construct<sup>19</sup>. To facilitate production of soluble TokK protein with maximal occupancy of [4Fe–4S] cluster and cobalamin cofactors, the strain was transformed with two additional plasmids, pDB1282 and pBAD42-BtuCEDFB<sup>19,29–31</sup>. A 100 mL LB starter culture with 50 µg/mL kanamycin (pET29b) containing *tokK*, 50 µg/mL spectinomycin (pBAD42-BtuCEDFB), and 100 µg/mL 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 OD<sub>600</sub> 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<sub>3</sub> and 150 µM cysteine were added at the initiation of induction with arabinose. The cultures were then grown to OD<sub>600</sub> 0.6 followed by incubation in an ice–water bath for 1 h. IPTG was added to a final concentration of 1 mM, and the cultures were incubated at 18 °C for an additional 18 h before harvesting the cells by centrifugation at 6000 × *g*. The resulting cell paste (~50 g) was flash frozen in liquid N<sub>2</sub> and stored in liquid N<sub>2</sub> prior to 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 of lysis buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 10% glycerol, 5 mM imidazole, 10 mM BME). The cell suspension was incubated with lysozyme (1 mg/mL), DNase (0.1 mg/mL), and PMSF (0.18 mg/mL) for 30 min at room temperature and then cooled to 4 °C<sup>19,31</sup>. Cells were then lysed by sonic disruption (70% amplitude, 45 s on, 59 s off, ~15 min) and then centrifuged at 50000 × *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 of lysis buffer prior to addition of elution buffer (lysis buffer supplemented with 500 mM imidazole). Dark colored elution fractions were pooled and concentrated to 1 mL in an Amicon 10 kDa MWCO ultrafiltration device (EMD Millipore; Billerica, MA). The protein fractions were then exchanged into TEV protease cleavage buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 15% glycerol, 10 mM BME) using a PD-10 pre-poured gel-filtration column from GE Biosciences (Pittsburgh, PA). The exchanged protein sample was allowed to react with 10 units of TEV protease (~2 mg/mL) in 25 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM TCEP, and 50% glycerol (Millipore Sigma; St Louis, MO) for two days on ice to generate TokK containing only 6 additional amino acids (ENLYFQ) on its C-terminus.

On the second day, the [4Fe–4S] cluster and cobalamin cofactors were reconstituted as previously described<sup>31</sup>. 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, 15% glycerol) for size-exclusion chromatography. In this step, the protein was applied to a HiPrep 16/60 S200 column using an ÅKTA fast protein liquid chromatography (FPLC) system (GE Biosciences) housed in the anaerobic chamber. TokK elutes as an apparent monomer. Fractions were pooled based on UV-vis absorption at 280 and 410 nm and concentrated to 38 mg/mL.

### 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<sup>28</sup>.

Characterization matched that previously reported.

### X-ray crystal structure determination 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<sup>32–35</sup>. Model building and refinement were performed with Coot and phenix.refine<sup>32,36</sup>. Figures were prepared using PyMOL<sup>37,38</sup>. Substrate channel figures were prepared using Hollow<sup>39</sup>. Ligplot was utilized to visualize the binding of substrate **1**<sup>40</sup>.

**Crystallization and structure solution of 5'-dAH•Met•TokK.**—Purified TokK protein aliquots were diluted to 8 mg/mL in 34 mM HEPES, pH 7.5. 5'-dAH and Met (Millipore Sigma, St. Louis, MO) 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 vapor-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  $\mu$ l of protein with 1  $\mu$ l precipitating solution, followed by equilibration against 500  $\mu$ l of a 0.5 M LiCl well solution at room temperature. Crystals were prepared for data collection by mounting on rayon loops followed by 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 ( $\lambda = 1.72194 \text{ \AA}$ ) with 360° of data measured using a 0.5° oscillation range to 2.52  $\text{\AA}$  resolution. Additionally, a 1.79- $\text{\AA}$  resolution native dataset was collected at  $\lambda = 1.03313 \text{ \AA}$  (Extended Data Table 1). Heavy atom sites were identified using HySS implemented within Phenix Autosol<sup>32</sup>. The initial overall



figure-of-merit (FOM) was 0.265 and the Bayes CC was 18.2<sup>32</sup>. 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  $R_{\text{work}}/R_{\text{free}}$  of 0.25/0.31. Iterative manual model building and refinement were performed in Coot and Phenix<sup>36</sup>. This model was used to obtain phase information for the 1.79-Å resolution native dataset by molecular replacement using Phenix Phaser-MR<sup>32</sup>. Geometric restraints for 5'-dAH and cobalamin were obtained from the Grade Web Server (Global Phasing).  $R_{\text{free}}$  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 hydroxycobalamin cofactor, one methionine (Met), and one 5'-deoxyadenosine (5'-dAH) in Chain A; residues 7–413, 416–672, one [4Fe-4S] cluster, one hydroxycobalamin cofactor, one Met, and one 5'-dAH in Chain B. The final model also contains two chloride ions, two potassium ions, 21 molecules of ethylene glycol, and 887 waters. The Ramachandran plot shows that 97.4% residues are in favored regions with the remaining 2.6% in allowed regions<sup>37</sup>. 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 TokK in 34 mM HEPES, pH 7.5. 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 vapor 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. Prior to looping, substrate 1 concentration 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<sup>32,36</sup>. Initial geometric restraints for (2*R*)-pantetheinylated carbapenem substrate were generated by eLBOW in Phenix<sup>32</sup>. Final geometric restraints for the carbapenem substrate were created by using the PRODRG 2 server<sup>41</sup>. The final model consists of residues 8–566, 571–672, one [4Fe-4S] cluster, one hydroxycobalamin 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 hydroxycobalamin cofactor, one Met, and one 5'-dAH in chain B. The final structure also contains two glycerol molecules, four potassium ions, and 1193 waters. The Ramachandran plot showed that 97.7% of residues are in favored regions with the remaining 2.3% in allowed regions<sup>37</sup>. Data collection and refinement statistics are shown in Extended Data Table 1.

### **Cobalamin ligand assignment in protein crystals**

The protocol was adapted from previously published studies<sup>29</sup>. Approximately 15 TokK + 5'-dAH + Met crystals were looped from the crystallization drop into 40 µL of mother

solution in a darkly colored Eppendorf tube. In the dark, 50 mM H<sub>2</sub>SO<sub>4</sub> was added to the tube. The tube was vortexed and then centrifuged for 20 min. 5 µl 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. Cobalamin forms were detected by ESI<sup>+</sup>, scanning from *m/z* 150–1700 with a resolution of 120,000. A calibration curve (0.1 µM – 5 µM) of cobalamin standards was run concurrently to quantify the cobalamin forms in the sample.

### Generation of sequence similarity networks (SSNs)

A pool of annotated Cbl-dependent radical SAM enzymes was generated by merging representative annotated B<sub>12</sub>-dependent radical SAM enzyme sequences from [radicalsam.org](https://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)<sup>42,43</sup>. Additional sequences were included for the following functionally characterized enzymes (with Uniprot accession codes): *STsrM* (C0JRZ9), *KsTsrM* (E4N8S5), GenK (Q70KE5), GenD1 (Q2MG55), PhpK (A0A0M3N271), Fom3 (Q56184), CysS (A0A0H4NV78), OxsB (O24770), PoyC (J9ZXD6), swb7 (D2KTX6), ArgMe (Q8THG6), ThnK (F8JND9), TokK (A0A6B9HEI0), BchE (Q7X2C7), CouN6 (A0A1H2F7M3), and CloN6 (Q9F8U1)<sup>8,13,15,16,19,44–52</sup>. The pool of B<sub>12</sub>-dependent radical SAM enzyme sequences from [radicalsam.org](https://radicalsam.org) contains 9,724 representative entries reflecting 53,470 unique sequences. The class B and B<sub>12</sub>-binding radical SAM 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<sup>42</sup>. 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 >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<sup>43,53–55</sup> 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 and Fig. S1) is depicted as a representative node network in which each node reflects sequences with >40% identity. All networks were visualized with the Organic layout in Cytoscape<sup>56</sup>. The SSN in Figure 1C represents 36 sequence clusters extracted from the full SSN shown in Fig. S1. The clusters in Figure 1C were selected based on 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 photolyzed on ice for 45 minutes 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 prior to 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 (Concord, MA) variable-temperature helium-flow cryostat. All measurements were taken at 70K, with 1mT modulation amplitude, and 1mW power.

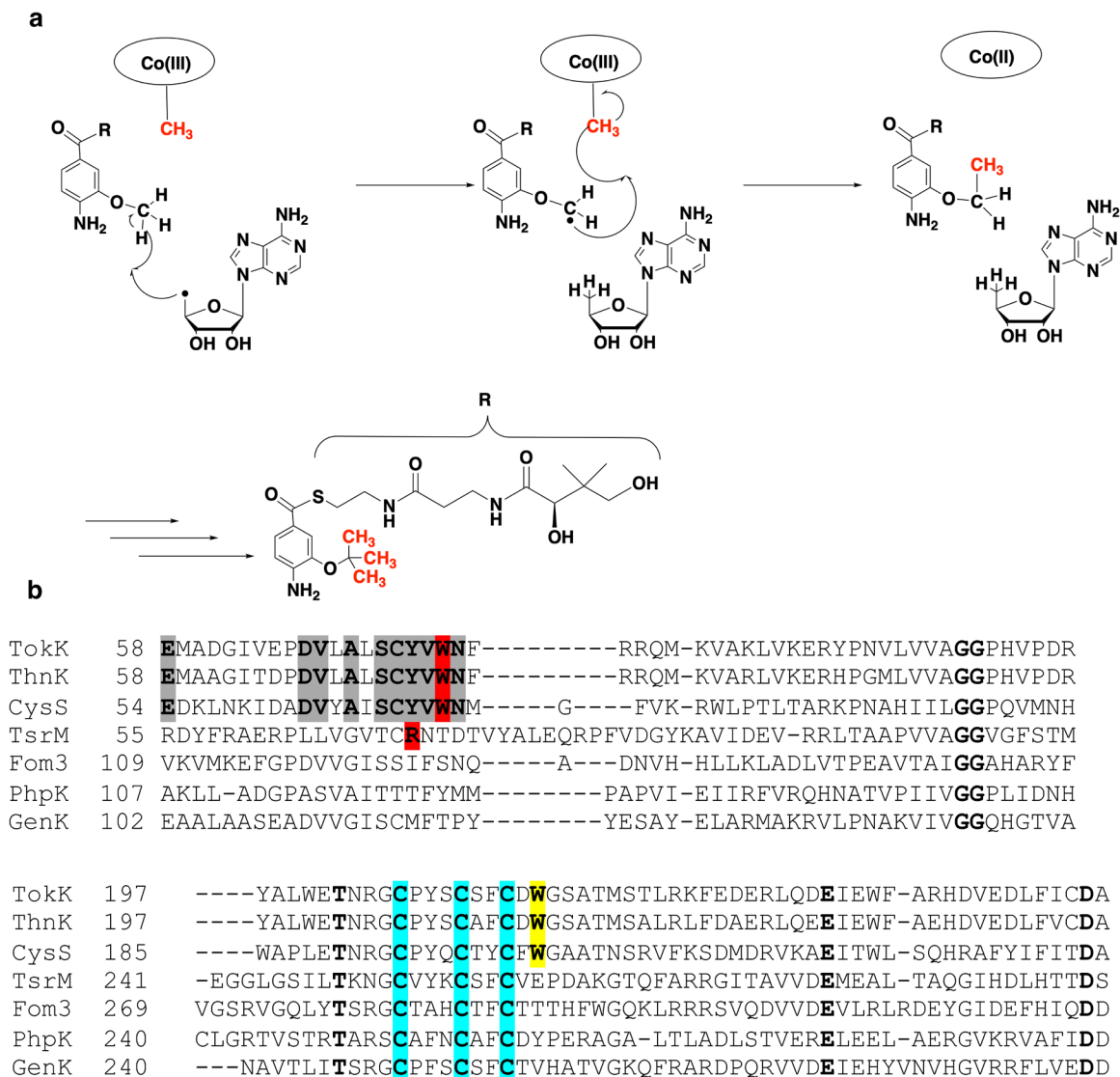
### Construction of TokK Variants

TokK variants were generated by overlap extension PCR using pET29b:*tokK*Tev as a template and primers described in Table S1. 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.

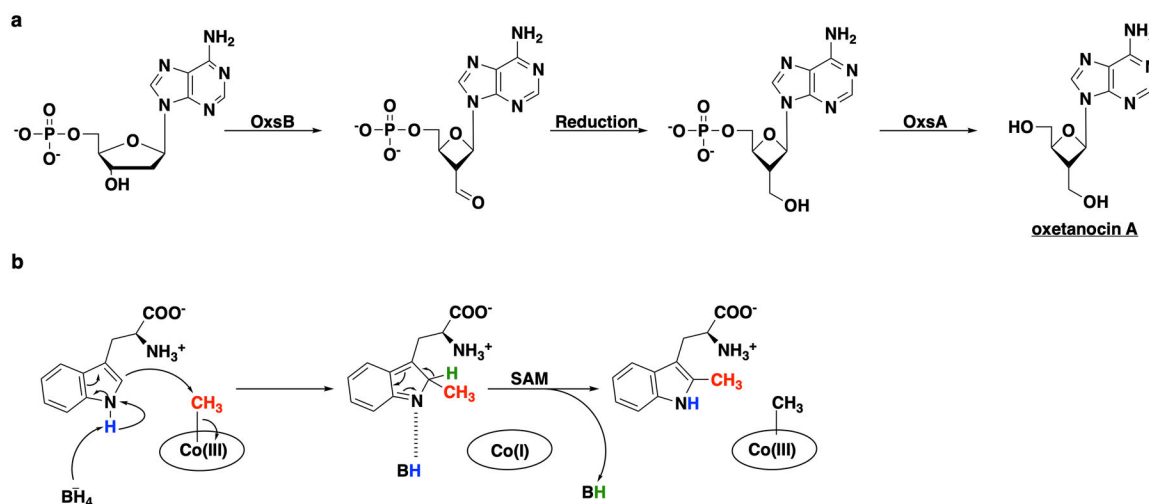
### Methylation Assays of TokK Variants

Expression and purification of TokK variants was carried out as previously described for wild-type TokK<sup>19</sup>, except reconstitution<sup>31</sup> 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 methylcobalamin, 100  $\mu$ M substrate, and 100  $\mu$ M enzyme. At each time point, an aliquot of the reaction mixture was diluted 5 $\times$ , filtered through a 10 kDa Amicon ultrafiltration device, and analyzed for product formation using UPLC-HRMS as previously described<sup>19</sup>. First order rate constants were determined using the COPASI parameter estimation tool<sup>57</sup>, and curves were simulated using Vcell<sup>58</sup>.

## Extended Data

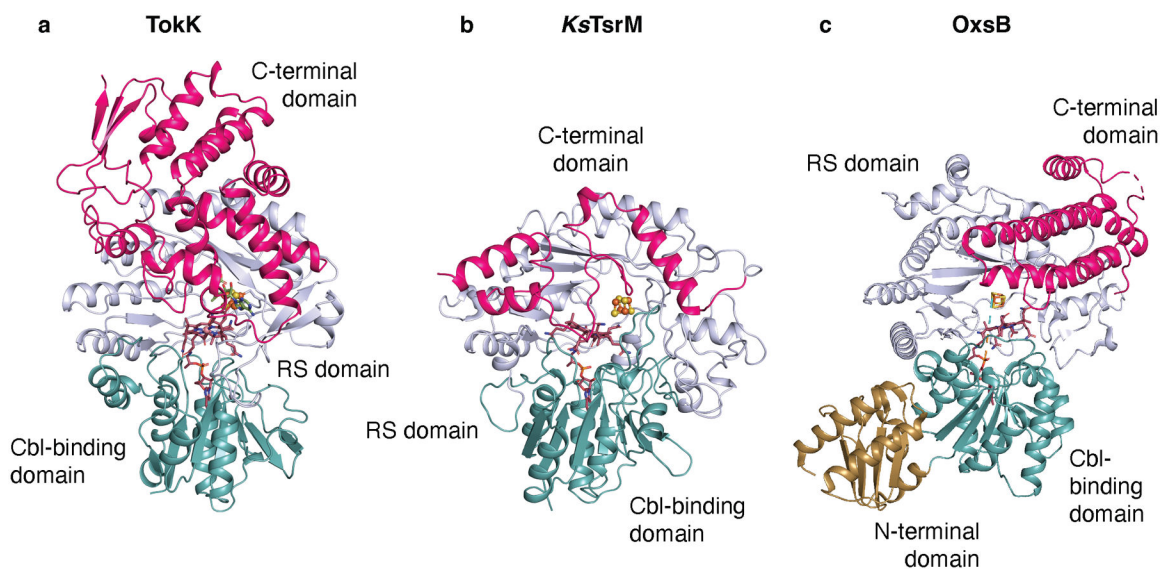
**Extended Data Fig. 1. Comparison to CysS, another class B sequential methylase.**

CysS performs sequential radical methylations like TokK and ThnK (a). 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) (b). 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 cobalamin-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.



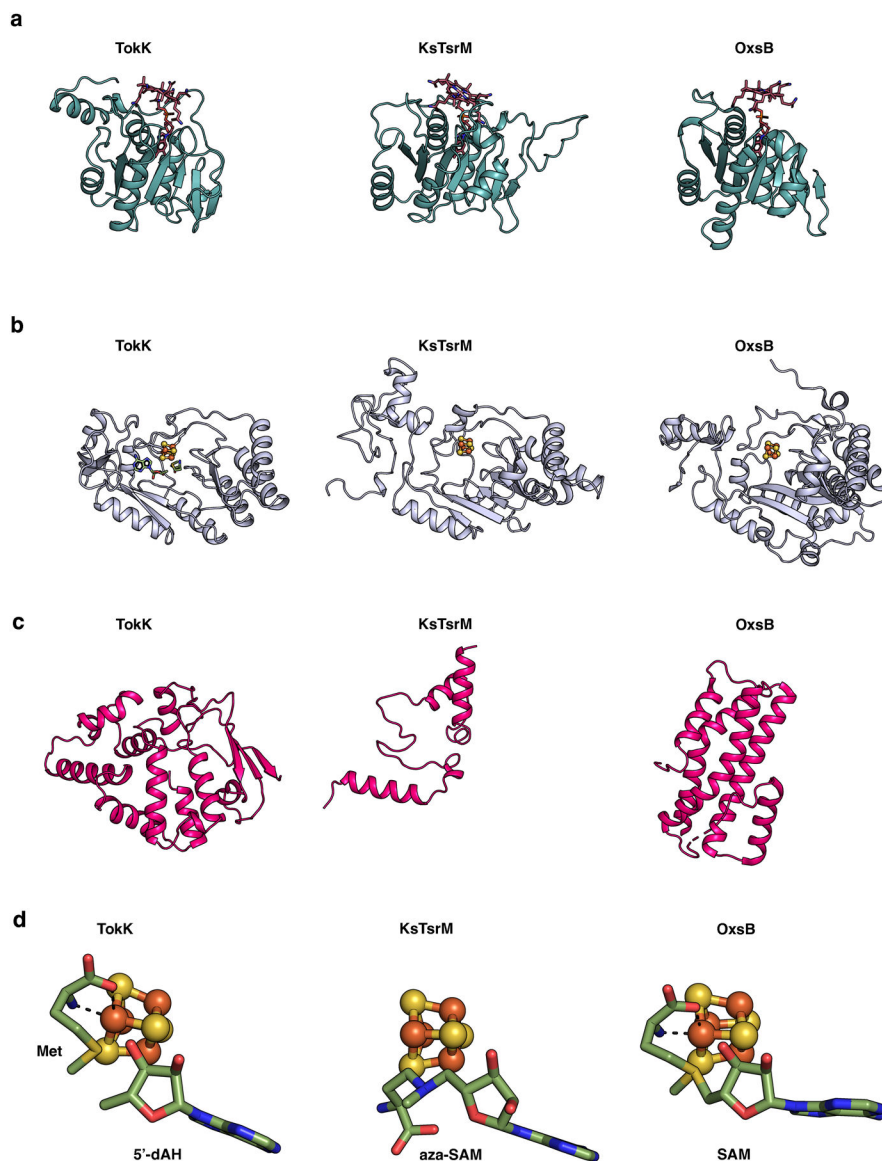
**Extended Data Fig. 2. Reactions performed by OxsB and TsrM, notable Cbl-binding radical SAM enzymes.**

(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.



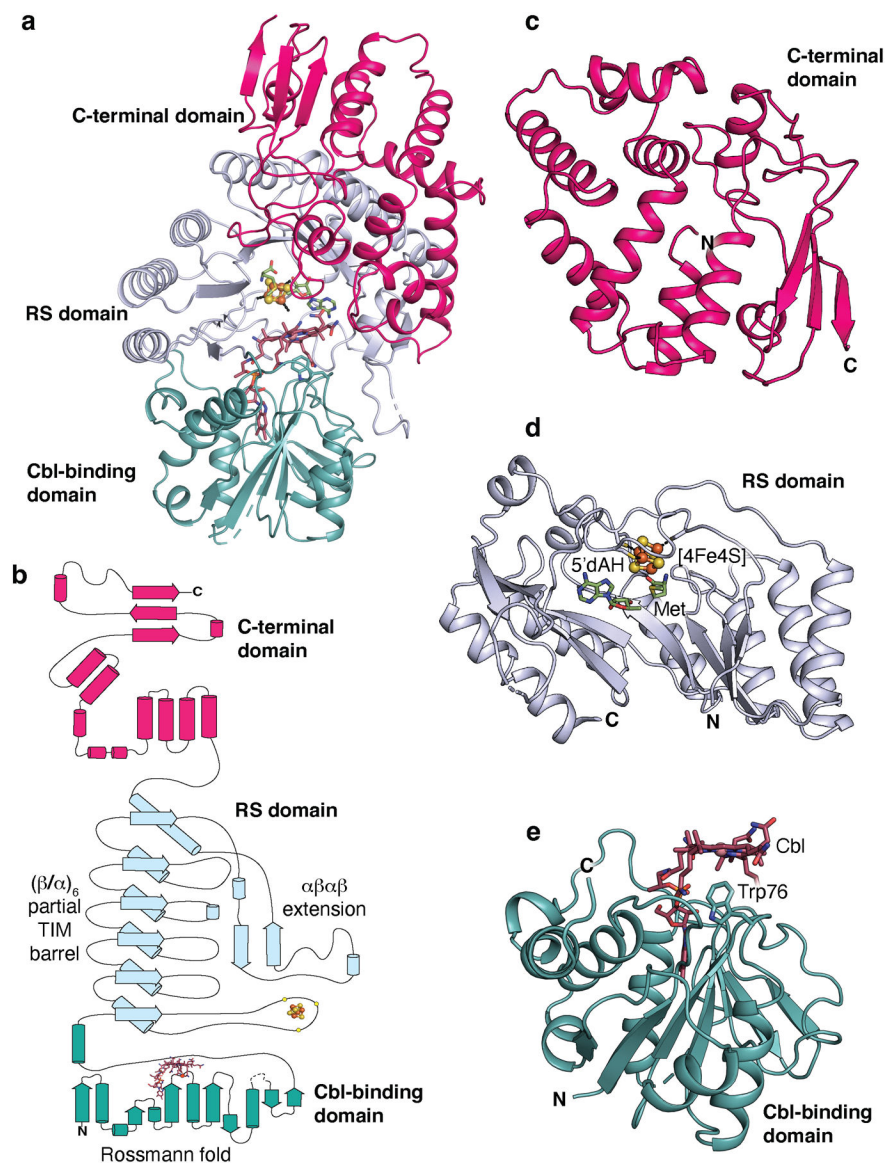
**Extended Data Fig. 3. Structural comparisons of Cbl-dependent RS enzymes.**

The overall structures of TokK (a), KsTsrM (PDB ID: 6WTE) (b), and OxsB (PDB ID: 5UL3) (c) are shown as ribbon diagrams and colored 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.



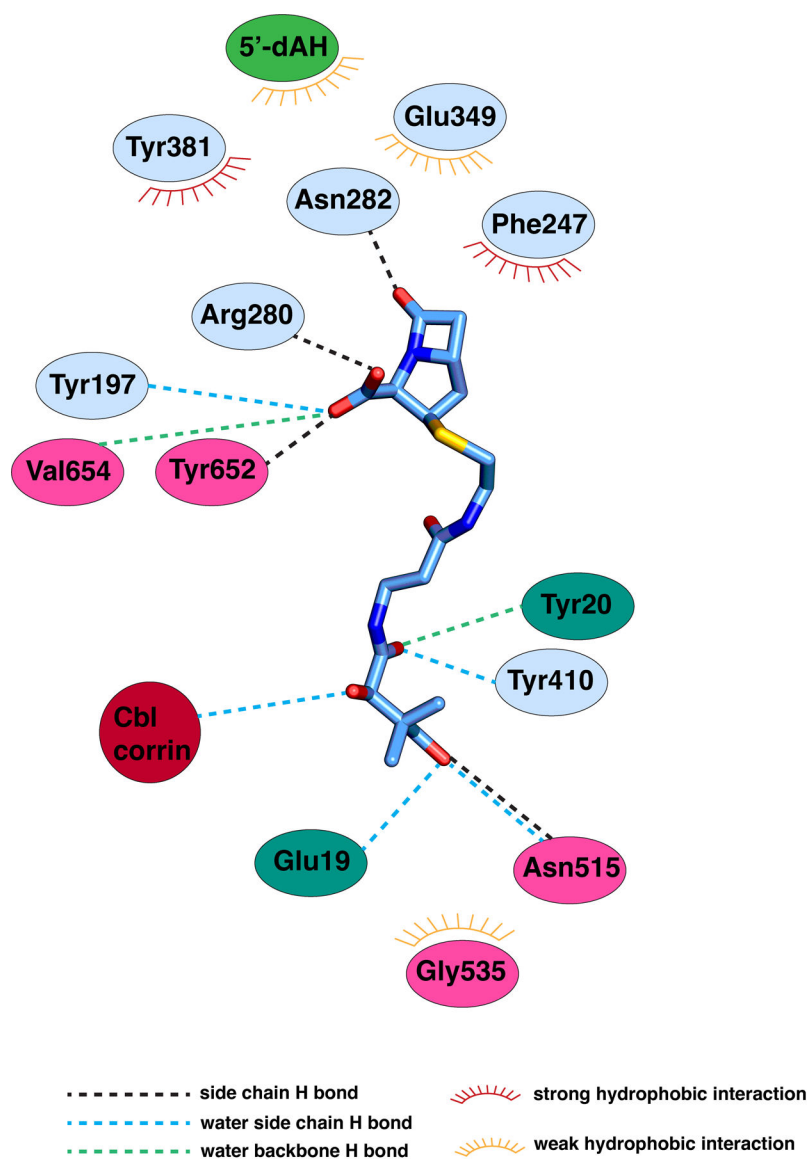
**Extended Data Fig. 4. Comparison of the Cbl-binding, RS, and C-terminal domains of TokK, KsTsrM, and OxsB.**

The domains are colored as in Extended Data Fig. 3. (a) The Rossmann fold, in teal, is highly similar among TokK (PDB ID: 7KDY), KsTsrM (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 KsTsrM, and all three have unique extra secondary structure features. (c) The C-terminal domains for TokK, KsTsrM, and OxsB are vastly different in architecture. (d) Comparison of the binding of Met and 5'-dAH, aza-SAM, and SAM for TokK, KsTsrM, 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.



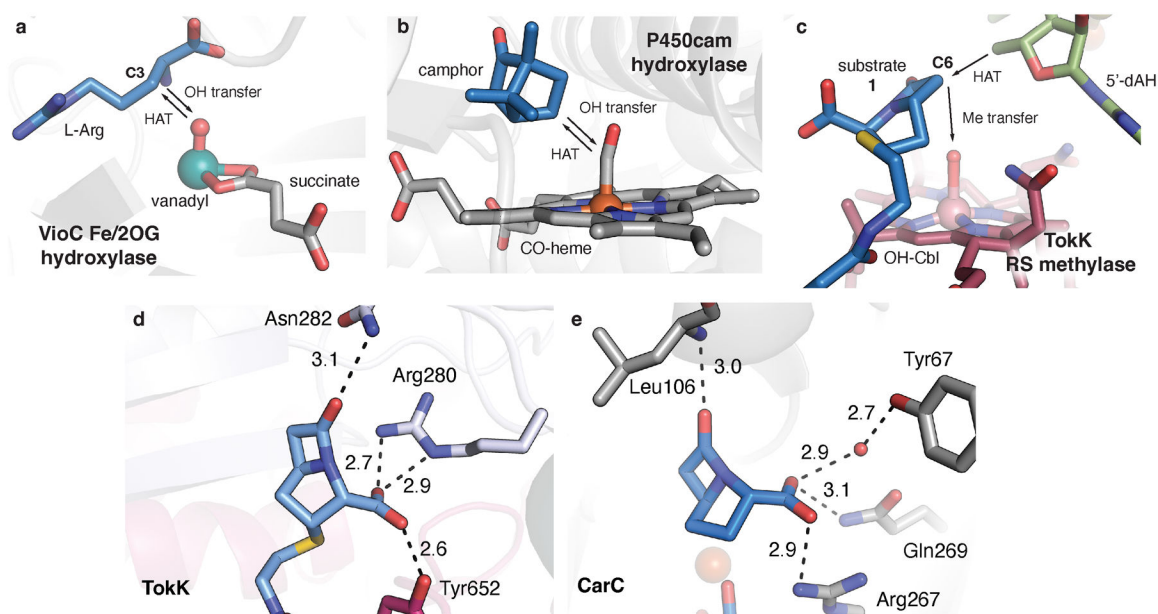
**Extended Data Fig. 5.**

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 colored as in panel A. The cobalamin 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. Zoomed in views of the C-terminal domain (c), RS domain (d), and Cbl-binding domain (e) are shown as ribbon diagrams.



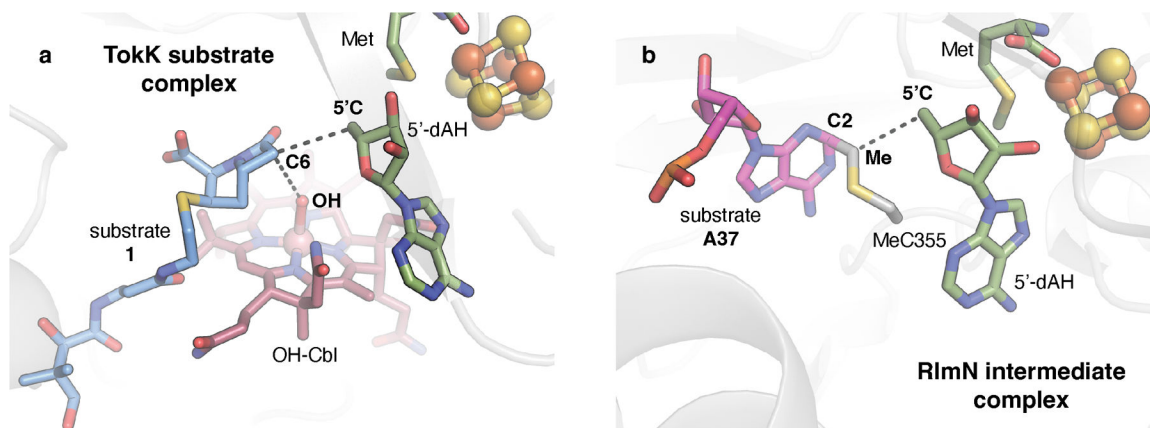
**Extended Data Fig. 6. Diagram of carbapenam interactions in the TokK complex with substrate.** Carbapenam substrate **1** shown in blue sticks and colored by atom type. Polar and hydrophobic interactions were mapped with the LigPlot program and indicated with dashed lines or a starburst symbol.





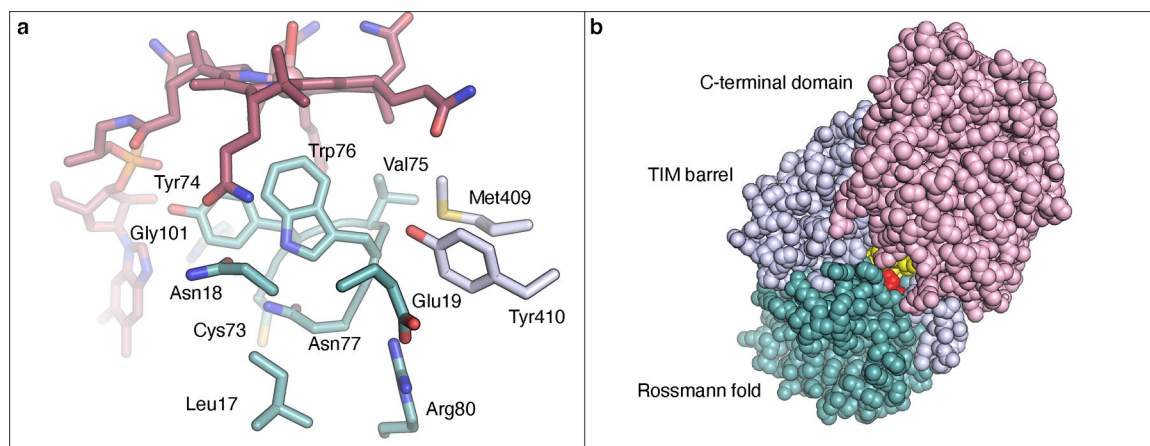
**Extended Data Fig. 7. Comparison of TokK structure with those of enzymes using alternative radical-generating mechanisms.**

Residues that interact with the polar substituents in the  $\beta$ -lactam ring are shown in stick format for TokK (a) and the epimerase CarC (PDB ID: 4OJ8) (b). While the two enzymes are structurally distinct, they use a similar number and type of functional groups to anchor the  $\beta$ -lactam by using direct and water-mediated contacts to the C7 carbonyl and C3 carboxylate substituents. Additionally, comparison of the active site with those of hydroxylases reveals differences in the orientation of hydrogen atom transfer (HAT) intermediates and  $-\text{OH}$  or  $-\text{CH}_3$  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  $-\text{OH}$  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 heme-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  $-\text{OH}$  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 ( $5'-\text{dA}\bullet$ ) 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  $\sim 120^\circ$  angle between the HAT acceptor ( $5'\text{C}$  of  $5'-\text{dAH}$ ), the substrate target carbon (C6), and the Cbl ligand ( $-\text{OH}$  of OH-Cbl, a surrogate for Me-Cbl).



**Extended Data Fig. 8. A comparison of the substrate-binding sites in two structurally characterized RS methylases.**

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 ( $\sim 6$  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 (colored yellow) is solvent accessible, with most of the Cbl being buried within the Rossmann fold. Trp76 is colored red.

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

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

	5'-dAH•Mef•TokK Fe anomalous (peak)	5'-dAH•Met•TokK native	Substrate-bound 5'- dAH•Met•TokK native
<b>Data collection</b>			
Space group	$P2_1$	$P2_1$	$P2_1 2_1 2_1$
Wavelength (Å)	1.72194	1.03313	1.03320
Cell dimensions			
<i>a, b, c</i> (Å)	78.18, 138.58, 78.18	65.64, 137.74, 77.91	84.50, 132.82, 155.51
$\alpha, \beta, \gamma$ (°)	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
$R_{\text{sym}}$ or $R_{\text{merge}}$	0.08 (0.42)	0.08 (0.84)	0.11 (0.59)
$R_{\text{pim}}$	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)
<b>Refinement</b>			
Resolution (Å)		1.79	1.94
No. reflections		130,677	129,773
$R_{\text{work}} / R_{\text{free}}$		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 <sup>2</sup>			
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	-

	5'-dAH•Mef•TokK Fe anomalous (peak)	5'-dAH•Met•TokK native	Substrate-bound 5'- dAH•Met•TokK native
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	0.006	0.007
Bond angles (°)	1.11	1.11	1.12
Clashscore	4.49	4.49	3.06
Rotamer outliers (%)	0.27	0.27	0.09
Ramachandran			
Most favored (%)	97.41	97.41	97.73
Allowed (%)	2.59	2.59	2.27
Outliers (%)	0.00	0.00	0.00
Number of TLS groups	1	1	1
PDB accession code	7KDX	7KDX	7KDY

\* Values in parentheses are for highest-resolution shell.

\*\* all structures result from one single crystal

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

### Funding Sources

This work was supported by NIH (GM122595 to SJB, GM119707 to AKB, AI121072 to CAT and GM080189 to EKS), 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 U.S. 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 National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169. EKS and CAT thank M. S. Lichstrahl for providing a synthetic intermediate, and Drs. I. P. Mortimer and J. Tang for their help with ESI-MS and NMR experiments, respectively. We thank Dr. David Iwig for his help in the MS analysis of the TokK crystals.

## 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).

## REFERENCES

1. US Food and Drug Administration, Supplemental New Drug Approval: RECARBRIO (imipenem, cilastatin, and relebactam) for injection, available at <https://www.fda.gov/RegulatoryInformation/Guidances/default.htm> (2020).
2. Fisher JF & Mobashery S Three decades of the class A beta-lactamase acyl-enzyme. *Curr Protein Pept Sci* 10, 401–407, doi:10.2174/138920309789351967 (2009). [PubMed: 19538154]
3. Sinner EK, Marous DR, Townsend CA Evolution of Methods for the Study of Cobalamin-Dependent Radical SAM Enzymes. *ACS Bio & Med Chem Au* 0, 10.1021/acsbiochemau.1021c00032, doi:10.1021/acsbiochemau.1c00032 (2021).
4. Li R, Lloyd EP, Moshos KA & Townsend CA Identification and characterization of the carbapenem MM 4550 and its gene cluster in *Streptomyces argenteolus* ATCC 11009. *Chembiochem* 15, 320–331, doi:10.1002/cbic.201300319 (2014). [PubMed: 24420617]
5. Nunez LE, Mendez C, Brana AF, Blanco G & Salas JA The biosynthetic gene cluster for the beta-lactam carbapenem thienamycin in *Streptomyces cattleya*. *Chem Biol* 10, 301–311, doi:10.1016/s1074-5521(03)00069-3 (2003). [PubMed: 12725858]
6. Sinner EK, Marous DR, Townsend CA Evolution of Methods for the Study of Cobalamin-Dependent Radical SAM Enzymes. *ACS Bio & Med Chem Au* 10.1021/acsbiochemau.1c00032, doi:10.1021/acsbiochemau.1c00032 (2021).
7. Coulthurst SJ, Barnard AM & Salmond GP Regulation and biosynthesis of carbapenem antibiotics in bacteria. *Nat Rev Microbiol* 3, 295–306, doi:10.1038/nrmicro1128 (2005). [PubMed: 15759042]
8. Wang Y, Schnell B, Baumann S, Muller R & Begley TP Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM Enzyme. *J Am Chem Soc* 139, 1742–1745, doi:10.1021/jacs.6b10901 (2017). [PubMed: 28040895]
9. Broderick JB, Duffus BR, Duschene KS & Shepard EM Radical S-Adenosylmethionine Enzymes. *Chemical Reviews* 114, 4229–4317, doi:10.1021/cr4004709 (2014). [PubMed: 24476342]
10. Holliday GL et al. Atlas of the radical SAM superfamily: divergent evolution of function using a “plug and play” domain. *Methods Enzymol* 606, 1–71, doi:10.1016/bs.mie.2018.06.004 (2018). [PubMed: 30097089]
11. Wang SC Cobalamin-dependent radical S-adenosyl-l-methionine enzymes in natural product biosynthesis. *Nat. Prod. Rep* 35, 707–720 (2018). [PubMed: 30079906]
12. Miller DV et al. in *Comprehensive Natural Products III: Chemistry and Biology Vol. 5* (ed H.-W. and Begley Liu TP) 24–69 (Elsevier, 2020).
13. Knox HL et al. Structural basis for non-radical catalysis by TsrM, a radical SAM methylase. *Nat Chem Biol*, doi:10.1038/s41589-020-00717-y (2021).
14. Pierre S et al. Thiostrepton tryptophan methyltransferase expands the chemistry of radical SAM enzymes. *Nat. Chem. Biol* 8, 957–959, doi:10.1038/nchembio.1091 (2012). [PubMed: 23064318]
15. Blaszczyk AJ 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, doi:10.1021/jacs.5b12592 (2016). [PubMed: 26841310]
16. Bridwell-Rabb J, Zhong A, Sun HG, Drennan CL & Liu HW A B12-dependent radical SAM enzyme involved in oxetanocin A biosynthesis. *Nature* 544, 322–326, doi:10.1038/nature21689 (2017). [PubMed: 28346939]
17. Blaszczyk AJ, Wang B, Silakov A, Ho JV & Booker SJ 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). [PubMed: 28747433]
18. Chang W. c. et al. Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science* 343, 1140–1144 (2014). [PubMed: 24604200]
19. Sinner EK, Lichstrahl MS, Li R, Marous DR & Townsend CA Methylations in complex carbapenem biosynthesis are catalyzed by a single cobalamin-dependent radical S-adenosylmethionine enzyme. *Chem Commun (Camb)* 55, 14934–14937, doi:10.1039/c9cc07197k (2019). [PubMed: 31774078]
20. Albers-Schoenberg GA, Byron H; Hensens Otto D.; Hirshfield Jordan; Hoogsteen Karst; Kaczka Edward A.; Rhodes Robert E.; Kahan Jean S.; Kahan Frederick M.; Ratcliff Ronald W.; Walton

Edward; Ruswinkle Linda J.; Morin Robert B.; and Christensen Burton G. Structure and absolute configuration of thienamycin. *J Am Chem Soc* 100, 6491–6499, doi:DOI: 10.1021/ja00488a038 (1978).

21. Mitchell AJ et al. Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-glutarate-Dependent Hydroxylase. *J Am Chem Soc* 139, 13830–13836, doi:10.1021/jacs.7b07374 (2017). [PubMed: 28823155]
22. Poulos TL, Finzel BC & Howard AJ High-resolution crystal structure of cytochrome P450cam. *J Mol Biol* 195, 687–700, doi:10.1016/0022-2836(87)90190-2 (1987). [PubMed: 3656428]
23. Schwalm EL, Grove TL, Booker SJ & Boal AK Crystallographic capture of a radical S-adenosylmethionine enzyme in the act of modifying tRNA. *Science* 352, 309–312, doi:10.1126/science.aad5367 (2016). [PubMed: 27081063]
24. Drennan CL, Huang S, Drummond JT, Matthews RG & Ludwig ML How a Protein Binds B-12 - a 3.0-Angstrom X-Ray Structure of B-12-Binding Domains of Methionine Synthase. *Science* 266, 1669–1674, doi:DOI 10.1126/science.7992050 (1994). [PubMed: 7992050]
25. Kräutler B Thermodynamic *trans*-effects of the nucleotide base in the B12 coenzymes. *Helv. Chim. ACTA* 70, 1268–1278 (1987).
26. Krautler B Chemistry of methylcorrinoids related to their roles in bacterial C1 metabolism. *FEMS Microbiol Rev* 7, 349–354, doi:10.1111/j.1574-6968.1990.tb04935.x (1990). [PubMed: 2094289]
27. Milton PA & Brown TL The kinetics of benzimidazole dissociation in methylcobalamin. *J Am Chem Soc* 99, 1390–1396, doi:10.1021/ja00447a018 (1977). [PubMed: 838998]
28. Marous DR et al. Consecutive radical S-adenosylmethionine methylations form the ethyl side chain in thienamycin biosynthesis. *Proc. Natl. Acad. Sci* 112, 10354–10358, doi:10.1073/pnas.1508615112 (2015). [PubMed: 26240322]
29. Lanz ND et al. Enhanced Solubilization of Class B Radical S-Adenosylmethionine Methylases by Improved Cobalamin Uptake in *Escherichia coli*. *Biochemistry* 57, 1475–1490, doi:10.1021/acs.biochem.7b01205 (2018). [PubMed: 29298049]
30. Lanz ND et al. RImN and AtsB as Models for the Overproduction and Characterization Radical SAM Proteins. *Natural Product Biosynthesis by Microorganisms and Plants, Pt B* 516, 125–152, doi:10.1016/B978-0-12-394291-3.00030-7 (2012).
31. Blaszczyk AJ, Wang RX & Booker SJ TsrM as a Model for Purifying and Characterizing Cobalamin-Dependent Radical S-Adenosylmethionine Methylases. *Methods Enzymol.* 595, 303–329, doi:10.1016/bs.mie.2017.07.007 (2017). [PubMed: 28882204]
32. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* 66, 213–221, doi:10.1107/S0907444909052925 (2010).
33. Bunkoczi G et al. Phaser.MRage: automated molecular replacement. *Acta Crystallographica Section D-Biological Crystallography* 69, 2276–2286, doi:10.1107/S0907444913022750 (2013).
34. 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 Crystallographica Section D-Biological Crystallography* 62, 859–866, doi:10.1107/S0907444906019949 (2006).
35. Otwinowski Z & Minor W Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* 276, 307–326, doi:Doi 10.1016/S0076-6879(97)76066-X (1997).
36. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* 66, 486–501, doi:10.1107/S0907444910007493 (2010).
37. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Structural Biology* 66, 12–21, doi:10.1107/S0907444909042073 (2010).
38. The PyMOL Molecular Graphics Systems, Version 2.0 Schrödinger, LLC.
39. Ho BK & Gruswitz F HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* 8, 49, doi:10.1186/1472-6807-8-49 (2008). [PubMed: 19014592]

40. Wallace AC, Laskowski RA & Thornton JM LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8, 127–134, doi:10.1093/protein/8.2.127 (1995). [PubMed: 7630882]
41. Schuttelkopf AW & van Aalten DM PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60, 1355–1363, doi:10.1107/S0907444904011679 (2004). [PubMed: 15272157]
42. Akiva E et al. The Structure-Function Linkage Database. *Nucleic Acids Research* 42, D521–D530, doi:10.1093/nar/gkt1130 (2014). [PubMed: 24271399]
43. Gerlt JA Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-Function Space and Genome Context to Discover Novel Functions. *Biochemistry* 56, 4293–4308, doi:10.1021/acs.biochem.7b00614 (2017). [PubMed: 28826221]
44. Kim HJ 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, doi:10.1021/ja312641f (2013). [PubMed: 23679096]
45. Huang C et al. Delineating the biosynthesis of gentamicin x2, the common precursor of the gentamicin C antibiotic complex. *Chem Biol* 22, 251–261, doi:10.1016/j.chembiol.2014.12.012 (2015). [PubMed: 25641167]
46. Werner WJ et al. In Vitro Phosphinate Methylation by PhpK from *Kitasatospora phosalacinea*. *Biochemistry* 50, 8986–8988, doi:10.1021/bi201220r (2011). [PubMed: 21950770]
47. 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, doi:10.1021/acs.biochem.8b00693 (2018). [PubMed: 30036047]
48. Parent A et al. The B12-Radical SAM Enzyme PoyC Catalyzes Valine C $\beta$ -Methylation during Polytheonamide Biosynthesis. *J. Am. Chem. Soc* 138, 15515–15518, doi:10.1021/jacs.6b06697 (2016). [PubMed: 27934015]
49. 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, doi:10.1021/ja902261a (2009). [PubMed: 19514719]
50. Suzuki JY, Bollivar DW & Bauer CE Genetic analysis of chlorophyll biosynthesis. *Annu Rev Genet* 31, 61–89, doi:10.1146/annurev.genet.31.1.61 (1997). [PubMed: 9442890]
51. Westrich L, Heide L & Li SM CloN6, a novel methyltransferase catalysing the methylation of the pyrrole-2-carboxyl moiety of clorobiocin. *ChemBioChem* 4, 768–773, doi:10.1002/cbic.200300609 (2003). [PubMed: 12898629]
52. Radle MI, Miller DV, Laremore TN & Booker SJ Methanogenesis marker protein 10 (Mmp10) from *Methanosarcina acetivorans* is a radical S-adenosylmethionine methylase that unexpectedly requires cobalamin. *Journal of Biological Chemistry* 294, 11712–11725, doi:10.1074/jbc.RA119.007609 (2019).
53. Gerlt JA 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, doi:10.1016/j.bbapap.2015.04.015 (2015). [PubMed: 25900361]
54. Zallot R, Oberg N & Gerlt JA 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, doi:10.1021/acs.biochem.9b00735 (2019). [PubMed: 31553576]
55. Zallot R, Oberg NO & Gerlt JA 'Democratized' genomic enzymology web tools for functional assignment. *Curr Opin Chem Biol* 47, 77–85, doi:10.1016/j.cbpa.2018.09.009 (2018). [PubMed: 30268904]
56. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504, doi:10.1101/gr.1239303 (2003). [PubMed: 14597658]
57. Hoops S et al. COPASI--a COMplex PATHway SIMulator. *Bioinformatics* 22, 3067–3074, doi:10.1093/bioinformatics/btl485 (2006). [PubMed: 17032683]

58. Schaff J, Fink CC, Slepchenko B, Carson JH & Loew LM A general computational framework for modeling cellular structure and function. *Biophys J* 73, 1135–1146, doi:10.1016/S0006-3495(97)78146-3 (1997). [PubMed: 9284281]

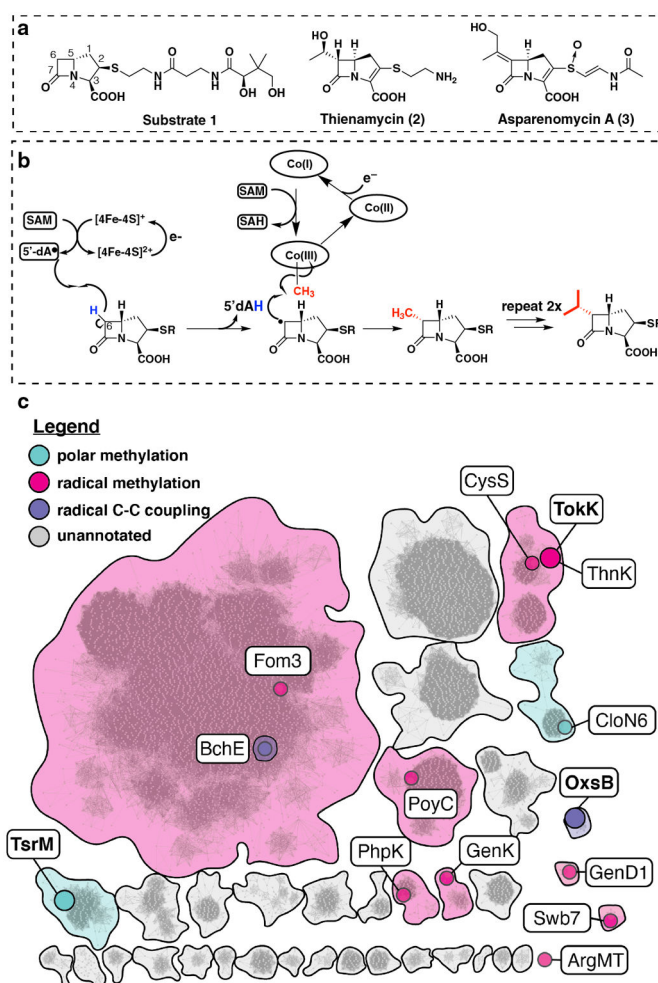
Author Manuscript

Author Manuscript

Author Manuscript

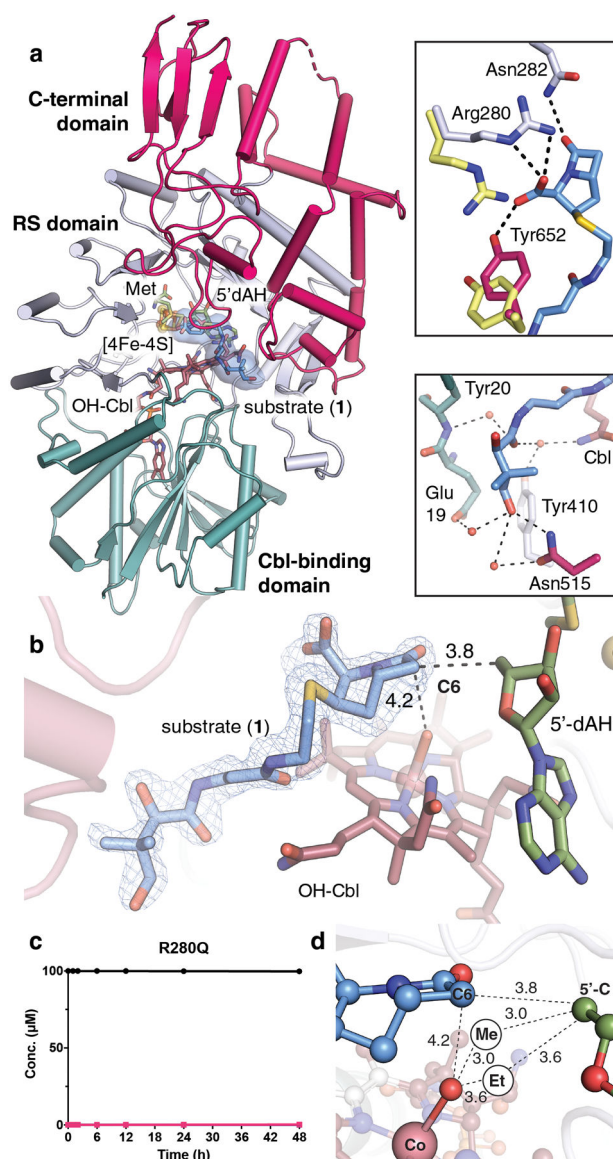
Author Manuscript





**Fig 1. Cobalamin-dependent radical-mediated methylations in carbapenem biosynthesis.**

(a) Structure of (2*R*) pantetheinylated carbapenem 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 catalyzed by TokK. (c) An abbreviated sequence similarity network (SSN) of cobalamin-binding RS enzymes, highlighting selected sequence clusters and nodes. The full network (Fig. S1) was generated from ~11,000 annotated cobalamin-dependent RS enzymes with an alignment score of 65. Each node represents a single sequence or a set of sequences with >40% sequence identity. Sequence clusters and nodes are colored by predicted reaction mechanism (see legend in figure panel). Nodes containing functionally annotated sequences are indicated in color. Structurally characterized enzymes are represented by enlarged nodes and labeled in boldface.



**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 colored by domain. The Cbl-binding Rossmann fold is shown in teal with the cobalamin cofactor in stick format and colored 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, colored by atom type. (b) An  $F_o - F_c$  omit electron density map is shown for (**1**) (blue mesh, contoured at  $3.0\sigma$ ). 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. 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

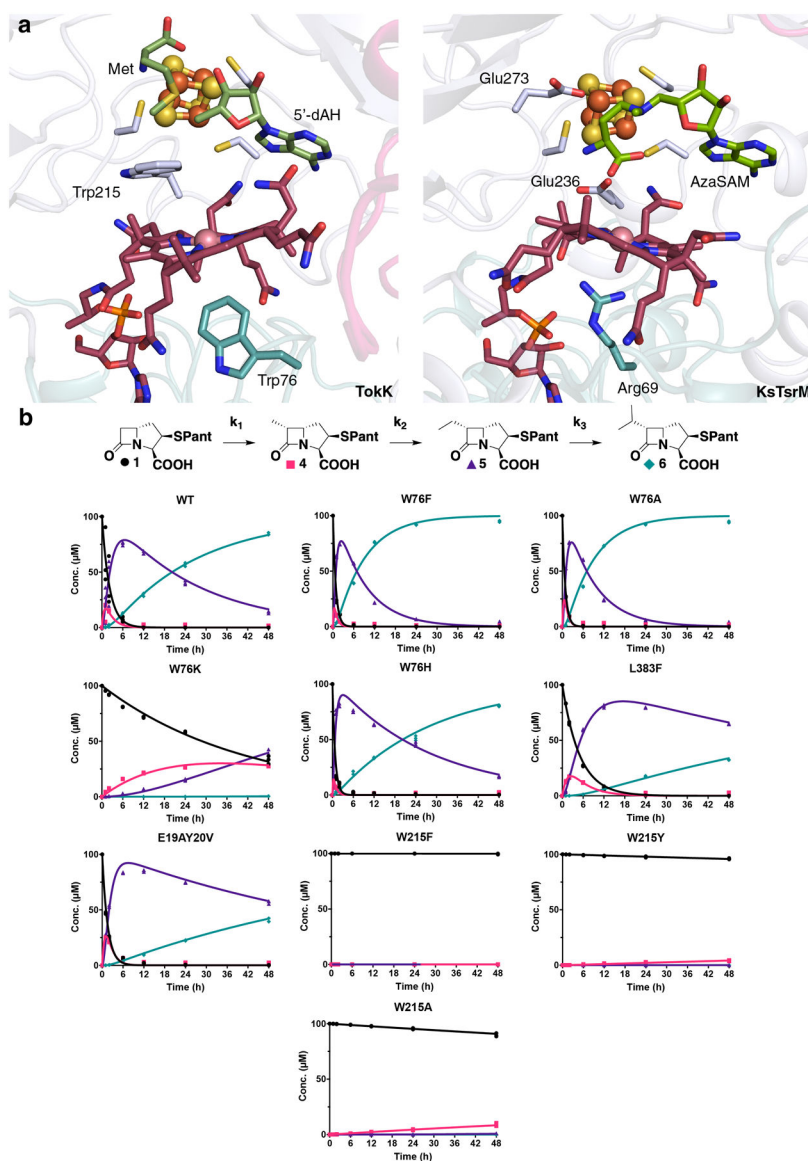
of OHCbl, the latter of which serves as a surrogate of the active MeCbl cofactor. Spheres labeled Me and Et represent the suggested positions of the newly installed carbon atoms in the mono-methylated and dimethylated products.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**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 cobalamin-binding RS enzymes, TokK and *KsTsrM* (PDB ID: 6WTF). Selected amino acid side chains are shown in stick format and the Co ion is shown as a pink sphere. (b) Schematic diagram of the three sequential methylations performed by TokK. 48 h time course experiments performed in triplicate (each replicate represented as a symbol) tracking substrate (black spheres), methyl (pink squares), ethyl (purple triangles), and isopropyl (blue diamonds) product formation. 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, W76F, W76A, W76K, W76H, L383F, E19AY20V, W215F, W215Y, and W215A.