

Caerulomycin and Collismycin Antibiotics Share a *trans* Flavin-Dependent Assembly Line for 2,2'-Bipyridine Formation and Sulfur Fate Differentiation

Bo Pang,^{1,#} Rijin Liao,^{2,#} Zhijun Tang,^{1,#} Shengjie Guo,¹ Zhuhua Wu,¹ and Wen Liu^{1,3,*}

¹ State Key Laboratory of Bioorganic and Natural Products Chemistry, Center for Excellence in Molecular Synthesis, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China;

² Shanghai Institute of Precision Medicine, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200125, China;

³ Huzhou Center of Bio-Synthetic Innovation, 1366 Hongfeng Road, Huzhou 313000, China;

* To whom correspondence should be addressed: Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Rd., Shanghai 200032, China. Wen Liu, Email: wliu@sioc.ac.cn, Tel: 86-21-54925111, Fax: 86-21-64166128

These authors equally contribute to this work.

ABSTRACT:

Linear nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) template the modular biosynthesis of numerous nonribosomal peptides, polyketides and their hybrids through assembly line chemistry. This chemistry can be complex and highly varied, and thus challenges the understanding in the diverse polymerization processes of amino acid and carboxylate monomers programmed by various NRPSs and PKSs in nature. Here, we report that caerulomycin and collismycin peptide-polyketide hybrid antibiotics share an unusual assembly line that involves NRPS activity to recruit a flavoprotein acting *in trans* and catalyze C-C bond formation and heterocyclization during 2,2'-bipyridine formation. Simultaneously, this assembly line provides dethiolated and thiolated 2,2'-bipyridine intermediates through differential treatment of the sulfhydryl group arising from L-cysteine incorporation. Subsequent L-leucine extension, which does not contribute any atoms to either caerulomycins or collismycins, plays a key role in sulfur fate determination by selectively advancing one of the two 2,2'-bipyridine intermediates down a path to the final products with or without sulfur decoration. These findings further the appreciation of assembly line chemistry and will facilitate the development of related molecules using synthetic biology approaches.

INTRODUCTION

Linear nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) are often large molecular machines that are composed of multidomain modules¹⁻⁶. They have evolved through functional unit (e.g., protein, module or domain) combination and variation to afford various assembly lines that program diverse polymerization and modification processes using amino acids, short carboxylic acids or both as monomers. Assembly line chemistry can be complex and highly varied, and thereby challenges our understanding of how nature creates numerous nonribosomal peptides, polyketides and hybrids thereof that historically play critical roles in medicinal chemistry and chemical biology. The biosynthesis of 2,2'-bipyridine, the core structure of a large class of synthetic bidentate metal-chelating ligands with a variety of applications in many areas of chemistry^{7,8}, occurs via such a process.

2,2'-Bipyridine natural products include caerulomycins (e.g., CAE-A, **Fig. 1**), a group of structurally related antibiotics that possess potent immunosuppressive activity⁹⁻¹³. While the process through which 2,2'-bipyridine is formed remains poorly understood, we and others unraveled the biogenesis of CAEs and identified a NRPS-PKS assembly line that contains three modular proteins, i.e., CaeA1, CaeA2 and CaeA3, to template a peptide-polyketide hybrid skeleton for 2,2'-bipyridine formation (**Fig. 1**)¹⁴⁻¹⁶. CaeA1, a bifunctional enzyme composed of a peptidyl carrier protein (PCP) and an adenylation (A) domain, was supposed to prime picolinyl, which provides the unmodified pyridine unit (Ring **B**) of CAEs, as the starter unit for molecular assembly. CaeA2 is a hybrid protein, with a typical PKS module containing a ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) and an

atypical NRPS module containing a condensation/cyclization (Cy) domain, A domain, PCP and terminal C (Ct) domain. Logically, this protein can incorporate malonyl-*S*-CoA and L-cysteine in tandem and associate CaeA1 to provide the essential building atoms for constructing the di- or trisubstituted pyridine unit (Ring A) of CAEs^{14,17,18}. As proposed, the PKS module of CaeA2 catalyzes two-carbon elongation and installs the atoms C3 and C4 following C2, the carboxyl carbon of the starter picoliny unit. The atypical NRPS then catalyzes L-cysteine extension, which is unique and occurs at C β of the L-cysteine residue to form a C-C bond, instead of its α -amino group for amide bond formation as usual, and thus provides C5, C6 and N1 as well as exocyclic C7 after heterocyclization. To our knowledge, peptidyl elongation through a C-C linkage lacks precedent in NRPS catalysis. The domains of CaeA3 are organized as C-A-PCP for L-leucine extension^{14,15}. Despite not contributing any atoms in the mature products, CaeA3 activity is necessary for the formation of 2,2'-bipyridinyl-L-leucine (**1**), an off-line intermediate that can be specialized to individual CAE members after L-leucinyl removal¹⁹⁻²³. If **1** is the direct product of CaeA1, CaeA2 and CaeA3, 2,2'-bipyridine formation could occur in this assembly line through an unprecedented dethiolation-coupled heterocyclization process that involves NRPS-mediated C-C bond formation.

Interestingly, homologs of the above modular synth(et)ases CaeA1, CaeA2 and CaeA3 (i.e., ColA1a and ColA1b, ColA2 and ColA3) were observed in the biosynthetic pathway of collismycins (e.g., COL-A, **Fig. 1**)^{14,24,25}, the 2,2'-bipyridine antibiotics differing from CAEs with a sulfur-containing group at C5 of Ring A. Most likely, COLs share a similar assembly line for 2,2'-bipyridine formation in which, however, the dethiolation step can be skipped to retain the sulfur-containing group during assembly of the completely identical substrates picolinic acid, malonyl-*S*-CoA, L-cysteine and L-

leucine¹⁶. With great interest in the catalytic logic and the mechanism for sulfur fate determination during 2,2'-bipyridine formation, in this study, we reconstituted the assembly line for the dethiolated antibiotics CAEs, validated the generality of 2,2'-bipyridine formation in the biosynthesis of the thiolated antibiotics COLs and experimentally dissected the on-line process for 2,2'-bipyridine formation and differentiation. The findings reported here exemplify the complexity and variety of assembly line chemistry, which has long been an area of intense interest in the creation of structurally diverse nonribosomal peptides, polyketides and their hybrid molecules.

RESULTS & DISCUSSION

2,2'-Bipyridine assembly line functions with a *trans* flavoprotein

We began by reconstituting the activities of the didomain NRPS CaeA1 (650-aa), the PKS-NRPS hybrid CaeA2 (2484-aa) and the single-module NRPS CaeA3 (1047-aa) *in vitro* (**Fig. 2**). These multifunctional enzymes were individually produced in an engineered *Escherichia coli* strain expressing Sfp^{26,27}, a phosphopantetheinyl (Ppant) transferase (PPTase) from *Bacillus subtilis*, leading to the *in vivo* conversion of their PCP and ACP domains from inactive apo-form into active holo-form by the thiolation of each active-site L-serine residue with Ppant. The CaeA1, CaeA2 and CaeA3 proteins produced here were soluble (**Supplementary Fig. 1**), and believed to be active and able to sequentially incorporate the starter and extender units via thiolated PCP/ACP S-(amino)acylation for chain elongation (**Supplementary Fig. 2a**). However, no off-line products were observed during the incubation with the substrates picolinic acid, malonyl-S-CoA, L-cysteine and L-leucine in the presence

of adenosine triphosphate (ATP, necessary for A domain activity) (**Fig. 2a**, i). The CAE-related assembly line likely functions with certain partner(s) that act(s) *in trans*.

Analyzing the *cae* cluster revealed two genes, *caeB1* and *caeA4* (**Fig. 1a**), which encode a flavin-dependent protein and a Type II thioesterase, respectively. Both genes are immediately downstream of *caeA2* and *caeA3* and are likely cotranscribed with these upstream assembly line-encoding genes. To determine whether CaeB1 and CaeA4 are functionally related, we expressed them in *E. coli* (**Supplementary Fig. 1**). Purified CaeB1, which appeared light yellow, was determined to bind oxidized flavin adenine dinucleotide (FAD) in a noncovalent manner based on absorbance spectrum analysis and protein heating followed by high-performance liquid chromatography with high-resolution mass spectrometry (HPLC-HR-MS) (**Supplementary Fig. 3**). The combination of CaeB1 with CaeA1, CaeA2 and CaeA3 resulted in the CAE intermediate 2,2'-bipyridinyl-L-leucine (**1**), whose production however did not require CaeA4-associated, putative thioesterase activity (**Fig. 2a**). Consequently, the CAE NRPS-PKS assembly line requires the *trans* activity of flavoprotein CaeB1 in the assembly of picolinic acid, malonyl-S-CoA, L-cysteine and L-leucine. In the reaction mixture containing tris(2-carboxyethyl)phosphine (TCEP) as a reducing agent, we observed tris(2-carboxyethyl)phosphine-sulfide (**2**, **Fig. 2b**), which was indicative of H₂S production. H₂S was then trapped using compound **19** as a probe in the TCEP-free reaction mixture (**Fig. 2c**)²⁸, confirming the involvement of dethiolation in the production of **1**.

The *trans* flavoprotein oxidatively processes L-cysteinyl on PCP

Phylogenetically, CaeB1 is close to flavin-dependent dehydrogenases, particularly those acting on acyl thioester substrates (**Supplementary Fig. 3d**)^{29,30}. This observation led to the hypothesis that CaeB1 oxidatively processes L-cysteinyI on PCP prior to its noncanonical incorporation into Ring A of CAEs. We thus focused on CaeA2 and analyzed its changes in PCP *S*-aminoacylation using a Bottom-Up protein MS strategy^{31,32}. For method establishment, thiolated CaeA2 was subjected to protease hydrolysis, and peptide mixtures were analyzed by nanoLC-MS/MS to map the sequence containing the Ppant-modified L-serine residue. Extensive attempts revealed **SLGGDSIMGIQF₂₀₄₂VSR** (the relevant L-serine is underlined), a 15-aa sequence generated from complete trypsin treatment and partial chymotrypsin digestion (**Supplementary Fig. 4a**). This sequence can be detected by MS, largely due to its C-terminal alkaline residue, L-arginine, which facilitates peptide ionization. To eliminate the remaining chymotrypsin cleavage site and retain L-arginine, we engineered CaeA2 by mutating F2042 to L-leucine, L-isoleucine and L-valine based on residue conservation analysis (**Supplementary Fig. 2b**). The activity of each variant, i.e., CaeA2^{F2042L}, CaeA2^{F2042I} or CaeA2^{F2042V}, was then assayed by substituting for wild-type CaeA2 in the **1**-producing reaction mixture. In contrast to the mutations F2042I and F2042V, both of which caused a significant decrease in **1** production, F2042L had little effect on CaeA2 activity (**Supplementary Fig. 2b**). The CaeA2^{F2042L} variant can be completely digested using trypsin and chymotrypsin to efficiently produce **SLGGDSIMGIQL₂₀₄₂VSR**, facilitating analysis by HR-MS/MS to confirm the PCP sequence identity and covalently tethered substrate/intermediate (**Supplementary Fig. 4b**).

PCP *S*-aminoacylation-caused mass changes in **SLGGDSIMGIQL₂₀₄₂VSR** were then monitored (**Fig. 3**). The incubation of thiolated CaeA2^{F2042L} with L-cysteine yielded L-cysteinyI-*S*-CaeA2^{F2042L} (**3**)

(**Supplementary Fig. 5a**), consistent with the activities of the A and PCP domains of the CaeA2 NRPS module. The presence of L-cysteinyl in CaeA2^{F2042L} was confirmed by treatment with iodoacetamide (IAM, a thiol derivatization reagent), which yielded derivative **4** (with an MS signal ~30-fold stronger than that of **3**, **Fig. 3** (ii, left) and **Supplementary Fig. 5b**). Incubation of the above mixture with CaeB1 significantly lowered the production of **3** by ~70% (**Fig. 3**, iii, left). Accordingly, treatment with IAM generated **5**, a derivative of (3-sulfhydryl)-pyruvoyl-S-CaeA2^{F2042L} (**6**) (**Fig. 3** (iii, middle) and **Supplementary Fig. 6a**), indicating that CaeB1 acts on PCP for L-cysteinyl transformation. This conclusion was supported by the observation of a + 3Da derivative of **5** when L-[1,2,3-¹³C₃,¹⁵N]cysteine was used (**Supplementary Fig. 6**). Using L-[2,3,3-D₃]cysteine as the substrate resulted in deuterium-unlabeled **5** (**Supplementary Fig. 6**), consistent with the notion that CaeB1 catalyzes the α,β -dehydrogenation of L-cysteinyl and converts **3** to dehydrocysteinyl-S-CaeA2^{F2042L} (**7**), which appears to be unstable and readily undergoes epimerization followed by hydrolysis to produce **6**. Proton exchange can occur at the C β position of **6**, leading to a loss of deuterium.

The sequence alignment of CaeB1 with various flavin-dependent dehydrogenases revealed potential key residues including E372 at the active site and S168 related to FAD binding (**Supplementary Fig. 3d**). These residues were mutated, generating CaeB1^{E372L} and CaeB1^{S168A}. The former variant shares with the wild-type protein a light yellow color and an absorbance spectrum characteristic of FAD in oxidized form; in contrast, the latter variant was nearly colorless, indicating the loss of FAD-binding ability (**Supplementary Fig. 3**). CaeB1^{E372L} and CaeB1^{S168A} were individually used to replace wild-type CaeB1 in the **1**-producing reaction mixture. Both mutations abolished **1** production, and saturation with an excess of FAD compensated the mutation S168A only (**Supplementary Fig. 3**). These results

ascertained that CaeB1 activity is FAD-dependent. Mechanistically, CaeB1 catalyzes α,β -dehydrogenation likely utilizing E372 as a base for L-cysteinyl α -deprotonation, followed by the elimination of a β -hydride equivalent to FAD to yield dehydrocysteinyl and FADH₂ (**Supplementary Fig. 7**).

The atypical NRPS Ct domain recruits *trans* flavoprotein activity

Notably, CaeB1 activity strictly depends on the Ct domain of CaeA2, a C-like domain that is shortened by $\sim 1/3$ and lacks the catalytic L-histidine residue (**Supplementary Fig. 8**). Deleting this domain completely abolished the production of **1** (**Fig. 2a, v**). In the incubation of truncated CaeA2^{F2042L} Δ Ct with L-cysteine, L-cysteinyl-S-CaeA2^{F2042L} Δ Ct (**8**) was observed; however, the α,β -dehydrogenation of L-cysteinyl failed to occur when CaeB1 was introduced (**Fig. 3** (iv, left and middle) and **Supplementary Fig. 9**). We thus proposed that resembling the X domain of the NRPSs involved in glycopeptide biosynthesis³³, Ct is necessary for recruiting the activity of the flavoprotein CaeB1 that functions *in trans*. To validate this hypothesis, we expressed the N-terminally thioredoxin (Trx)-tagged Ct domain in *E. coli* and measured its interaction with CaeB1 by isothermal titration calorimetry (ITC). Because CaeB1 is relatively unstable and tends to be precipitated during the measurement, we prepared the variant of this flavoprotein by fusion with maltose binding protein (MBP) at the N-terminus. ITC analysis revealed a K_d value of 0.53 ± 0.12 μ M between MBP-fused CaeB1 variant and Trx-tagged Ct domain, indicating the existence of a moderate interaction (**Fig. 4**). In contrast, this interaction was not observed by titrating MBP-fused CaeB1 variant to control proteins, e.g., CaeA2 Δ Ct (**Supplementary Fig. 10a**). Consequently, the flavoprotein CaeB1 likely functions *in trans* through the interaction with

the Ct domain of CaeA2, leaving the Cy domain in the atypical NRPS module as the catalyst for dehydrocysteinylation and subsequent heterocyclization.

2,2'-Bipyridine formation proceeds in the atypical NRPS module

We then examined the elongating intermediate covalently tethered to the PCP domain of CaeA2 during **1** formation. The incubation of CaeA1, CaeA2^{F2042L} and CaeB1 with picolinic acid, malonyl-*S*-CoA and L-cysteine led to the accumulation of 2,2'-bipyridinyl-*S*-CaeA2^{F2042L} (**9**). In contrast, **9** was not observed in the presence of CaeA3 for **1** production (**Fig. 3** (iv and v, right) and **Supplementary Fig. 11a**), supporting that 2,2'-bipyridinyl is an intermediate formed in the atypical NRPS module of CaeA2 before L-leucine extension. The atypical NRPS Cy domain likely catalyzes the formation of 2,2'-bipyridinyl, because mutating the catalytic residue D1165 in its conserved motif DxxxxD(1165)xxS to L-alanine completely abolished the production of **1** (**Fig. 2a**, vi). H₂S was observed during **9** formation (**Fig. 2b**), indicating that dethiolation occurred. The production of 2,2'-bipyridinyl in the atypical NRPS module of CaeA2 was further confirmed by HR-MS/MS analysis of the positive control of 2,2'-bipyridinyl-*S*-CaeA2^{F2042L}, which was generated by incubating unthiolated CaeA2^{F2042L} with synthetic 2,2'-bipyridinyl-*S*-CoA in the presence of Sfp (**Fig. 3** (i, right) and **Supplementary Fig. 11b**). In addition, the combination of CaeA3 with truncated 2,2'-bipyridinyl-*S*-PCP_{CaeA2} (**10**, produced by thiolating the apo-form PCP_{CaeA2} domain with Sfp in the presence of 2,2'-bipyridinyl-*S*-CoA) and L-leucine produced **1** (**Supplementary Fig. 12**), validating that this single module NRPS accepts dethiolated 2,2'-bipyridinyl for L-leucine extension. Notably, 2,2'-bipyridinyl loses its chirality at C α . This feature is consistent with the fact that CaeA3 possesses a ^DC_L domain (**Supplementary Fig. 13**)³⁴.

A similar catalysis was observed in the biosynthesis of nocardicins³⁵, in which an NRPS ^DC_L domain mediates the assembly of a C α -achiral intermediate with an L-amino acid for β -lactam formation.

Timing of the dethiolation step in the formation of the CAE 2,2'-bipyridine core

Focusing on the timing of the dethiolation step, we attempted to dissect the specific 2,2'-bipyridine-forming process in the CAE assembly line (**Fig. 5**). Omitting CaeA1 for picolinic acid incorporation from the **1** or **9**-forming reaction mixture prevented H₂S production (**Fig. 2b**), inconsistent with *route a* in which condensation follows the dethiolation of dehydrocysteinyI to 2-amino-allyl-*S*-CaeA2 (**11**). In addition, derivatives of the possible intermediate **11** (e.g., pyruvoyl-*S*-CaeA2, which might arise from epimerization and hydrolysis in a manner similar to the conversion of **7** to **6**) in this route were not observed based on careful HR-MS/MS analysis. Removing the substrate picolinic acid from the **1** or **9**-forming reaction mixtures led to similar results, supporting the hypothesis that the condensation with dehydrocysteinyI precedes its dethiolation. The Cy domain of CaeA2 shares high sequence homology to the counterparts known for thiazoline formation (**Supplementary Figs. 8 and 14**), in which related Cy domains display dual activity for L-cysteine or L-serine extension by forming an amide bond first and then cyclization via the nucleophilic addition of the –SH or –OH side chain of the newly incorporated residue onto the preceding carbonyl group to yield a five-membered thiazoline or azoline ring after H₂O elimination^{3,5}. However, the Cy domain of CaeA2 appears to be functionally distinct because it does not utilize unprocessed L-cysteinyI for condensation (**Figs. 2 and 3**). This domain likely follows CaeB1 activity and condenses the resulting reactive dehydrocysteinyI unit with the upstream intermediate for C-C bond formation through unprecedented C β -nucleophilic

substitution, yielding picolinyl-acetyl-dehydrocysteinyl-*S*-CaeA2 (**12**, **Fig. 5**) before heterocyclization to form the six-membered pyridine ring. We then utilized L-[2,3,3-D₃]cysteine in the **1**-forming reaction mixture. This assay led to the production of the + 1Da derivative of **1** (**Supplementary Fig. 15**), which excludes *route b* in which cyclization precedes dethiolation. As shown in this route, cyclization requires the epimerization of **12** to the enamine, which would eliminate the C β deuterium of the dehydrocysteinyl residue and result in unlabeled **1**. Instead, *route c* is favored based on this observation because the dethiolation of **12** to picolinyl-acetyl-dehydroalanyl-*S*-CaeA2 (**13**) facilitates cyclization to afford 2,2'-bipyridinyl and retain the C β deuterium (**Fig. 5**). For dethiolation, reduced FADH₂ might serve as a hydride supplier and undergo oxidation to produce FAD for recycling, consistent with a net non-redox result from this process (**Supplementary Fig. 7a**).

2,2'-Bipyridine assembly line provides both dethiolated and thiolated intermediates

Intriguingly, although *route b* is not favored in CAE biosynthesis, skipping the dethiolation step allows the condensation of 5-sulfhydryl-2,2'-bipyridinyl-*S*-CaeA2 (**14**) with L-leucine to produce 5-sulfhydryl-2,2'-bipyridinyl-L-leucine (**15**), the proposed off-loading intermediate in the biosynthesis of COLs (**Fig. 5**). Distinct from CAEs, COLs are 2,2'-bipyridine antibiotics possessing a sulfur-containing group (**Fig. 1**). In addition to ColA1a and ColA1b, ColA2 and ColA3, which share high sequence homology with the CAE-related modular synth(et)ases CaeA1, CaeA2 and CaeA3, respectively, the biosynthetic pathway of COLs involves ColB1, the homolog of the flavoprotein CaeB1 that also noncovalently binds a FAD cofactor (**Supplementary Fig. 3**), and thus are assumed to share a similar *trans* flavin-dependent NRPS-PKS assembly line for the formation of the different

thiolated 2,2'-bipyridine core^{14,25}. To determine the difference in 2,2'-bipyridine formation, we attempted to replace the CAE enzymes with their counterparts from the COL pathway in the **1**-producing reaction mixture. These counterparts did not include those for picolinyl priming (i.e., CaeA1 or ColA1a/ColA1b) and subsequent two-carbon elongation (i.e., the PKS module of CaeA2 or ColA2), which are believed to be identical in the biosynthesis of CAEs or COLs. The *trans* partners ColB1 and CaeB1 were confirmed to be interchangeable in the production of **1** (**Fig. 2a**, vii), supporting that they both share the *trans* flavin-dependent activity for oxidatively processing L-cysteinyl to dehydrocysteinyl on PCP. Consistent with this finding, ITC analysis showed that ColB1 can interact with the Ct domain of the CaeA2, with a moderate K_d value of $1.18 \pm 0.15 \mu\text{M}$ (**Fig. 4**). Interestingly, changing the NRPS CaeA3 to ColA3 completely abolished **1** production (**Fig. 2a**, viii), indicating that ColA3 differs from CaeA3, even though they both mediate L-leucine extension.

In vitro testing the exchangeability between the PKS-NRPS hybrid proteins CaeA2 and ColA2 failed because ColA2 was highly resistant to various methods for soluble protein preparation. We thus turned to *in vivo* biochemical assays and examined this exchangeability in the COL-producing strain *Streptomyces roseosporus* (**Fig. 6**). A chimeric gene that encodes the variant harboring the PKS module of ColA2 and the NRPS module of CaeA2, *colA2_{PKS}-caeA2_{NRPS}*, was constructed and introduced into the previously developed, $\Delta colA2$ mutant *S. roseosporus* strain that is incapable of producing COLs¹⁴. Remarkably, this complementation restored the production of COLs, supporting that the atypical NRPS modules of CaeA2 and ColA2 are functionally identical in terms of L-cysteinyl incorporation, *trans* CaeB1/ColB1 activity recruiting, condensation and heterocyclization during COL biosynthesis. These findings demonstrate the generality of both COLs and CAEs in 2,2'-bipyridine formation and

particularly suggest that dethiolation can be skipped in the atypical NRPS module of CaeA2 to produce the thiolated 2,2'-bipyridine intermediate 5-sulfhydryl-2,2'-bipyridinyl. Notably, no dethiolated CAE-like products were observed in the *S. roseosporus* strain when *colA2* was replaced with *colA2_{PKS}*-*caeA2_{NRPS}* (**Fig. 6**). Most likely, the specificity of the L-leucine extension determines the fate of the sulfur as the substitution with the CaeA2 NRPS module does not change the product profile.

L-Leucine extension specializes 2,2'-bipyridine differentiation

Unlike 2,2'-bipyridinyl, sulfhydryl-2,2'-bipyridinyl formed in the assembly line is highly reactive and proved to be difficult for analysis by aforementioned HR-MS/MS approaches. To *in vitro* evidence this covalently tethered thiol intermediate, we attempted to examine the downstream off-loading intermediates by incubating CaeA1, CaeA2 and CaeB1 with ColA3 in the presence of the necessary substrates and ATP. Treating the reaction mixture with IAM revealed **17**, the derivative of the expected product 5-sulfhydryl-2,2'-bipyridinyl-L-leucine (**15**) (**Supplementary Fig. 16**). To further validate the production of **15**, we added ColG2, a COL pathway-specific methyltransferase^{14,25}, into this reaction mixture. After L-leucine extension, ColG2 is believed to catalyze the *S*-methylation of the highly reactive thiol of intermediate **15** to produce the less reactive, previously characterized intermediate 5-methylmercapto-2,2'-bipyridinyl-L-leucine (**16**). Indeed, **16** was observed as the sole product (**Fig. 2a**, ix). During this *in vitro* conversion, the cycling of FADH₂ to FAD can go through the formation of the oxidizing adduct FAD-4a-OOH, in agreement with a net oxidation occurring in sulfhydryl-2,2'-bipyridine formation (**Supplementary Fig. 7b**). In contrast, the combination of ColG2 with the complete CAE assembly line (i.e., CaeA1, CaeA2 and CaeA3 as well as CaeB1) failed to produce **16**

and had little effect on the production of **1** (**Fig. 2a, x**), indicating that **15** cannot be generated in the presence of CaeA3. Furthermore, we narrowed the specificity of the upstream intermediate selection by CaeA3 and ColA3 to their C domains by domain swapping. Extensive attempts pointed to active CaeA3_CColA3_{A-PCP}, the chimeric protein composed of the C domain of CaeA3 and the A-PCP didomain from ColA3. This protein was confirmed to be functionally identical to CaeA3 in the production of **1**, albeit with a lower yield (~70% decrease) (**Fig. 2a, xi**). It cannot be used to replace ColA3 because **16** was not observed in the presence of ColG2. Clearly, the CAE and COL assembly lines are identical for 2,2'-bipyridine formation and can simultaneously provide the dethiolated and thiolated intermediates (**9** and **14**). The only exception is the C domain-conferred specificity in L-leucine extension, which specializes the pathway to the production of either CAEs or COLs.

CONCLUSION

We demonstrate a common paradigm for 2,2'-bipyridine formation, which features a NRPS/PKS hybrid assembly line for sequential incorporation of the substrates picolinic acid, malonyl-S-CoA, L-cysteine and L-leucine with the association of a *trans* FAD-dependent partner (**Fig. 5**). This assembly line was first reconstituted *in vitro* for the biosynthesis of the dethiolated 2,2'-bipyridine antibiotics CAEs; then, the generality of 2,2'-bipyridine formation was validated in the biosynthesis of the distinct thiolated 2,2'-bipyridine antibiotics COLs, where the dethiolation step is skipped. In this assembly line, the atypical NRPS module Cy-A-PCP-Ct plays a central role in 2,2'-bipyridine formation. Following the A domain activity of this module for L-cysteine incorporation, the Ct domain recruits the activity of a flavoprotein that functions *in trans* for oxidatively processing L-cysteinyI on PCP. The Cy domain,

which differs from the counterparts well-known for thiazoline formation, then condenses the resulting reactive dehydrocysteinyl extender unit with the upstream picolinyl-acetyl intermediate to afford a C-C bond via unusual C β -nucleophilic substitution, followed by intramolecular heterocyclization of the linear picolinyl-acetyl-dehydrocysteinyl intermediate to form the 2,2'-bipyridine core. Surprisingly, dethiolation can proceed or not between the condensation and heterocyclization steps, allowing for the production of both dethiolated and thiolated 2,2'-bipyridine intermediates in the atypical NRPS module. Subsequent L-leucine extension, which does not contribute any atoms in the final molecules, determines the sulfur fate and thus differentiates the biosynthesis of CAEs and COLs. Specifically, the C domain of the NRPS module C-A-PCP confers this selectivity for L-leucine extension and advances one of the two intermediates down a path to the 2,2'-bipyridine antibiotics with or without sulfur decoration. Therefore, whether dethiolation proceeds or is skipped appears to be not specifically controlled (i.e., not the key step) in the 2,2'-bipyridine assembly line because advancing the either intermediates (dethiolated or thiolated) strictly depends on the subsequent C domain selectivity for L-leucine extension. Based on this study, we provide a unique flavin-dependent paradigm in the assembly line, where cyclization reactions are usually difficult to be approached and thus remains poorly understood compared to those occurring at the off-line stage^{3,4}. In addition, this study advances the understanding of the versatile functions of C domains in NRPS catalysis³⁶. These domains include the gatekeeping C domain for sulfur fate determination, the Ct domain for *trans* flavin activity recruiting and, particularly, the unprecedented Cy domain for C-C bond formation and heterocyclization. The findings reported here further our appreciation of assembly line enzymology and will facilitate the design, development and utilization of enzymatic molecular machinery to address synthetic challenges arising from structurally-related complex polypeptides, polyketides and their hybrids.

REFERENCES AND NOTES:

1. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805-1810 (2004).
2. Walsh, C. T. & Fischbach, M. A. Natural products version 2.0: connecting genes to molecules. *J. Am. Chem. Soc.* **132**, 2469-2493 (2010).
3. Sundaram, S. & Hertweck, C. On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems. *Curr. Opin. Chem. Biol.* **31**, 82-94 (2016).
4. Brown, A. S., Calcott, M. J., Owen, J. G. & Ackerley, D. F. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210-1228 (2018).
5. Pang, B., Wang, M. & Liu, W. Cyclization of polyketides and non-ribosomal peptides on and off their assembly lines. *Nat. Prod. Rep.* **33**, 162-173 (2016).
6. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524-12547 (2019).
7. Kaes, C., Katz, A. & Hosseini, M. W. Bipyridine: The Most Widely Used Ligand. A Review of Molecules Comprising at Least Two 2,2'-Bipyridine Units. *Chem. Rev.* **100**, 3553-3590 (2000).
8. Hapke, M., Brandt, L. & Lutzen, A. Versatile tools in the construction of substituted 2,2'-bipyridines-cross-coupling reactions with tin, zinc and boron compounds. *Chem. Soc. Rev.* **37**, 2782-2797 (2008).
9. Funk, A. & Divekar, P. V. Caerulomycin, a New Antibiotic from *Streptomyces Caeruleus* Baldacci: I. Production, Isolation, Assay, and Biological Properties. *Can. J. Microbiol.* **5**, 317-321 (1959).
10. Gurram, R. K., Kujur, W., Maurya, S. K. & Agrewala, J. N. Caerulomycin A enhances transforming growth factor-beta (TGF-beta)-Smad3 protein signaling by suppressing interferon-gamma (IFN-gamma)-signal transducer and activator of transcription 1 (STAT1) protein signaling to expand regulatory T cells (Tregs). *J. Biol. Chem.* **289**, 17515-17528 (2014).
11. Singla, A. K. *et al.* Cerulomycin A: A Potent Novel Immunosuppressive Agent. *Transplantation* **97**, E57-E59 (2014).

12. Kaur, S., Srivastava, G., Sharma, A. N. & Jolly, R. S. Novel immunosuppressive agent caerulomycin A exerts its effect by depleting cellular iron content. *Br. J. Pharmacol.* **172**, 2286-2299 (2015).
13. Kujur, W., Gurram, R. K., Haleem, N., Maurya, S. K. & Agrewala, J. N. Caerulomycin A inhibits Th2 cell activity: a possible role in the management of asthma. *Sci. Rep.* **5**, 15396-15405 (2015).
14. Qu, X. *et al.* Caerulomycins and collismycins share a common paradigm for 2,2'-bipyridine biosynthesis via an unusual hybrid polyketide-peptide assembly Logic. *J. Am. Chem. Soc.* **134**, 9038-9041 (2012).
15. Zhu, Y. *et al.* Identification of caerulomycin A gene cluster implicates a tailoring amidohydrolase. *Org. Lett.* **14**, 2666-2669 (2012).
16. Chen, D., Zhao, Q. & Liu, W. Discovery of caerulomycin/collismycin-type 2,2'-bipyridine natural products in the genomic era. *J. Ind. Microbiol. Biotechnol.* **46**, 459-468 (2019).
17. McInnes, A. G., Smith, D. G., Walter, J. A., Vining, L. C. & Wright, J. L. C. The biosynthesis of caerulomycin A in *Streptomycescaeruleus*. Incorporation of ¹⁴C- and ¹³C-labeled precursors and analyses of labeling patterns by ¹³C nmr. *Can. J. Chem.* **57**, 3200-3204 (1979).
18. Vining, L. C., McInnes, A. G., McCulloch, A. W., Smith, D. G. & Walter, J. A. The biosynthesis of caerulomycins in *Streptomyces caeruleus*. Isolation of a new caerulomycin and incorporation of picolinic acid and glycerol into caerulomycin A. *Can. J. Chem.* **66**, 191-194 (1988).
19. Zhu, Y. *et al.* Biochemical and Structural Insights into the Aminotransferase CrmG in Caerulomycin Biosynthesis. *ACS Chem. Biol.* **11**, 943-952 (2016).
20. Zhu, Y. *et al.* Insights into caerulomycin A biosynthesis: a two-component monooxygenase CrmH-catalyzed oxime formation. *J. Am. Chem. Soc.* **135**, 18750-18753 (2013).
21. Fu, P. *et al.* Acyclic congeners from *Actinoalloteichus cyanogriseus* provide insights into cyclic bipyridine glycoside formation. *Org. Lett.* **16**, 4264-4267 (2014).
22. Chen, M. *et al.* Enzymatic Competition and Cooperation Branch the Caerulomycin Biosynthetic Pathway toward Different 2,2'-Bipyridine Members. *Org. Biomol. Chem.* **15**, 5472-5475 (2017).
23. Chen, M., Pang, B., Du, Y. N., Zhang, Y. P. & Liu, W. Characterization of the metallo-dependent amidohydrolases responsible for "auxiliary" leucinyI removal in the biosynthesis of 2,2'-bipyridine antibiotics. *Synth Syst Biotechnol* **2**, 137-146 (2017).
24. Shindo, K., Yamagishi, Y., Okada, Y. & Kawai, H. Collismycins A and B, novel non-steroidal

- inhibitors of denamethasone-glucocorticoid receptor binding. *J. Antibiot.* **47**, 1072-1074 (1994).
25. Garcia, I. *et al.* Elucidating the biosynthetic pathway for the polyketide-nonribosomal peptide collismycin A: mechanism for formation of the 2,2'-bipyridyl ring. *Chem. Biol.* **19**, 399-413 (2012).
 26. Quadri, L. E. *et al.* Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595 (1998).
 27. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790-1792 (2001).
 28. Wei, C. *et al.* Dual-Reactable Fluorescent Probes for Highly Selective and Sensitive Detection of Biological H₂S. *Chem. Asian J.* **11**, 1376-1381 (2016).
 29. Ghisla, S. & Thorpe, C. Acyl-CoA dehydrogenases - A mechanistic overview. *Eur. J. Biochem.* **271**, 494-508 (2004).
 30. Watanabe, K., Khosla, C., Stroud, R. M. & Tsai, S.-C. Crystal Structure of an Acyl-ACP Dehydrogenase from the FK520 Polyketide Biosynthetic Pathway: Insights into Extender Unit Biosynthesis. *J. Mol. Biol.* **334**, 435-444 (2003).
 31. Bumpus, S. B. & Kelleher, N. L. Accessing natural product biosynthetic processes by mass spectrometry. *Curr. Opin. Chem. Biol.* **12**, 475-482 (2008).
 32. Gu, L. *et al.* GNAT-like strategy for polyketide chain initiation. *Science* **318**, 970-974 (2007).
 33. Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E. & Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **521**, 105-109 (2015).
 34. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78-92 (2007).
 35. Gaudelli, N. M., Long, D. H. & Townsend, C. A. beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **520**, 383-387 (2015).
 36. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *BBA-Proteins and Proteomics* **1865**, 1587-1604 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Dr. L. Yi (Beijing University of Chemical Technology) and Prof. Z. Xi (Nankai University) for providing the fluorescent probe in H₂S examination. This work was supported in part by grants from NSFC (21750004, 21520102004, and 21621002), CAS (QYZDJ-SSW-SLH037 and XDB20020200), STCSM (17JC1405100), MST (2018ZX091711001-006-010) and K. C. Wang Education Foundation.

FIGURE LEGENDS

Figure 1. Biogenesis of the polyketide-peptide hybrid, 2,2'-bipyridine antibiotics CAE-A and COL-A. **(a)** Organization of the genes conserved in the CAE and COL biosynthetic gene clusters that codes for the hybrid NRPS/PKS assembly line and its associated *trans* FAD-dependent partner (gray) and other proteins (white). The didomain encoding gene *caeA1* in the *cae* cluster corresponds to the two discrete genes, *colA1a* and *colA1b*, in the *col* cluster. **(b)** Hybrid NRPS/PKS assembly line for 2,2'-bipyridine formation. The functional domains (AT and A domains) for substrate specificity are indicated in different colors with their associated building blocks: blue for picoliny, yellow for malonyl, red for L-cysteinyl, and light gray for L-leuciny. The sulfur atoms in the molecules are highlighted in red.

Figure 2. *In vitro* reconstitution of the 2,2'-bipyridine assembly line. **(a)** Analyses of 2,2'-bipyridine production by HPLC (λ 315 nm). The control reaction was conducted with the combination of the proteins CaeA1, CaeA2 and CaeA3 with the substrates picolinic acid, malonyl-S-CoA, L-cysteine, L-leucine, and ATP under aerobic conditions (i). To determine the *trans* partner, the reactions were conducted with CaeB1 (ii), CaeA4 (iii), or CaeB1 and CaeA4 (iv). To validate the necessity of Ct, CaeA2 was replaced with CaeA2 Δ Ct (v). To determine the catalytic role of the Cy domain of the atypical NRPS module, CaeA2 was replaced with CaeA2_{D1165A}. (vi). To evaluate the component generality, the reactions were conducted with ColB1 (vii), CaeB1 and ColA3 (replacing CaeA3) (viii), CaeB1, ColA3 (replacing CaeA3) and ColG2 (ix), CaeB1 and ColG2 (x), or CaeB1 and CaeA3_CColA3_{A-PCP} (replacing CaeA3) (xi). Synthetic **1** (xii) and **16** (xiii) were used as standards. **(b)** Examination of H₂S production in TCEP-containing reaction mixtures by HPLC-MS. Electrospray

ionization (ESI) m/z $[M + H]^+$ at 283.0405 (calcd.) for **2**, which is generated by a reaction of H_2S with TCEP. The initial reaction was conducted by combining CaeA1, CaeA2, CaeA3 and CaeB1 with picolinic acid, malonyl- S -CoA, L-cysteine, L-leucine, and ATP (i), and reactions in the absence of CaeA3 (ii), picolinic acid and CaeA3 (iii) or CaeA1 and CaeA3 (iv) were also conducted. (C) Examination of H_2S production in TCEP-free reaction mixtures by HPLC with excitation at 370 nm and monitoring emission at 450 nm. The reaction of the fluorescent probe **19** with Na_2S to yield **20** was used as the positive control (i). H_2S production was examined in the reactions where CaeA1, CaeA2, CaeA3 and CaeB1 were combined with picolinic acid, malonyl- S -CoA, L-cysteine and L-leucine in the presence (ii) and absence (iii) of ATP.

Figure 3. Examination of PCP-tethered intermediates during 2,2'-bipyridine formation (top) by nanoLC-MS/MS (below). The sequence **SLGGDSIMGIQL₂₀₄₂VSR** (in the rectangle) arises from the complete digestion of CaeA2^{F2042L} using trypsin and chymotrypsin. For details of the HR-MS/MS analyses, see **Figs. S4~S6, S9 and S11**. The two IAM-treated sequences (ESI m/z $[M + 3H]^{3+}$ for the left, calcd. 678.3143; for the middle, calcd. 677.9704) come from L-cysteinyl- S -CaeA2^{F2042L} (**3**) and dehydrocysteinyl- S -CaeA2^{F2042L} (**7**), respectively, which were examined by incubating CaeA2^{F2042L} and L-cysteine in the absence (i, negative control) and presence (ii) of ATP, with CaeB1 (iii), CaeB1 and CaeA2^{F2042L} Δ Ct (replacing CaeA2^{F2042L}) (iv), or CaeB1, CaeA1, picolinic acid and malonyl- S -CoA (v). The right sequence (ESI m/z $[M + H]^+$, calcd. 690.9854) comes from 2,2'-bipyridinyl- S -CaeA2^{F2042L} (**9**). The positive control was prepared by incubating CaeA2^{F2042L} (with PCP in apo form) and 2,2'-bipyridinyl- S -CoA (i). **9** was examined by incubating CaeA2^{F2042L} and L-cysteine (ii), with CaeB1 (iii), CaeB1, CaeA1, picolinic acid and malonyl- S -CoA (iv), or CaeB1, CaeA1, CaeA3,

picolinic acid, malonyl-*S*-CoA and L-leucine (v).

Figure 4. Measurement of the interactions of the Ct domain of CaeA2 with related flavoproteins by ITC. Raw dates were shown on top, and the integrated curves containing experimental points and the best fitting line obtained from the single binding site model were shown on bottom. This measurement was conducted in triplicate. For negative controls, see **Supplementary Fig. 10**. **(a)** Titrating MBP-fused CaeB1 to Trx-tagged Ct. **(b)** Titrating ColB1 to Trx-tagged Ct.

Figure 5. Proposed mechanisms for 2,2'-bipyridine formation. The PCP domain of CaeA2 is labeled in red. The C domains of CaeA3 and ColA3 are highlighted in green and blue, respectively. The favored pathways for CAEs and COLs are highlighted with light yellow and light purple backgrounds, respectively.

Figure 6. Domain organization of the NRPSs CaeA2, ColA2 and ColA2_{PKS}-CaeA2_{NRPS} (Top) and HPLC analysis of COL-related products in *S. roseosporus* strains (below). i, the wild-type COL-producing strain (positive control); ii, the $\Delta colA2$ mutant in which a chimeric gene *colA2_{PKS}-caeA2_{NRPS}* is expressed *in trans*; and iii, the $\Delta colA2$ mutant (negative control).

Figure 1.

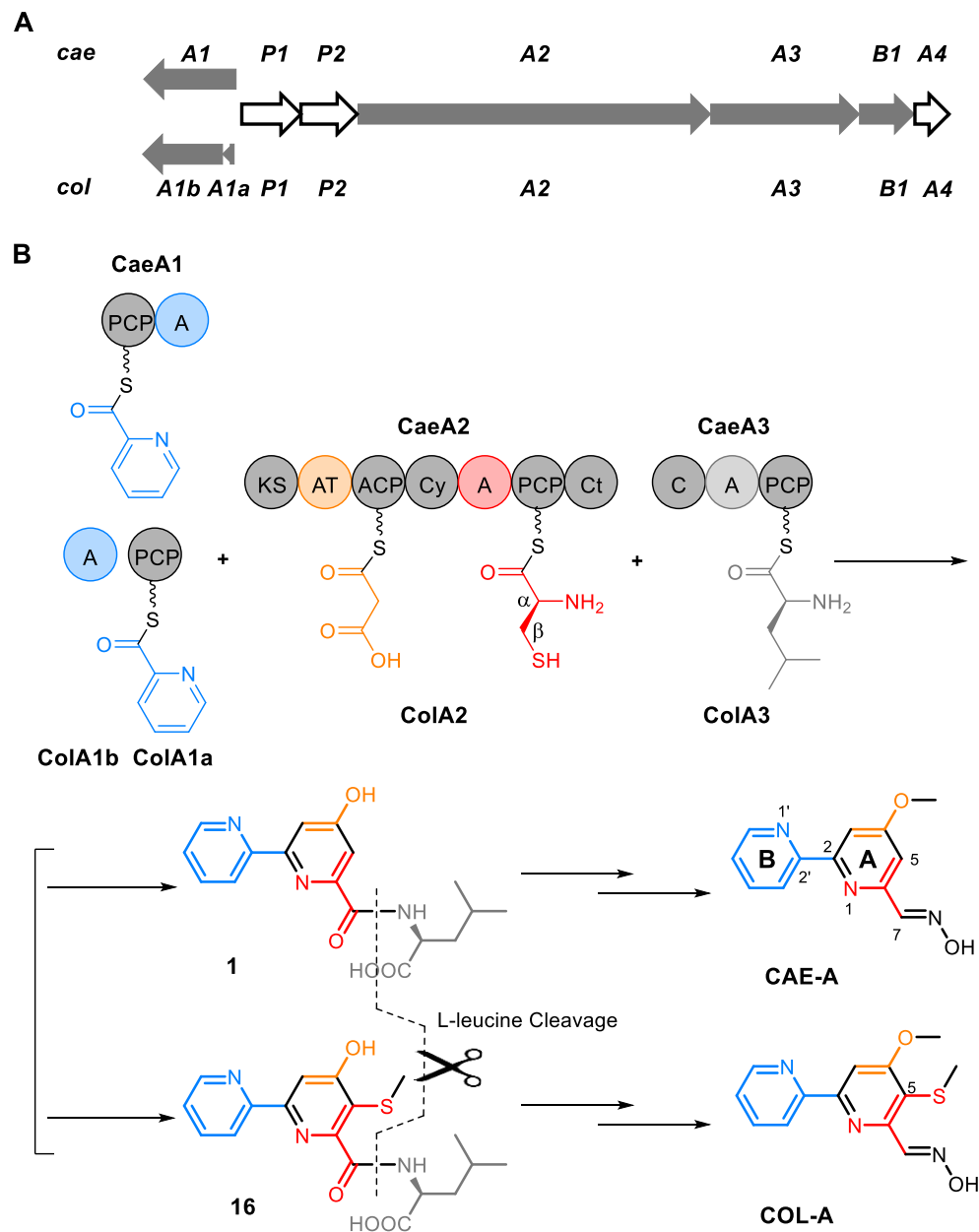


Figure 2.

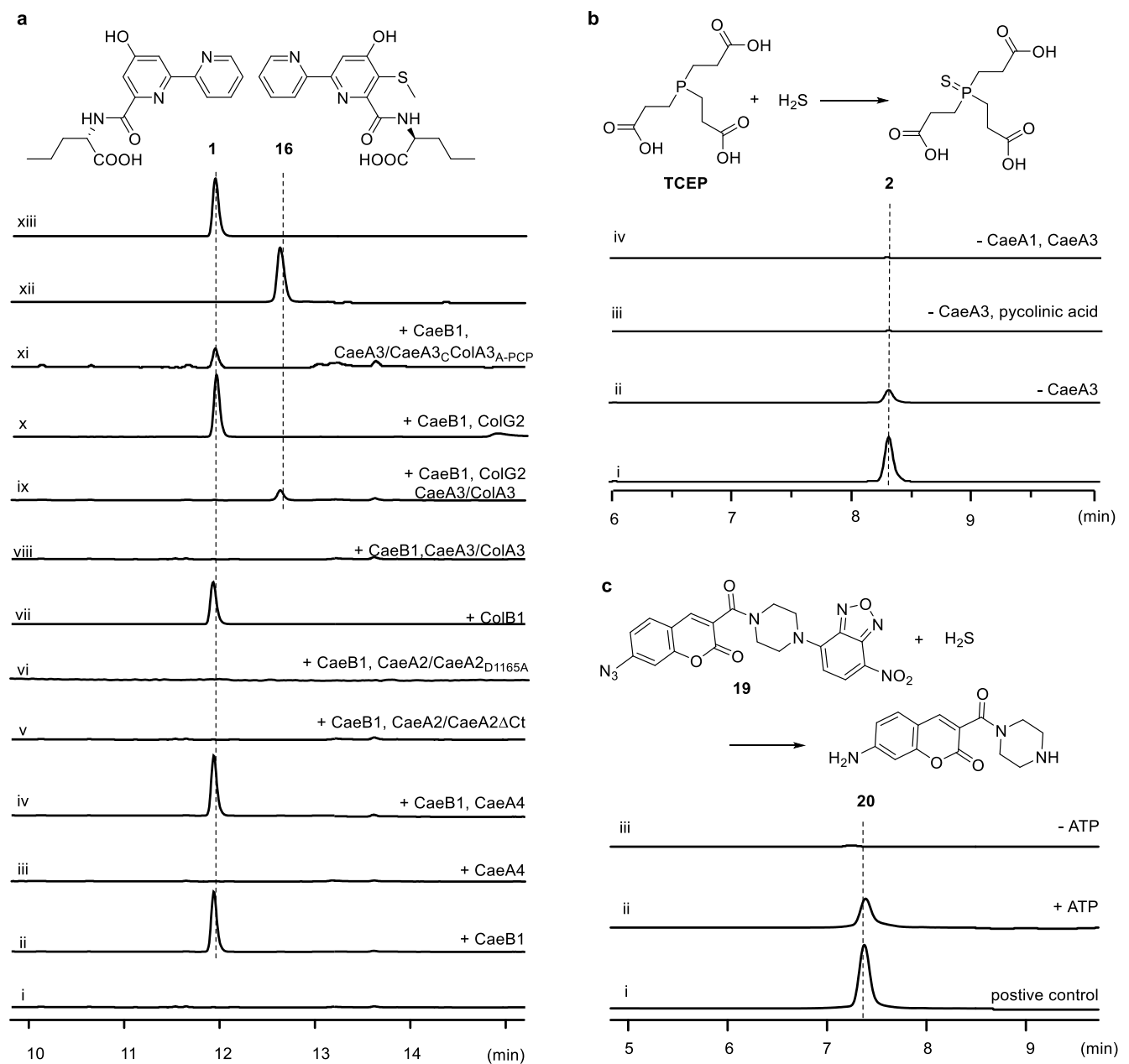


Figure 3.

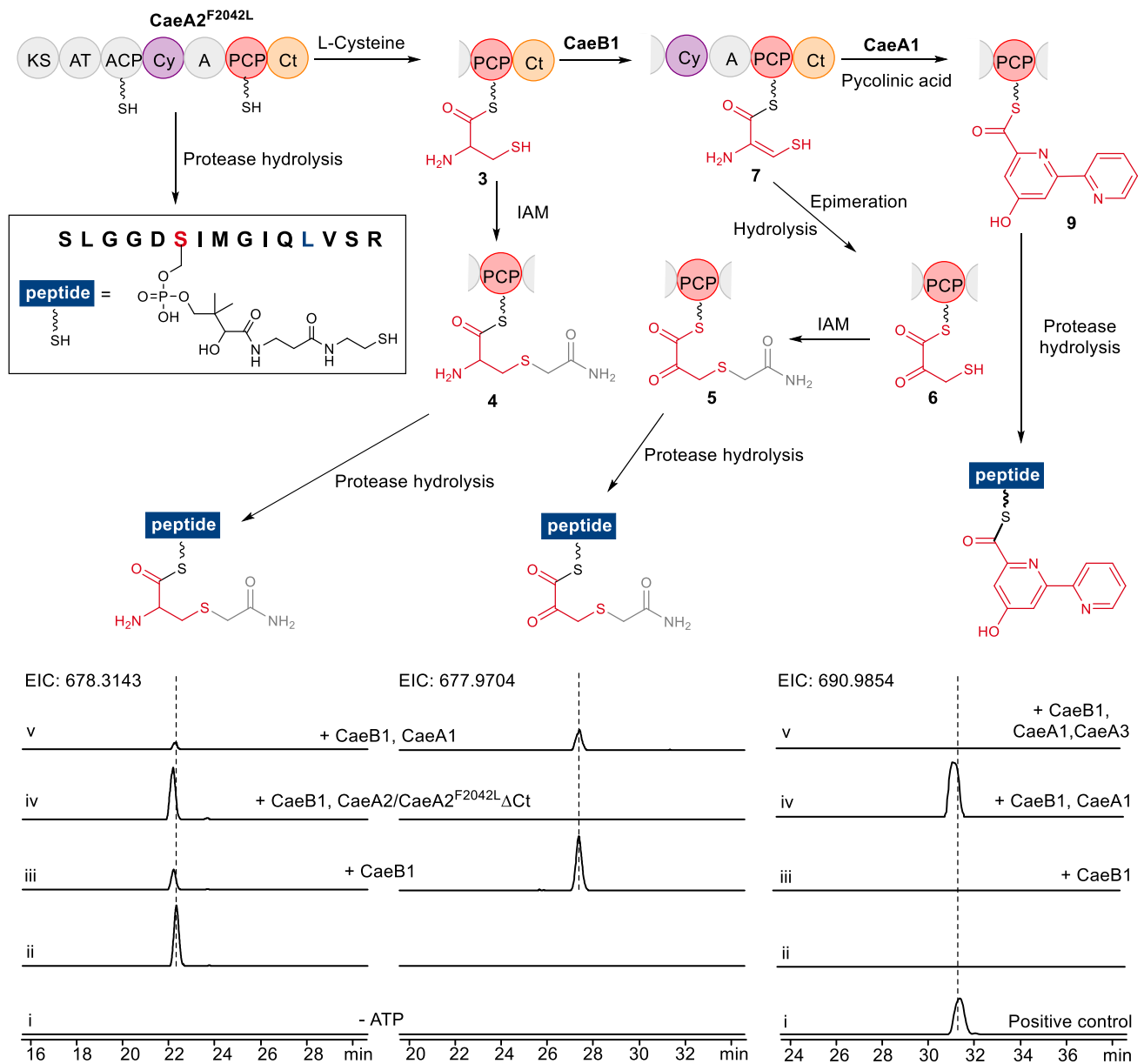


Figure 4.

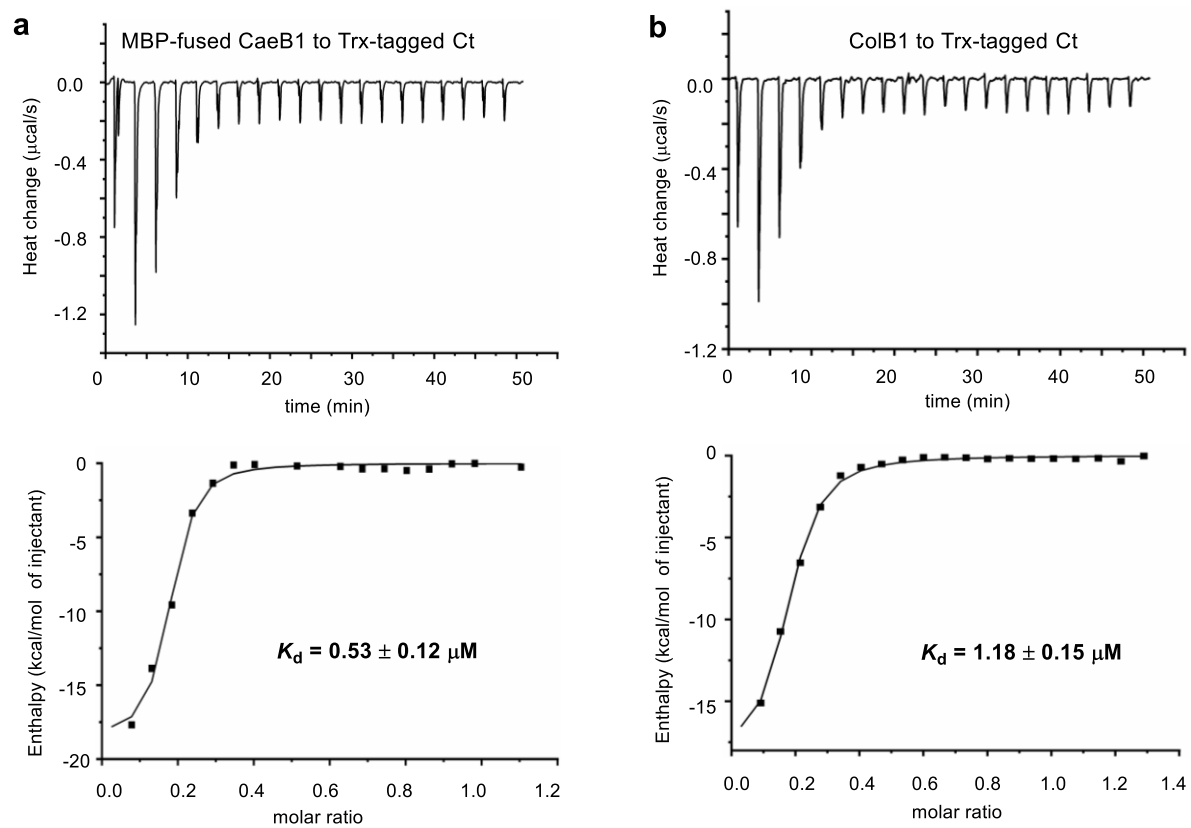


Figure 5.

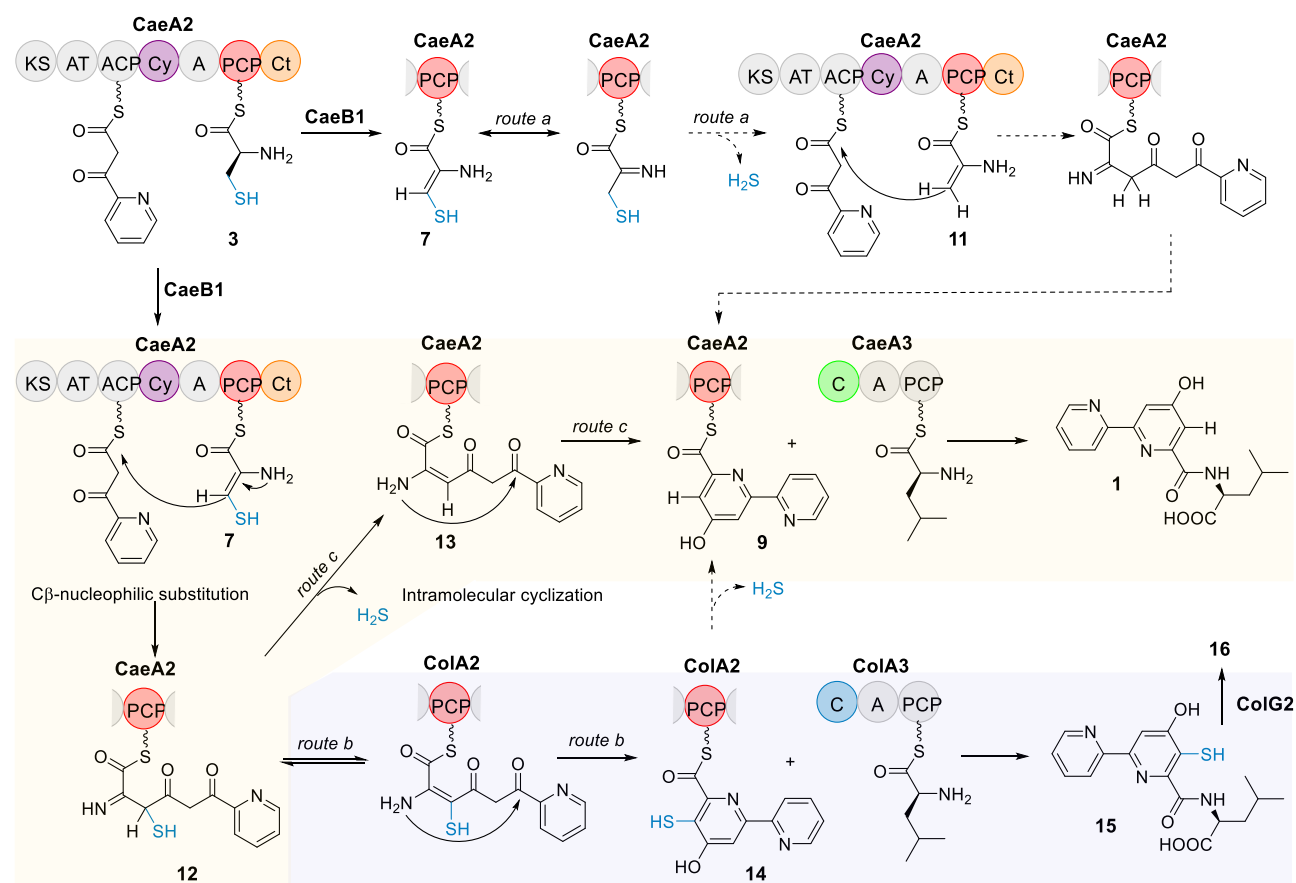
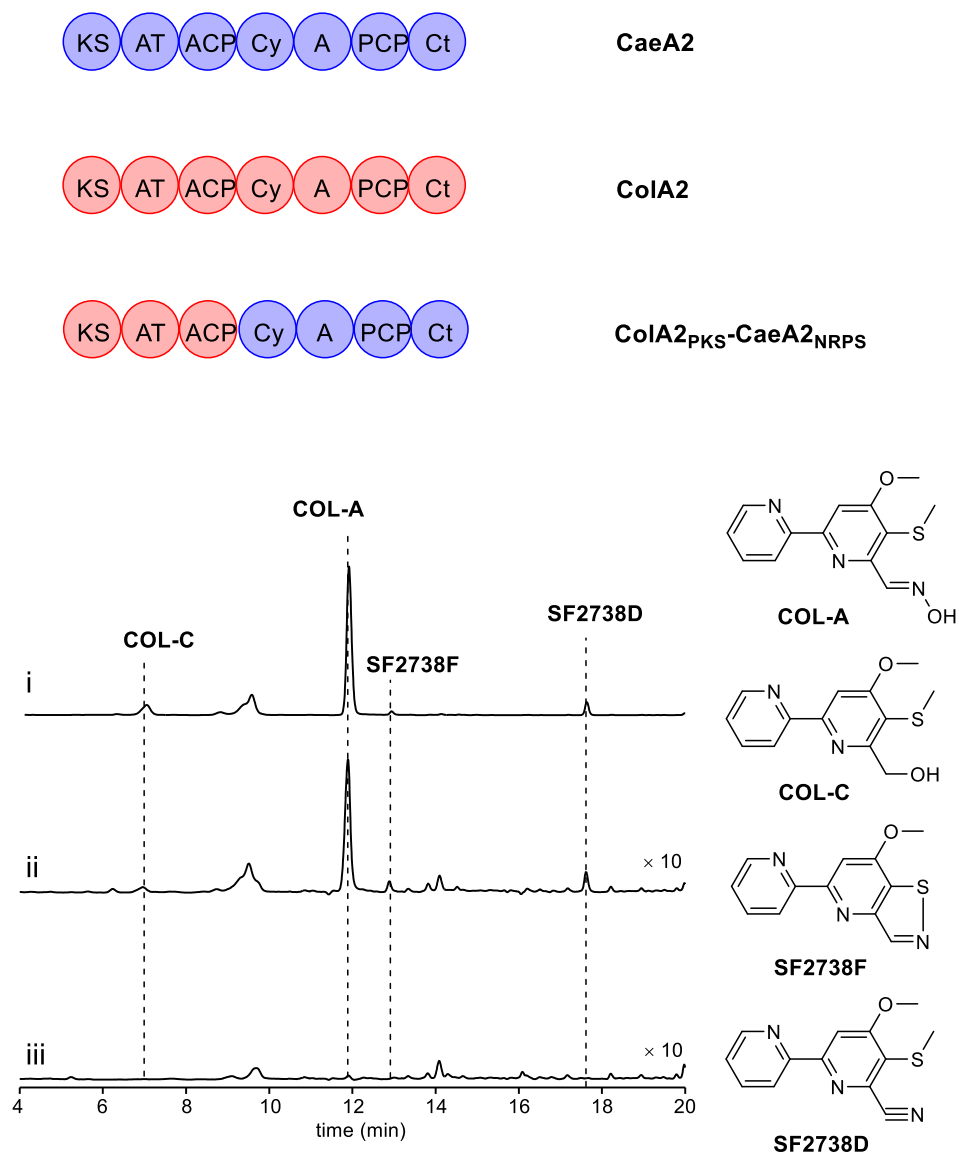


Figure 6.



Caerulomycin and Collismycin Antibiotics Share a *trans* Flavin-Dependent Assembly Line for 2,2'-Bipyridine Formation and Sulfur Fate Differentiation

Bo Pang,^{1,#} Rijin Liao,^{2,#} Zhijun Tang,^{1,#} Shengjie Guo,¹ Zhuhua Wu,¹ and Wen Liu^{1,3,*}

¹ State Key Laboratory of Bioorganic and Natural Products Chemistry, Center for Excellence in Molecular Synthesis, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China;

² Shanghai Institute of Precision Medicine, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200125, China;

³ Huzhou Center of Bio-Synthetic Innovation, 1366 Hongfeng Road, Huzhou 313000, China;

* To whom correspondence should be addressed: Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Rd., Shanghai 200032, China. Wen Liu, Email: wliu@sioc.ac.cn, Tel: 86-21-54925111, Fax: 86-21-64166128

These authors equally contribute to this work.

SUPPLEMENTARY METHODS

General materials and methods. Biochemicals and media were purchased from Sinopharm Chemical Reagent Co., Ltd. (China), Oxoid Ltd. (U.K.) or Sigma-Aldrich Co. LLC. (USA) unless otherwise stated. Restriction endonucleases were purchased from Thermo Fisher Scientific Co. Ltd. (USA). Chemical reagents were purchased from standard commercial sources. The bacterial strains, plasmids and primers used in this study are summarized in **Supplementary Tables 1, 2, and 3**, respectively.

DNA isolation and manipulation in *E. coli* or actinobacteria were carried out according to standard methods^{1,2}. PCR amplifications were carried out on an Applied Biosystems Veriti Thermal Cycler (Thermo Fisher Scientific Inc., USA) using either Taq DNA polymerase (Vazyme Biotech Co. Ltd, China) for routine verification or PrimeSTAR HS DNA polymerase (Takara Biotechnology Co., Ltd. Japan) for high fidelity amplification. Primer synthesis was performed at Shanghai Sangon Biotech Co. Ltd. (China). DNA sequencing was performed at Shanghai Majorbio Biotech Co. Ltd. (China).

High performance liquid chromatography (HPLC) analysis was carried out on an Agilent 1260 HPLC system (Agilent Technologies Inc., USA) equipped with a DAD detector. Semi-preparative HPLC was performed on an Agilent 1100 system equipped with a DAD detector (Agilent Technologies Inc., USA). HPLC Electrospray ionization MS (HPLC-ESI-MS) and tandem MS (MS/MS) for small molecules were performed on a Thermo Fisher LTQ Fleet ESI-MS spectrometer (Thermo Fisher Scientific Inc., USA), and the data were analyzed using Thermo Xcalibur software. NanoLC-MS/MS and MS/MS for peptides were performed on an EASY-nLC 1200 (Thermo Fisher Scientific Inc., USA) coupled with a

Q Exactive HF mass spectrometer (Thermo Fisher Scientific Inc., USA), and the data were analyzed using pymzML, pFind³ and Thermo Xcalibur. High resolution ESI-MS (HR-ESI-MS) analysis for small molecules was carried out on an Agilent 6230B Accurate Mass TOF LC/MS System (Agilent Technologies Inc., USA), and the data were analyzed using Agilent MassHunter Qualitative Analysis software. NMR data were recorded on an Agilent 500 MHz PremiumCompact+ NMR spectrometer (Agilent Technologies Inc., USA).

Protein expression and purification. The recombinant proteins CaeA1, PCP_{CaeA2}, CaeB1, ColB1, ColG2, CaeA3, ColA3, and the chimeric NRPS protein CaeA3_CColA3_{A-PCP}, were produced in N-terminal 6 × His-tagged forms, while CaeA2 and its variants (e.g., CaeA2^{F2042L}, CaeA2^{F2042I}, CaeA2^{F2042V}, Ct_{CaeA2}, CaeA2-ΔCt and CaeA2^{F2042L}ΔCt) were expressed in C-terminal 8 × His-tagged forms or N-terminal 6 × His and Trx-tagged forms.

In general, the genes coding for the above proteins were amplified by PCR individually from the CAE-producing strain *Actinoalloteichus cyanogriseus* or the COL-producing strain *Streptomyces roseosporus*, both of which are listed in **Supplementary Table 1**, using the corresponding primers listed in **Supplementary Table 3**. The PCR products were first cloned into the vector pMD19-T for sequencing to confirm the fidelity and then into the expression vector pET-28a(+) (for N-terminal 6 × His-tagged proteins) or pET-37b(+) (for C-terminal 8 × His-tagged proteins). For expression of the chimeric NRPS protein CaeA3_CColA3_{A-PCP} (containing the C domain of CaeA3 and the A and PCP domains of ColA3), a 1.4 kb DNA fragment amplified by PCR from *A. cyanogriseus* using the primers CaeA3-C-For and CaeA3-C-Rev was cloned into pMD19-T, yielding pQL1046. Similarly, a 1.8 kb

DNA fragment amplified by PCR from *S. roseosporus* using the primers ColA3-APCP-For and ColA3-APCP-Rev was cloned into pMD19-T, yielding pQL1047. The 1.4 kb NdeI-NheI fragment and the 1.8 kb NheI-HindIII fragment were recovered and co-ligated into the NdeI-HindIII site of pET-28a(+) to give pQL1048.

The resulting recombinant plasmids, which are listed in **Supplementary Table 2**, were transferred into *E. coli* BL21(DE3) for expression of CaeA2^{F2042L}, PCP_{CaeA2}, Trx-Ct_{CaeA2}, CaeB1, ColG2 and ColB1 or into *E. coli* BAP1, an engineered strain capable of co-expressing the PPTase Sfp⁴, for expression of CaeA1, CaeA3, ColA3, and CaeA2 and its variants CaeA2^{F2042L}, CaeA2^{F2042I}, CaeA2^{F2042V}, CaeA2-ΔCt, CaeA2^{F2042L}ΔCt and CaeA3_CColA3_A-PCP.

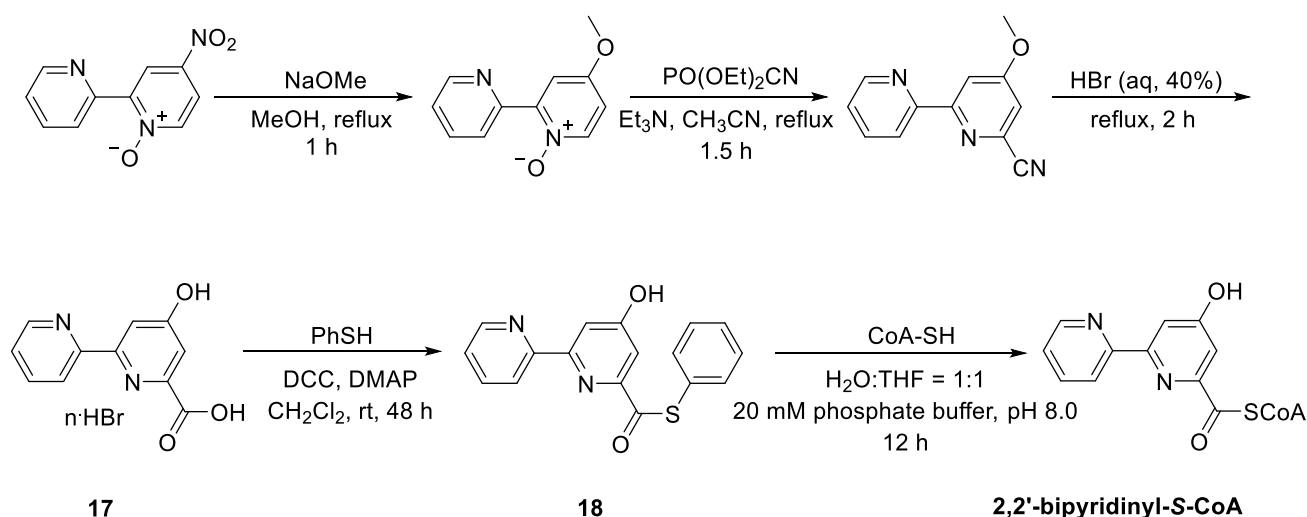
The culture of each *E. coli* transformant was incubated in Luria-Bertani (LB) medium containing 50 μg/mL kanamycin at 37°C and 220 rpm until the cell density reached 0.6-0.8 at OD₆₀₀. Protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to the final concentration of 0.1 mM, followed by further incubation for 20 h at 25°C. The cells were harvested by centrifuging at 5000 rpm for 20 min at 4°C and were re-suspended in 30 mL of lysis buffer (50 mM Tris-HCl, 300 mM NaCl, 10% glycerol, pH 7.8). After disruption by FB-110X Low Temperature Ultra-pressure Continuous Flow Cell Disrupter (Shanghai Litu Mechanical Equipment Engineering Co., Ltd, China), the soluble fraction was collected and subjected to purification of each target protein using a HisTrap FF column (GE Healthcare, USA). The desired protein fractions, as determined by SDS-PAGE, were concentrated and desalted using a PD-10 Desalting Column (GE Healthcare, USA) according to the manufacturer's protocols. The concentration of protein was determined by Bradford

assay using bovine serum albumin (BSA) as the standard.

Sequence analysis. Open reading frames (ORFs) were identified using the FramePlot 4.0beta program (<http://nocardia.nih.go.jp/fp4/>). The deduced proteins were compared with other known proteins in the databases using available BLAST methods (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Amino acid sequence alignments were performed using Vector NT1 and ESPript 3.0 (<http://esprict.ibcp.fr/ESPript/ESPript/>).

Chemical synthesis of 4-hydroxy-2,2'-bipyridine-6-carboxyloyl-S-CoA (2,2'-bipyridinyl-S-CoA).

The precursor 4-hydroxy-2,2'-bipyridinyl-6-carboxylic acid hydrobromide (**17**) was synthesized according to the method described previously⁴.



4-Hydroxy-2,2'-bipyridinyl-6-carboxylic acid hydrobromide (415 mg) was suspended in 15 mL of anhydrous CH₂Cl₂. Then, 721 mg of dicyclohexylcarbodiimide (DCC, 3.5 mmol), 0.58 mL of thiophenol (5.6 mmol) and 427 mg of 4-dimethylaminopyridine (DMAP, 3.5 mmol) were added into

this CH₂Cl₂ suspension stirred at 0°C under argon atmosphere. The mixture was stirred at room temperature for 48 h before adding 0.3 mL of AcOH. After filtering, the organic phase of the reaction mixture was treated with saturated aqueous solution of NaHCO₃ and brine, and subsequent concentrated to dryness under reduced pressure. The resulting residue was re-dissolved in EtOAc, washed with 5% (w/v) citric acid solution, and filtered over diatomite. The organic phase was then dried over anhydrous Na₂SO₄ overnight, filtered, and concentrated under reduced pressure. Further purification was carried out by chromatography on silica gel using 2:1 *n*-hexane/EtOAc as the eluent, followed by concentration to yield *S*-phenyl 4-hydroxy-2,2'-bipyridinyl-6-carbothioate (**18**, 50.2 mg). for the purified compound, ¹H NMR (500 MHz, DMSO-*d*₆) δ 11.48 (s, 1H), 8.72 (ddd, *J* = 4.3, 1.5, 0.8 Hz, 1H), 8.49 (d, *J* = 7.9 Hz, 1H), 8.07 (d, *J* = 2.3 Hz, 1H), 8.04 (td, *J* = 7.7, 1.8 Hz, 1H), 7.57 – 7.48 (m, 6H), 7.30 (d, *J* = 2.3 Hz, 1H); and ¹³C NMR (125 MHz, DMSO-*d*₆) δ 191.7, 166.9, 157.1, 154.2, 152.4, 149.9, 138.1, 135.3, 129.9, 129.7, 128.6, 125.3, 121.2, 112.1, 108.5. ESI-HRMS Calcd. for C₁₇H₁₃N₂O₂S⁺ 309.0698 [M+H]⁺, found 309.0696.

Next, 30 mg of CoA-SH (0.039 mmol, Sangon Biotech, China) was dissolved in 1.8 mL phosphate buffer (40 mM, pH = 8.0), and the pH was adjusted to 8.0 with 1 M NaOH, followed by the addition of the solution of 18 mg of the above obtained *S*-phenyl 4-hydroxy-2,2'-bipyridinyl-6-carbothioate in 1.8 mL THF, stirring at room temperature under argon atmosphere for 12 h. After removal of THF under reduced pressure, the mixture was washed with ether (to remove the unreacted thioester), and then subjected to semi-preparation by HPLC to give 2,2'-bipyridinyl-*S*-CoA as a white powder (8 mg). This semi-preparation was carried out on an Agilent Zorbax column (SB-C18, 5 μm, 9.4 x 250 mm, Agilent Technologies Inc., USA) by gradient elution of solvent A (H₂O containing 5 mM NH₄Ac) and

solvent B (CH₃CN) at a flow rate of 3 mL/min over a 30-min period as follows: T = 0 min, 5% B; T = 2 min, 5% B; T = 20 min, 90% B; T = 25 min, 90% B; T = 26 min, 5% B and T = 30 min, 5% B (mAU at 254 nm). For purified 2,2'-bipyridinyl-S-CoA, ESI-HRMS Calcd. for C₁₇H₁₃N₂O₂S⁺ 966.1682 [M+H]⁺, found 966.1670.

The CAE 2,2'-bipyridine intermediate **1** and the COL 2,2'-bipyridine intermediate **16** were synthesized using the methods previously reported⁴.

Site-specific mutation of CaeA2 and its truncated proteins. Rolling cycle PCR amplification (using the primers listed in **Supplementary Table 3**) followed by subsequent DpnI digestion was conducted according to the standard procedure of the Mut Express II Fast Mutagenesis Kit (Vazyme Biotech Co. Ltd, China). The yielded recombinant plasmids were listed in **Supplementary Table 2**. Each mutation was confirmed by Sanger sequencing. CaeA2^{F2042L}, CaeA2^{F2042L}ΔCt, CaeA2^{F2042I}, and CaeA2^{F2042V} were expressed in *E.coli* BAP1. All the mutant proteins were purified to homogeneity, and then concentrated according to the procedures for the native proteins described above.

Determination of the flavin cofactor. Each protein (CaeB1 or ColB1) solution at the concentration of 1 mg/ml was incubated at 100°C for 5 min for denaturation and then subjected to HPLC-DAD analysis on an Agilent Zorbax column (SB-C18, 5 μm, 4.6 x 250 mm, Agilent Technologies Inc., USA) by gradient elution of solvent A (H₂O containing 20 mM ammonium acetate) and solvent B (CH₃CN) at a flow rate of 1 mL/min over a 35-min period as follows: T = 0 min, 5% B; T = 2 min, 5% B; T = 20 min, 90% B; T = 25 min, 90% B; T = 30 min, 5% B and T = 35, 5% B (λ at 448 nm), using

standard FAD as control. The supernatant of CaeB1 or ColB1 was subjected to ESI-HRMS analysis to confirmed the identity of FAD (ESI m/z $[M+H]^+$, calcd. 786.1644; found 786.1593).

Reconstitution of the 2,2'-bipyridine assembly line *in vitro*. The initial reaction was conducted at 30°C for 1 h in a 100 μ L reaction mixture containing 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 1 mM picolinic acid, 1 mM malonyl-CoA, 100 μ M L-cysteine, 1 mM L-leucine, 10 μ M CaeA1, 1 μ M CaeA2, 1 μ M CaeA3, and 4 mM ATP. To determine the necessary *trans* partner, CaeB1, CaeA4 or both of them were added into the above mixture, respectively, with the final concentration of 1 μ M for each protein. Each reaction was quenched with 100 μ L of CH₃CN after incubation. To examine the production of the dethiolated 2,2'-bipyridine intermediate **1**, reaction mixtures were subjected to HPLC analysis on an Agilent Zorbax column (SB-C18, 5 μ m, 4.6 x 250 mm, Agilent Technologies Inc., USA) using a DAD detector, by gradient elution of solvent A (H₂O containing 0.1% TFA) and solvent B (CH₃CN containing 0.1% TFA) at a flow rate of 1 mL/min over a 35-min period as follows: T = 0 min, 5% B; T = 2 min, 5% B; T = 20 min, 90% B; T = 25 min, 90% B; T = 30 min, 5% B and T = 35, 5% B (λ at 315 nm). For HPLC-ESI-MS analysis, TFA was replaced by 0.1% formic acid.

To evaluate the participation of the enzymes and substrates in the production of **1**, each component was omitted from the 100 μ L reaction mixture that contains 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 1 mM picolinic acid, 1 mM malonyl-CoA, 100 μ M L-cysteine, 1 mM L-leucine, 10 μ M CaeA1, 1 μ M CaeA2, 1 μ M CaeA3, 1 μ M CaeB1, and 4 mM ATP. To examine O₂ dependence, gas exchange for O₂ elimination was conducted in an anaerobic glovebox overnight before incubation under anaerobic conditions. To evaluate the mutation effects of CaeA2 on **1** production, CaeA2 was

replaced with its variants CaeA2^{F2042L}, CaeA2^{F2042I}, CaeA2^{F2042V}, CaeA2-ΔCt or CaeA2^{F2042L}ΔCt. To mechanistically trace the transformation process, L-[1,2,3-¹³C₃, ¹⁵N]cysteine and L-[2,3,3-D₃]cysteine were used to replace unlabeled L-cysteine, respectively. To evaluate the changeability of the enzymes in the reaction mixture, 1) CaeB1 was replaced with ColB1 (with the final concentration of 10 μM); and 2) CaeA3 was replaced with ColA3 and CaeA3_CColA3_{A-PCP}, respectively. Reactions are conducted at 30°C for 1 h and then quenched with 100 μL of CH₃CN. The production of **1** was monitored by HPLC or HPLC-ESI-MS under conditions as described above.

For H₂S examination during **1** production, the reaction was conducted at 30°C for 1 h in the 100 μL, TCEP-involving reaction mixture that contains 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 1 mM picolinic acid, 1 mM malonyl-CoA, 100 μM L-cysteine, 1 mM L-leucine, 10 μM CaeA1, 1 μM CaeA2, 1 μM CaeA3, 1 μM CaeB1, and 4 mM ATP. Then, the TCEP derivative **2** was analyzed by HPLC-ESI-MS under conditions as described above. Alternatively, H₂S examination was conducted in the reaction mixture where TCEP was omitted, in the presence of **19** (1 μM), a dual-reactable fluorescent probe used for highly selective and sensitive detection of biological H₂S⁵. The reaction of **19** with Na₂S to yield **20** serves as the control reaction. **20** was examined by HPLC-FLD under conditions as described above with excitation at 370 nm and relative emission at 450 nm.

To examine whether the CAE assembly line provides the thiolated intermediate in COL biosynthesis, ColG2 (with the final concentration of 1 μM) was added into the 100 μL reaction mixture that contains 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 1 mM picolinic acid, 1 mM malonyl-CoA, 100 μM L-cysteine, 1 mM L-leucine, 10 μM CaeA1, 1 μM CaeA2, 1 μM CaeB1, 1 μM CaeA3 (ColA3

or CaeA3_CColA3_{A-PCP}), and 4 mM ATP. The production of the thiolated 2,2'-bipyridine intermediate **16** was examined by HPLC or HPLC-ESI-MS under conditions as described above.

For sulfhydryl-2,2'-bipyridinyl-L-leucine (**15**) examination, the reaction was conducted at 30°C for 1 h in the 50 µL reaction mixture contains 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 1 mM picolinic acid, 1 mM malonyl-CoA, 100 µM L-cysteine, 1 mM L-leucine, 10 µM CaeA1, 1 µM CaeA2, 1 µM ColA3, 1 µM CaeB1, and 4 mM ATP. The reaction mixture was treated with 1% SDS and 5 mM TCEP at 55°C for 30 min to release the free thiol of **15**, and then was incubated with 25 mM iodoacetamide (IAM) at 30°C for 30 min. The production of **17** was examined by HPLC or HPLC-ESI-MS under conditions as described above.

***In vitro* assays of PCP S-aminoacylation on CaeA2 by nanoLC-MS/MS.** The CaeA2 recombinant protein that was purified from *E.coli* BAP1 was subjected to complete or partial protease hydrolysis with trypsin, Glu-C or chymotrypsin as well as a variety of their combinations to map **SLGGDSIMGIQF₂₀₄₂VSR** of CaeA2, the MS-detectable sequence that contains the Ppant-modified active-site L-serine residue (underlined). The digestion mixtures were filtrated using Microcon YM-10 (MilliporeSigma, USA) by centrifugation and stored at - 80°C before analysis. For nanoLC-MS/MS analysis, each sample was loaded on a trap column (75 µm i.d., 2 cm, C18, 5 µm, 100 Å, Thermo Fisher Scientific Inc., USA) for online desalting, and then was separated using a reversed phase column (75 µm i.d., 10.2 cm, C18, 3 µm, 120 Å, New Objective Inc., USA) by gradient elution of solvent A (H₂O containing 0.1% formic acid) and solvent B (80% CH₃CN containing 0.1% formic acid) at a flow rate of 300 nL/min over a 1.5 h period as follows: T = 0 min, 18% B; T = 45 min, 45% B; T = 50 min,

100% B; and T = 90 min, 100% B. For MS analysis, the nano-ESI voltage and capillary temperature were set at 2.2 kV and 275°C, respectively. The MS data were acquired in data-dependent mode. Each full-scan MS (m/z 350–2000, resolution of 60 k) was followed with 10 HCD MS/MS scans (normalized collision energy of 33, resolution of 15 k) for the most intense precursor ions. The maximum ion injection time for MS and MS/MS were 50 and 45 ms, and the auto gain control target for MS and MS/MS were 3×10^6 and 5×10^4 , respectively. The dynamic-exclusion time was set as 40 s.

To obtain the sequence **SLGGDSIMGIQF₂₀₄₂VSR**, the 50 μ L solution that contains 50 mM Tris-HCl (pH 7.5) and 5 μ M CaeA2 was treated with 4 μ g of trypsin (sequencing grade, Promega Corp., USA) at 30°C for 20 min (leading to complete digestion) and then with 0.6 μ g of chymotrypsin (sequencing grade, Promega Corp., USA) at 30°C for 10 min (leading to partial digestion that retains the C-terminal sequence **F₂₀₄₂VSR**). To obtain the engineered sequence **SLGGDSIMGIQL₂₀₄₂VSR**, the CaeA2^{F2042L} recombinant protein that was purified from *E.coli* BAP1 and its variants different in *S*-(amino)acylation underwent complete digestion with trypsin (4 μ g) and chymotrypsin (0.6 μ g) at 30°C for 20 min.

To identify the target sequence that contains the Ppant-modified active-site L-serine residue, the raw MS data was processed using the pFind software by setting Ppant as a variable posttranslational modification. The selected sequence was then validated by HR-MS/MS analysis of the parent ion, its associated fragmented peptide ions and particularly the characteristic Ppant ejection ion⁶.

To prepare L-cysteinyl-S-CaeA2^{F2042L} (**3**), the reaction was conducted at 30°C for 10 min in a 50 µL reaction mixture containing 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 100 µM L-cysteine, 5 µM CaeA2^{F2042L} (or CaeA2^{F2042L}ΔCt) and 4 mM ATP. To prepare (3-sulfhydryl)-pyruvoyl-S-CaeA2^{F2042L} (**6**), the reaction was conducted at 30°C for 10 min in a 50 µL reaction mixture containing 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 100 µM L-cysteine, 5 µM CaeA2^{F2042L} (or CaeA2^{F2042L}ΔCt), 1 µM CaeB1 and 4 mM ATP. To prepare 2,2'-bipyridinyl-S-CaeA2^{F2042L} (**9**), the reaction was conducted at 30°C for 10 min in a 50 µL reaction mixture containing 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 1 mM picolinic acid, 1 mM malonyl-CoA, 100 µM L-cysteine, 1 µM CaeA1, 1 µM CaeB1, 5 µM CaeA2^{F2042L} (or CaeA2^{F2042L}ΔCt) and 4 mM ATP. For isotope labeling, L-[1,2,3-¹³C₃,¹⁵N]cysteine and L-[2,3,3-D₃]cysteine were used to replace unlabeled L-cysteine, respectively. For thiol derivatization, each reaction mixture was treated with 10 mM iodoacetamide (IAM) at 30°C for 2 min. Protease digestion and subsequent nanoLC-MS/MS analyses were conducted using approaches as described above.

To probe potential intermediates, all MS/MS data of the derivatives from the sequence (SLGGDSIMGIQL₂₀₄₂VSR) were extracted from raw MS data using script written by Python (<https://github.com/billpb610/AlFinder/blob/master/AlFinderS2.py>). The script also recorded information about precursor ions, fragment ion and Pparent ejection ions for further analysis.

***In vitro* preparation of 2,2'-bipyridinyl-S-CaeA2^{F2042L} (**9**).** To prepare **9** as a positive control in S-aminoacylation assays, the reaction was conducted at 30°C for 30 min in a 50 µL reaction mixture containing 50 mM Tris-HCl (pH 7.5), 1 mM TCEP, 10 mM MgCl₂, 100 µM 2,2'-bipyridinyl-S-CoA, 5 µM Sfp, and 10 µM CaeA2^{F2042L} in apo-form. Protease digestion and subsequent nanoLC-MS/MS

analyses were conducted using approaches as described above.

***In vitro* preparation of 2,2'-bipyridinyl-S-PCP_{CaeA2} (10).** The reaction was conducted at 30°C for 1 h in an 80 µL reaction mixture containing 62.5 mM Tris-HCl (pH = 7.5), 1.25 mM TCEP, 12.5 mM MgCl₂, 1 mM 2,2'-bipyridinyl-S-CoA, 100 µM PCP_{CaeA2} (derived from wild-type CaeA2) in apo-form, and 4 µM Sfp.

To produce the dethiolated 2,2'-bipyridine intermediate **1**, the reaction mixture was combined with the following components in a new 100 µL reaction mixture: 1 mM L-leucine, 5 µM CaeA3, and 4 mM ATP. After incubation at 30°C for 2 h, the reaction was quenched with 100 µL of CH₃CN for HPLC or HPLC-ESI-MS analysis under conditions as described above.

Measurement of protein-protein interactions by isothermal titration calorimetry (ITC). ITC was performed with MicroCal-ITC200 (Malvern) at 25°C. A 400 µl aliquot of 60 µM Ct or CaeA2ΔCt was placed in the stirred cell, and 120 µl aliquot of 400 µM MBP-fused CaeB1 or ColB1 was prepared in the syringe. All the recombinant proteins were prepared by Ni-affinity chromatography followed by size-exclusion chromatography (SEC) for purification, and then were exchanged in the 50 mM Tris-HCl (PH = 7.5) buffer containing 100 mM NaCl, 0.1 mM TCEP and 5 % glycerol. The titration was performed as follows: 1 µl of protein in the syringe over 0.8 s for the first injection, followed by 19 injections of 2 µl protein in the stirred cell at 120 s intervals. The heat of reaction per injection was determined by integration of the peak areas using the MicroCal-PEAQ-ITC software, which provides the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N), and the dissociation constant (K_d). The heats of dilution were determined by injecting flavoprotein alone into the buffer and

were subtracted from the corresponding experiments before curve fitting.

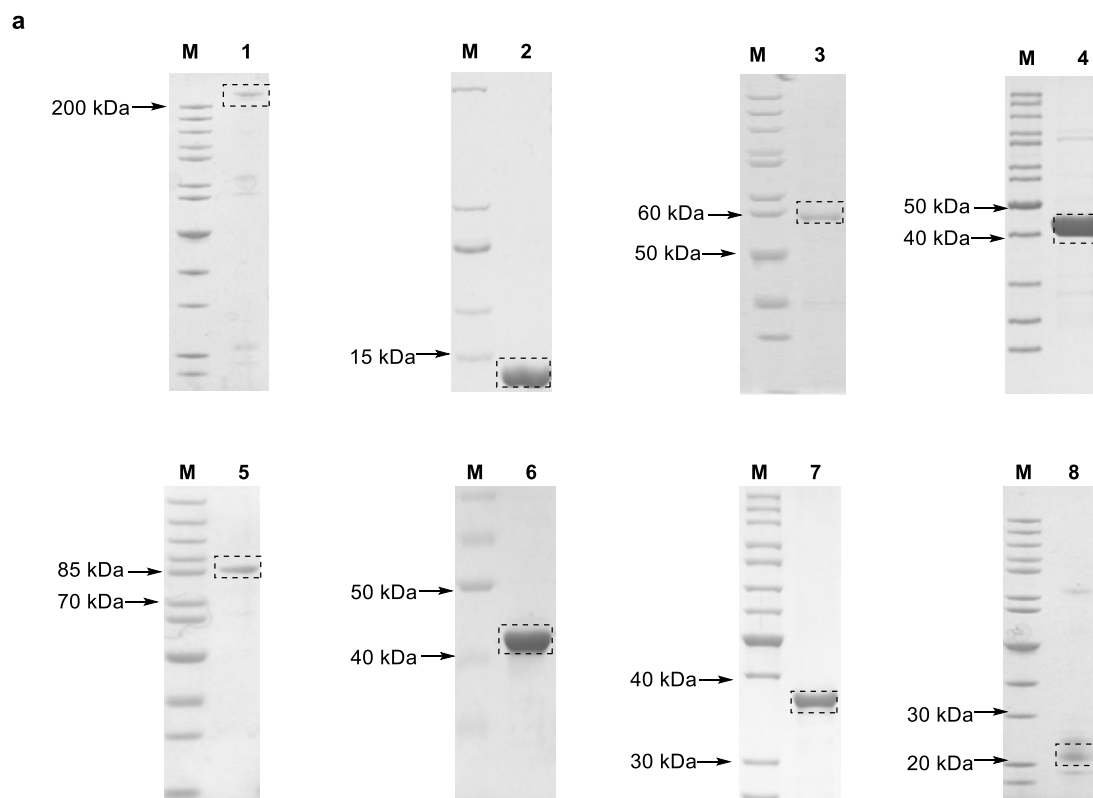
Analysis of the functional exchangeability between CaeA2 and ColA2 *in vivo*. A 4.4 kb DNA fragment amplified by PCR from the CAE-producing *A. cyanogriseus* strain using the primers cae-CO-for and cae-Ct-rev was cloned into pMD19-T, yielding pQL1053. After sequencing to confirm the fidelity, this PCR product was recovered by digestion with BglIII-EcoRV and then utilized to replace the 4.4 kb BglIII-EcoRV fragment of pQL1022⁷, a pSET152 derivative previously constructed for *colA2* expression under the control of *PermE** (the constitutive promoter for expressing the erythromycin-resistance gene in *Saccharopolyspora erythraea*). The resulting recombinant plasmid pQL1054, which carries the chimeric gene *col/caeA2* coding for the hybrid protein that harbors the PKS module from ColA2 and the NRPS module from CaeA2, was transferred by conjugation into the $\Delta colA2$ *S. roseosporus* mutant strain⁷, yielding the recombinant strain QL2006. The fermentation of QL2006 and the examination of CAE or COL production were conducted according the methods described previously⁷.

SUPPLEMENTARY FIGURES

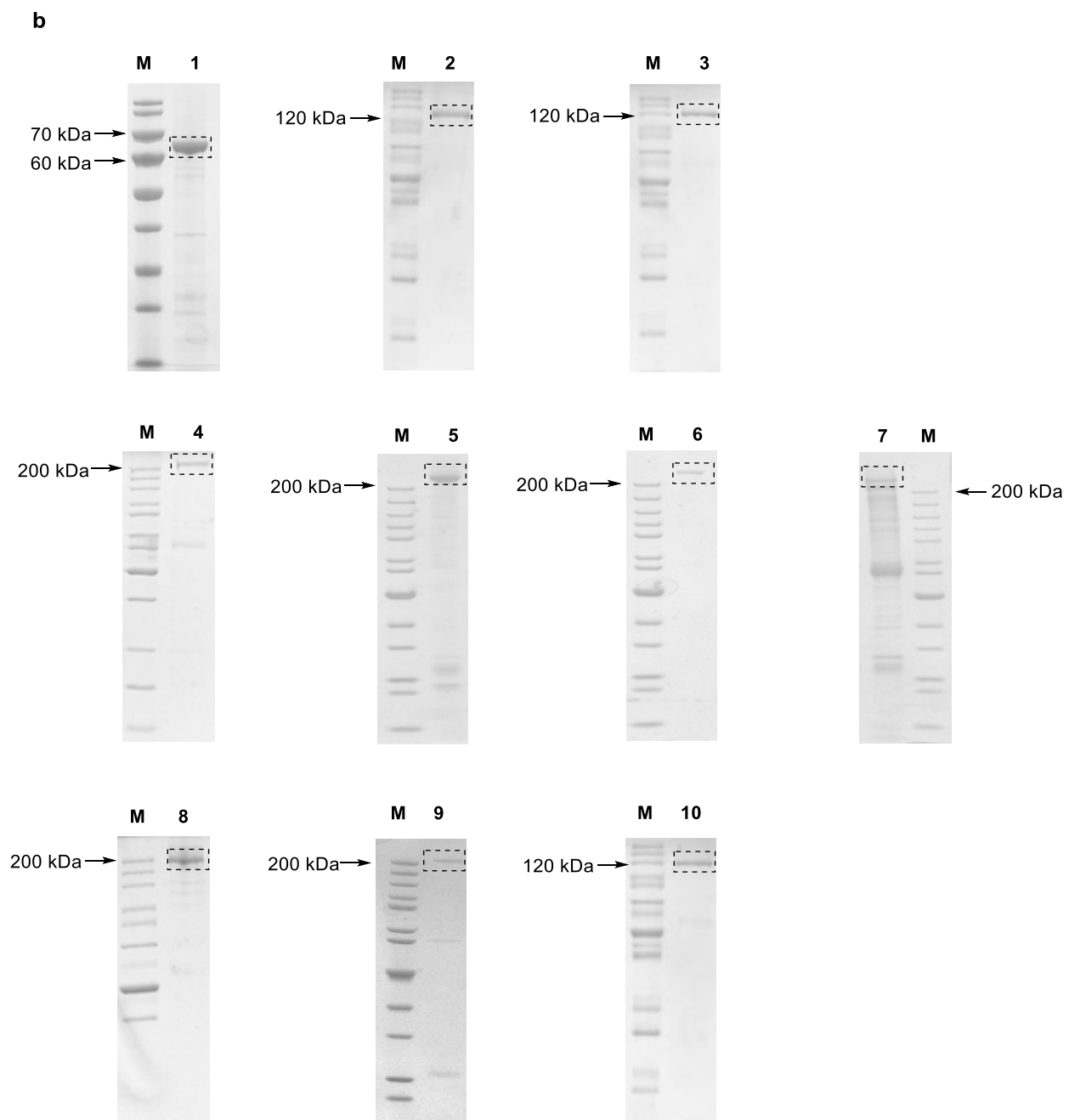
Supplementary Figure 1. Coomassie-stained SDS-PAGE analysis of purified recombinant proteins.

Lane M, protein standard. The target proteins are indicated in the dashed rectangles.

(a) Proteins produced in *E. coli* BL21(DE3). Lane 1, CaeA2^{F2042L} (269 kDa); Lane 2, PCP_{CaeA2} (derived from CaeA2, 13 kDa); Lane 3, Trx-tagged Ct_{CaeA2} (58 kDa), Lane 4, CaeB1 (43 kDa); Lane 5, MBP-fused CaeB1 (87 kDa), Lane 6, ColB1 (43 kDa); Lane 7, ColG2 (39 kDa); and Lane 8, CaeA4 (25 kDa).



(b) Proteins produced in *E. coli* BAP1. Lane 1, CaeA1 (69 kDa); Lane 2, CaeA3 (118 kDa); Lane 3, ColA3 (121 kDa); Lane 4, CaeA2 (269 kDa); Lane 5, CaeA2^{F2042L} (269 kDa); Lane 6, CaeA2^{F2042I} (269 kDa); Lane 7, CaeA2^{F2042V} (269 kDa); Lane 8, CaeA2- Δ Ct (225 kDa); Lane 9, CaeA2^{F2042L} Δ Ct (225 kDa); and Lane 10, CaeA3_CColA3_{A-PCP} (120 kDa).

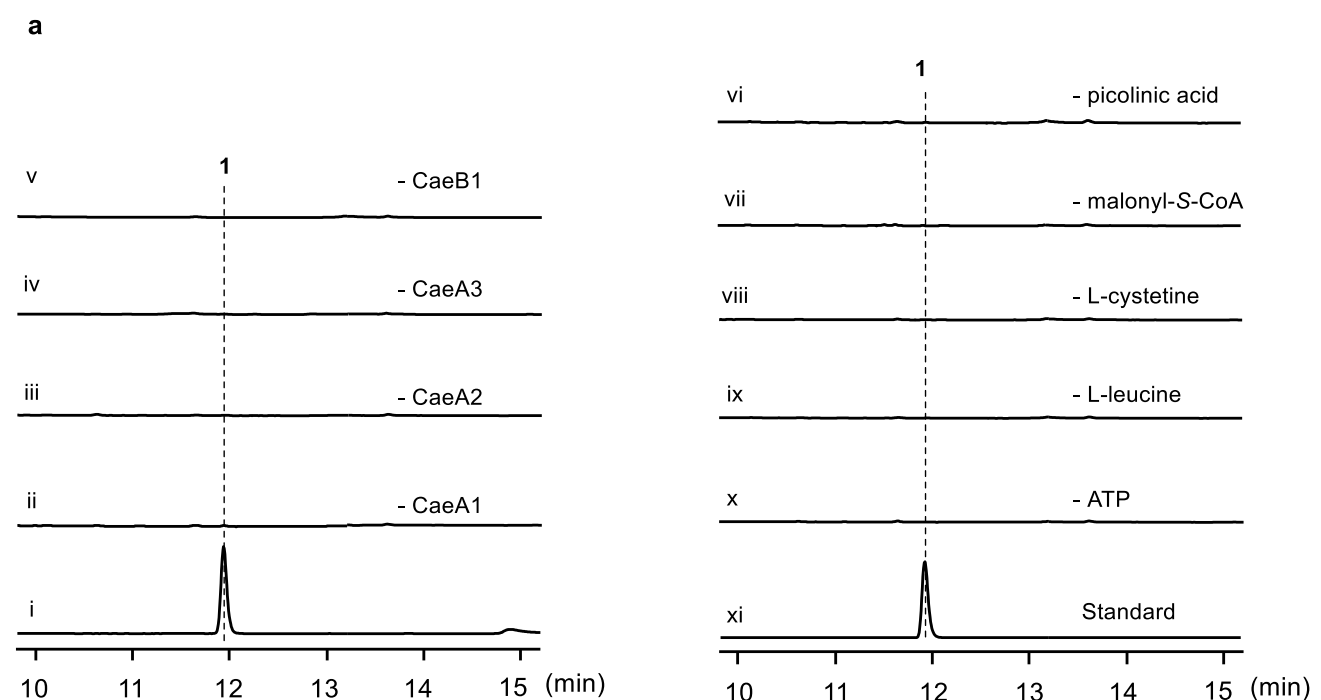


Supplementary Figure 2. Supplementary data for *in vitro* reconstitution of the 2,2'-bipyridine assembly line.

(a) Determination of component necessity in the production of the CAE 2,2'-bipyridine intermediate

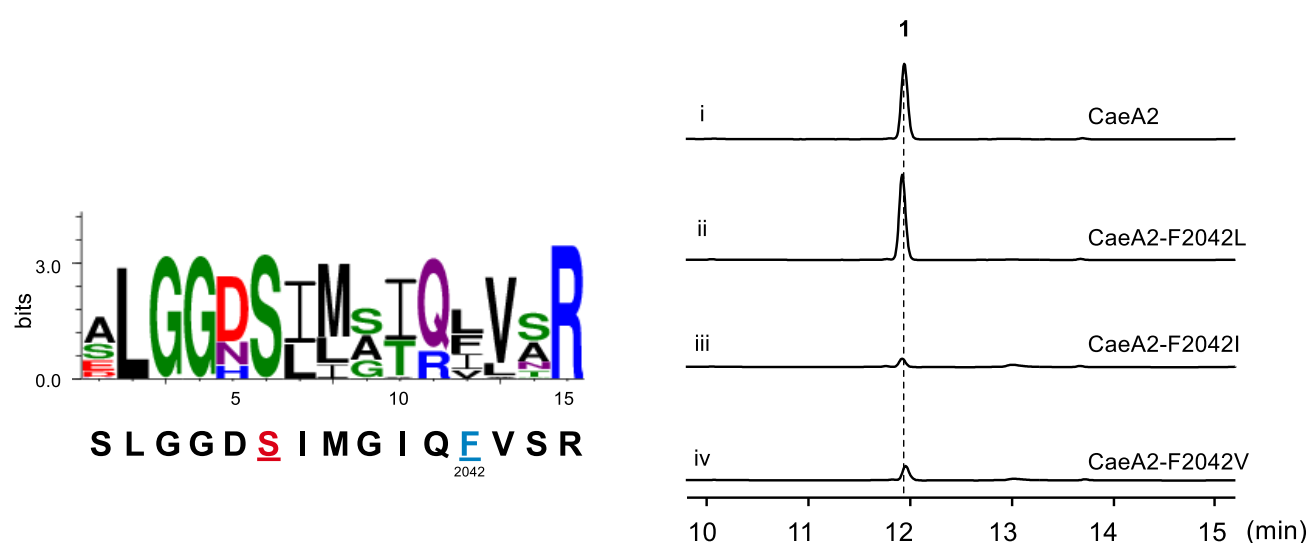
1. Tested reactions were derived from the combination of CaeA1, CaeA2, CaeA3 and CaeB1 with picolinic acid, malonyl-*S*-CoA, L-cysteine, L-leucine, and ATP (i), and included those lacking the enzyme CaeA1 (ii), CaeA2 (iii), CaeA3 (iv) or CaeB1 (v) and the substrate picolinic acid (vi), malonyl-*S*-CoA (vii), L-cysteine (viii), L-leucine (ix) or ATP (x, for the activity of the A domains), respectively.

Synthesized **1** was used as the standard (xi).



(b) Activity assays of CaeA2 variants in the production of **1**. The CaeA2 variants, i.e., CaeA2^{F2042L}, CaeA2^{F2042V} and CaeA2^{F2042I}, were designed according to residue conservative analysis (left), which was conducted by comparing the 15-aa sequence **SLGGD****S****IMGIQ****F****2042****V****SR** that is derived from the PCP domain of CaeA2 (the Ppant-modified active-site L-serine residue and the target L-phenylalanine residue are highlighted in color and underlined) with the corresponding sequences of 500 homologous PCP domains or proteins from NCBI NR database. Tested reactions were derived from the combination of CaeA1, CaeA2, CaeA3 and CaeB1 with picolinic acid, malonyl-*S*-CoA, L-cysteine, L-leucine, and ATP (i), and included those in which CaeA2 was replaced with CaeA2^{F2042L} (ii), CaeA2^{F2042V} (iii) and CaeA2^{F2042I} (iv), respectively.

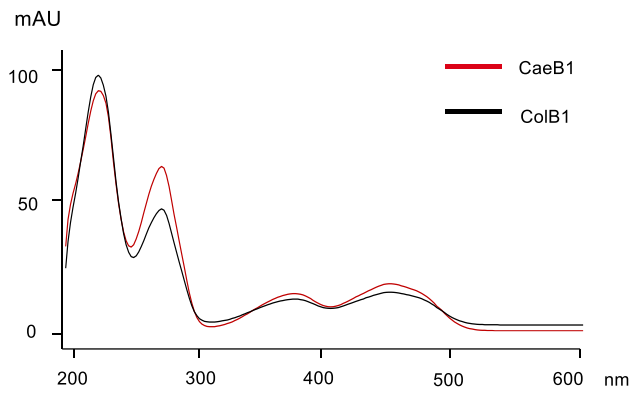
b



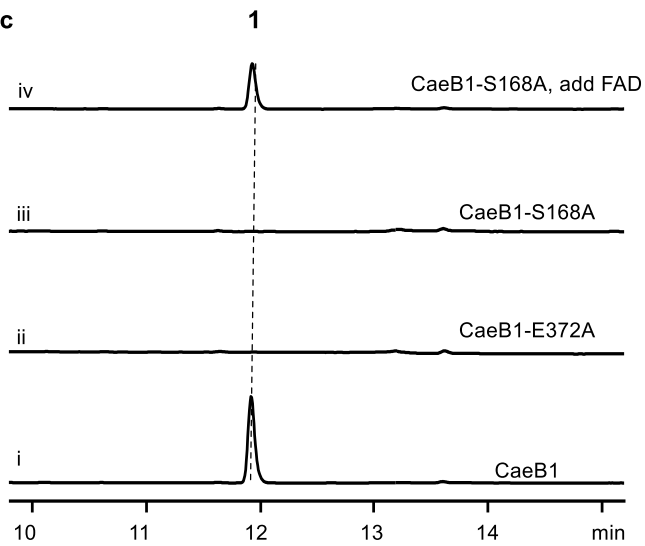
Supplementary Figure 3. Analysis of the flavoproteins CaeB1 and ColB1.

- (a) UV spectra of the recombinant proteins CaeB1 and ColB1 that were purified from *E. coli* BL21(DE3).
- (b) Determination of flavin cofactors associated with CaeB1 and ColB1. (i) authentic FMN; (ii) authentic FAD, (iii) boiled CaeB1, (iv) boiled CaeB1-S168A, (v) boiled CaeB1-E372A, and (vi) boiled ColB1. For examination by HPLC, UV absorbance at 375 nm.
- (c) Activity assays of CaeB1 variants in the production of **1**. Tested reactions were derived from the combination of CaeA1, CaeA2, CaeA3 and CaeB1 with picolinic acid, malonyl-*S*-CoA, L-cysteine, L-leucine, and ATP (i), and included those in which CaeB1 was replaced with CaeB1-E372A (ii), CaeB1-S168A (iii), and CaeB1-S168A with excess of FAD (iv), respectively.
- (d) Sequence alignment of CaeB1 and ColB1 with various FAD-dependent dehydrogenases. The homologs include FkbI (1R2G) in FK520 biosynthesis⁸, medium chain acyl-CoA dehydrogenase (MCAD, 1T9G_A)⁹, short chain acyl-CoA dehydrogenase (SCAD, 1JQI_A)¹⁰ and cyclohexanecarboxyl-CoA dehydrogenase (ChCAD, ABC76100.1)¹¹. Residues important for dehydrogenase activity, FAD-binding and phosphopentathione binding are indicated by star, black triangle and red triangle, respectively.

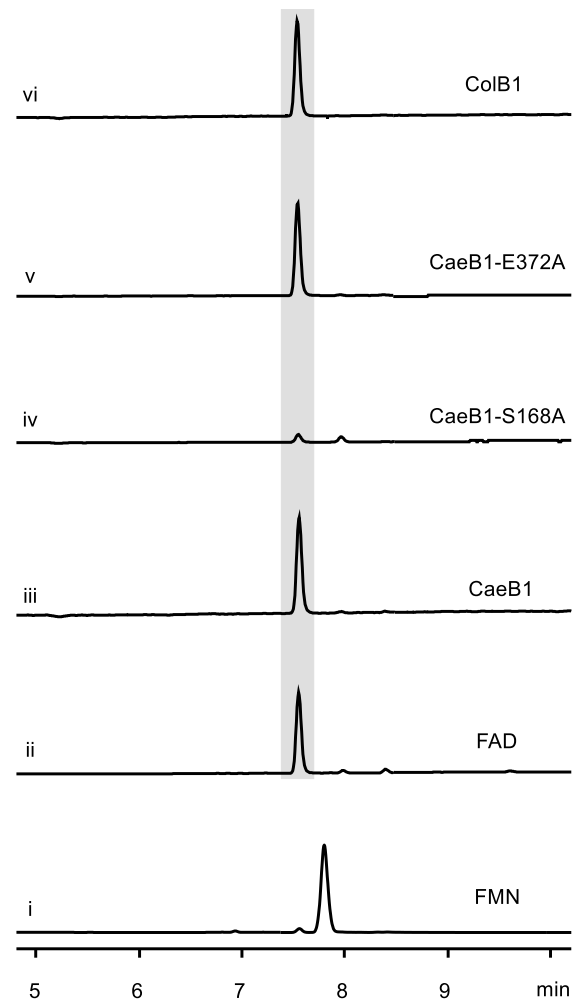
a



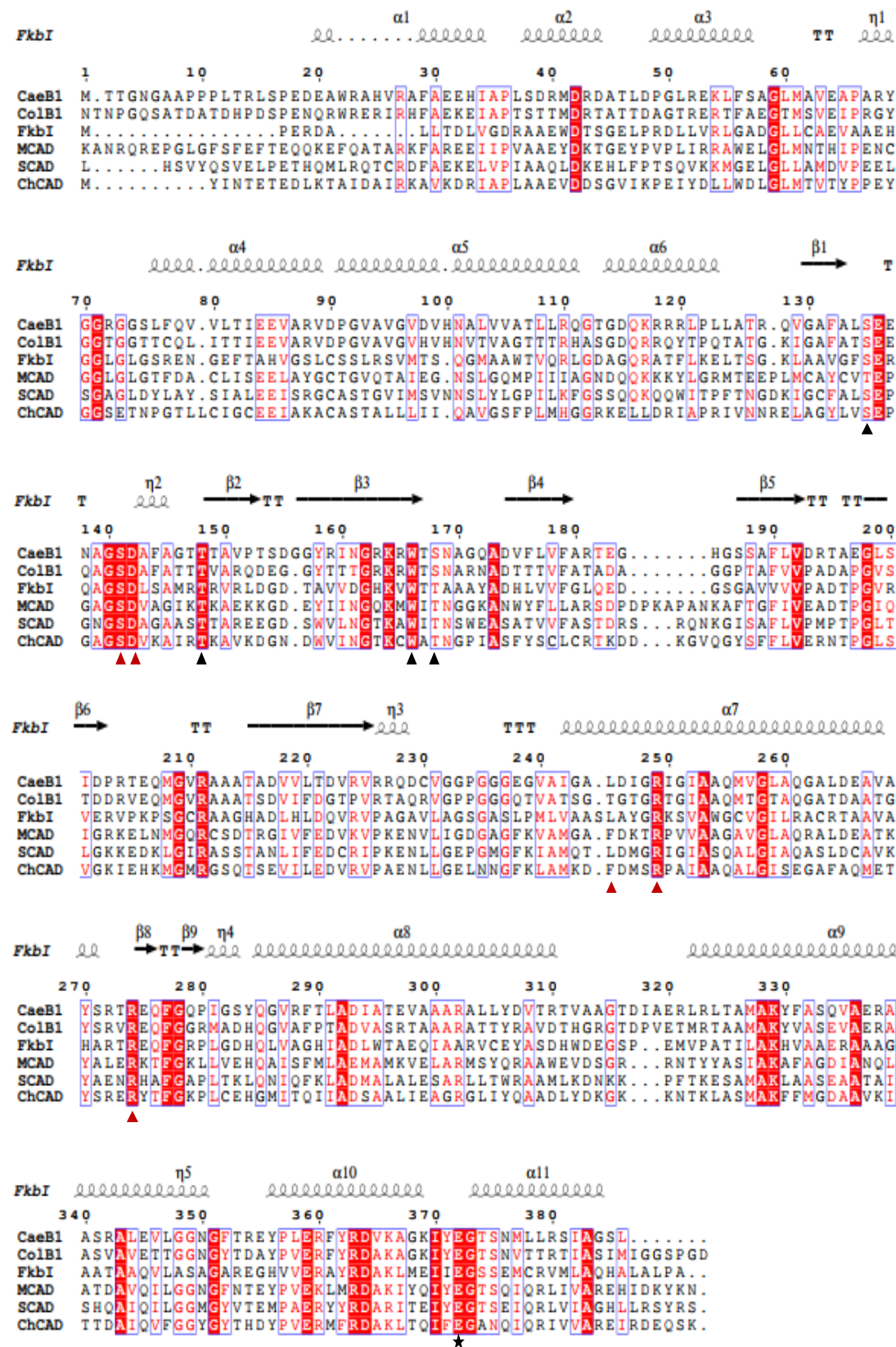
c



b

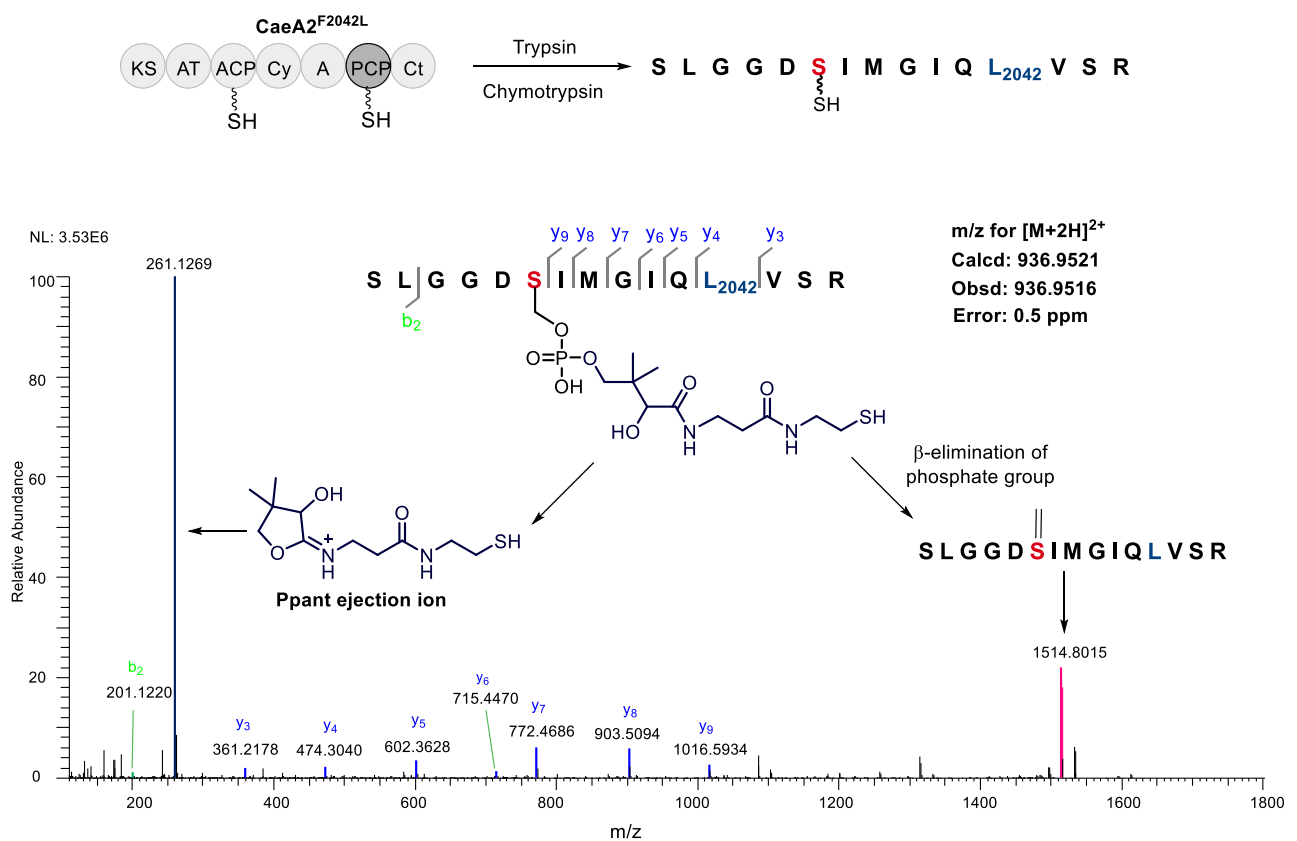


d

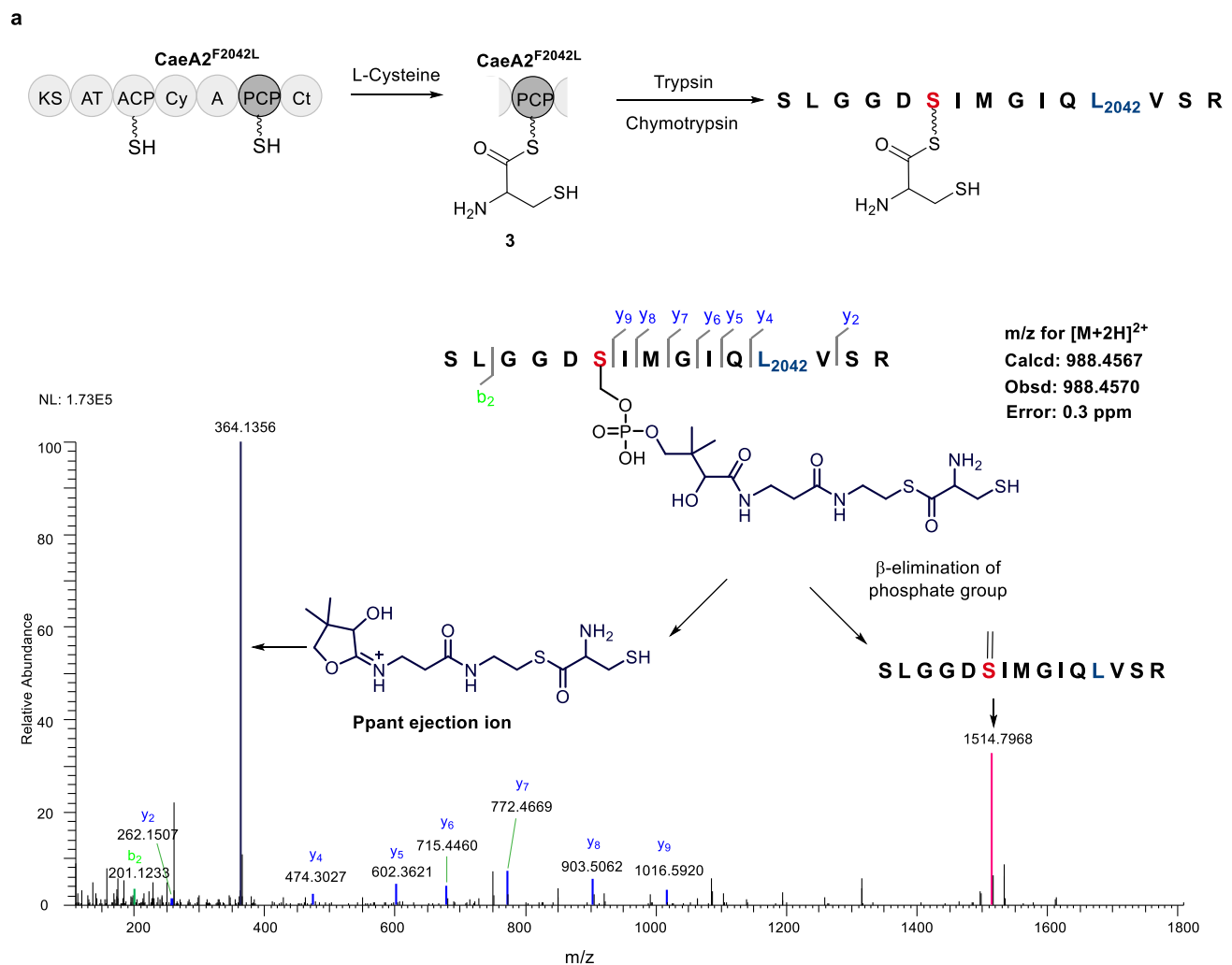


Supplementary Figure 4. Determination of the 15-aa sequence that is derived from the PCP domain of CaeA2 and contains the Ppant-modified active-site L-serine residue (red) by nanoLC-MS/MS. **(a)** Ppant-modified CaeA2 by treatment with trypsin (complete digestion) and chymotrypsin (partial digestion). **(b)** Ppant-modified CaeA2^{F2042L} by treatment with trypsin (complete digestion) and chymotrypsin (complete digestion).

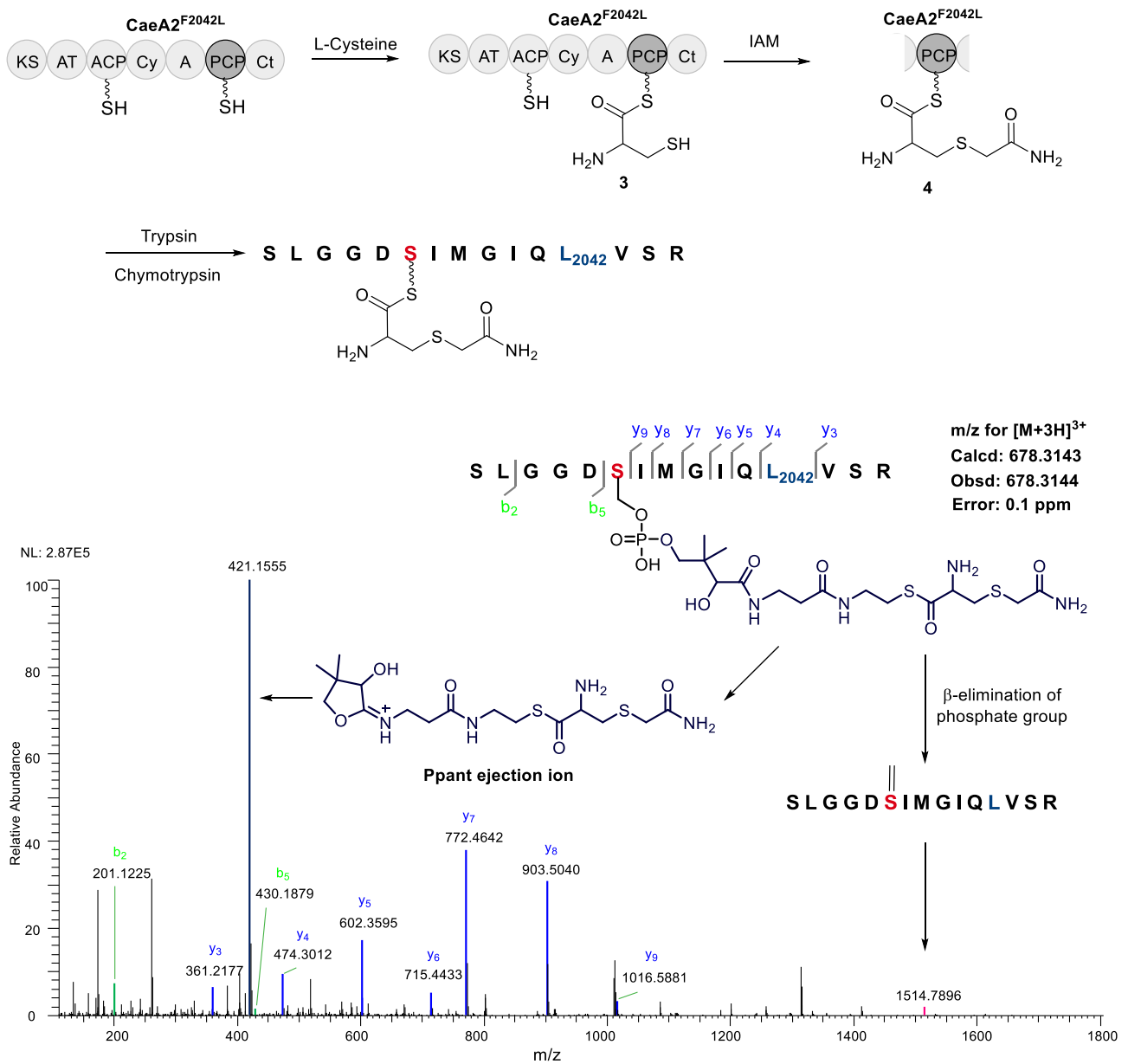
b



Supplementary Figure 5. Characterization of L-cysteinyl-S-CaeA2^{F2042L} (**3**) by nanoLC-MS/MS following complete digestion with trypsin and chymotrypsin. **(a)** Incubation of thiolated CaeA2^{F2042L} and L-cysteine. **(b)** Treatment of with IAM after the incubation of thiolated CaeA2^{F2042L} and L-cysteine.

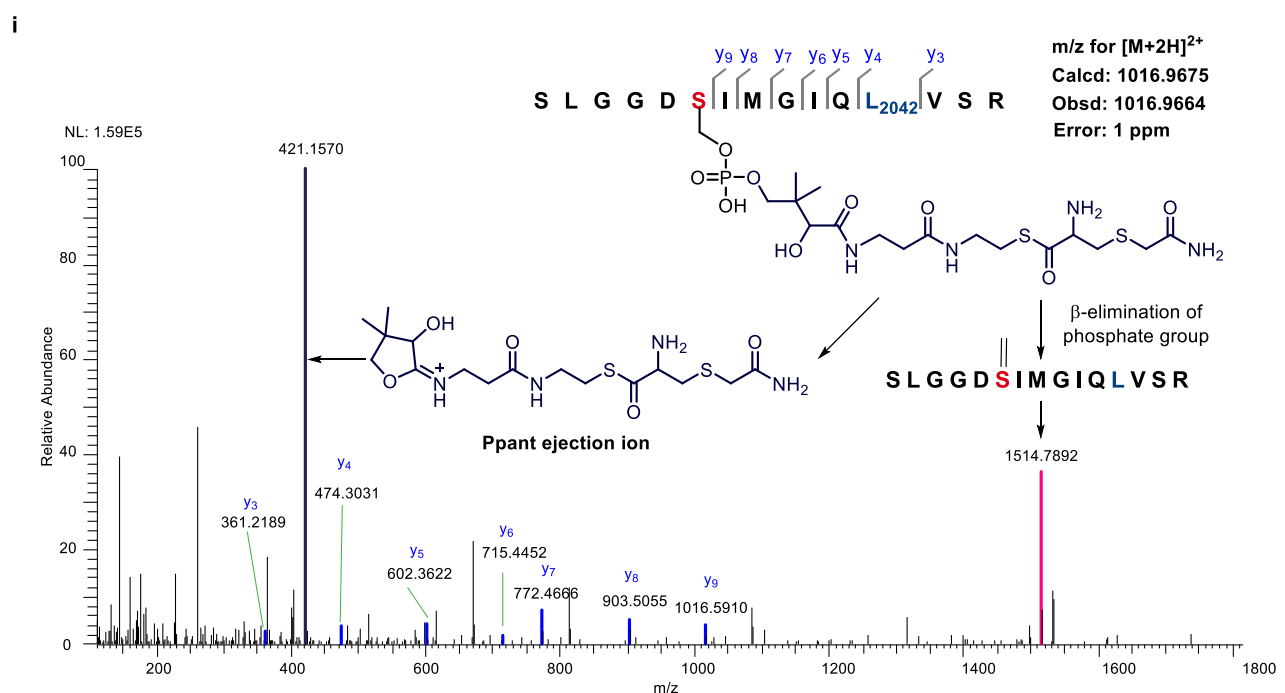
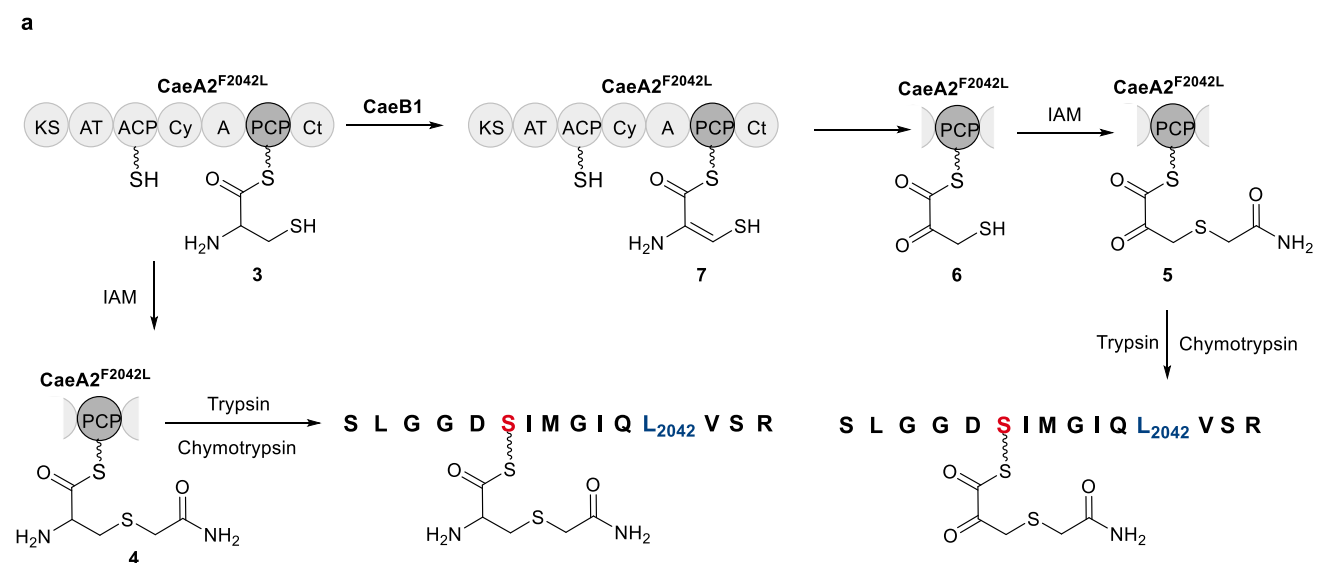


b

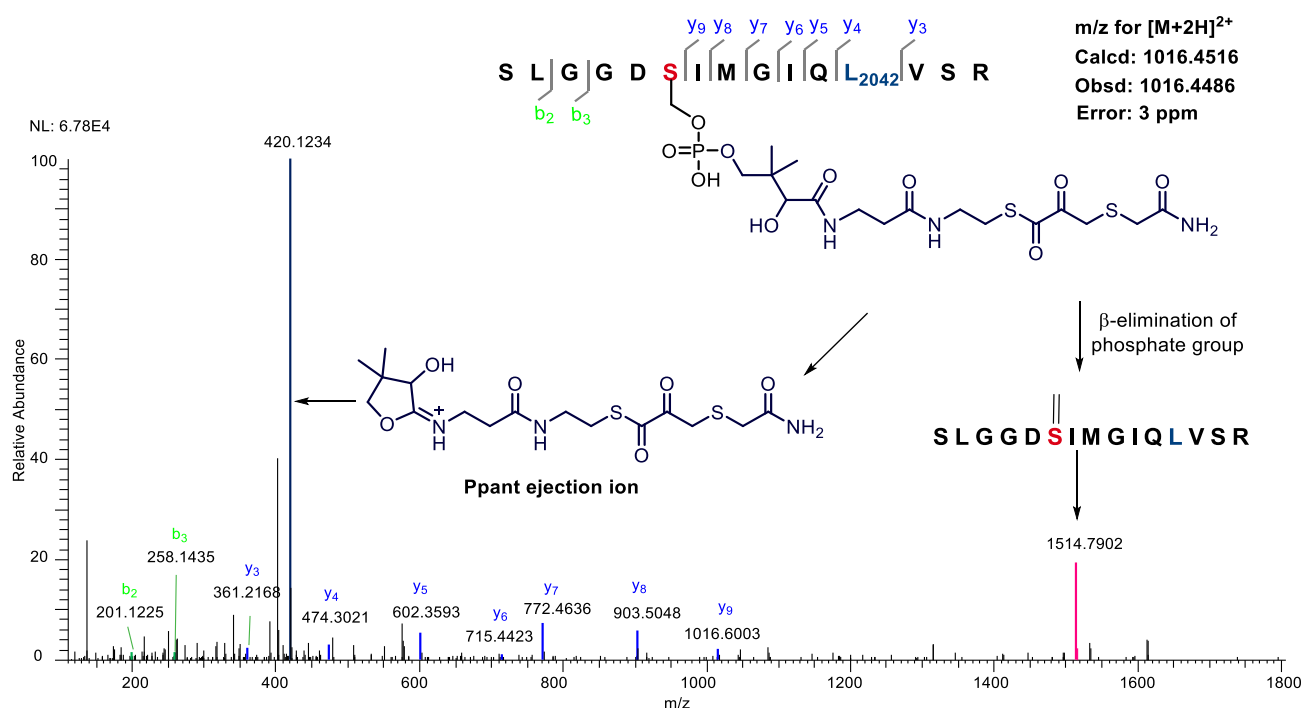


Supplementary Figure 6. Characterization of (3-sulfhydryl)-pyruvoyl-*S*-CaeA2^{F2042L} (**6**) by nanoLC-MS/MS following treatment with IAM and subsequent complete digestion with trypsin and chymotrypsin.

(a) Incubation of thiolated CaeA2^{F2042L}, CaeB1 and L-cysteine. MS-detectable products include **4** (i) and **5** (ii).

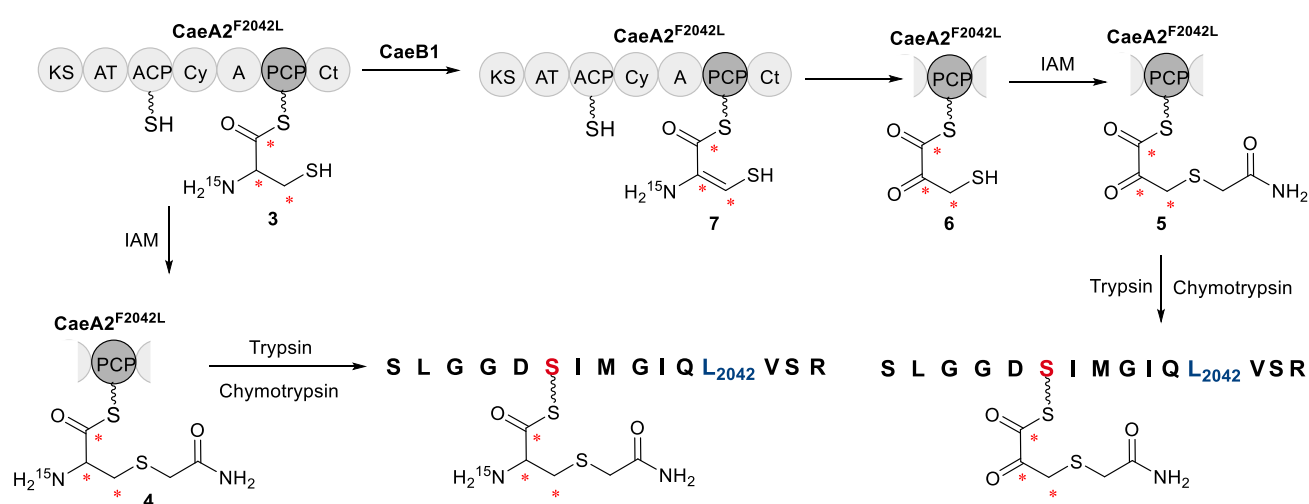


ii

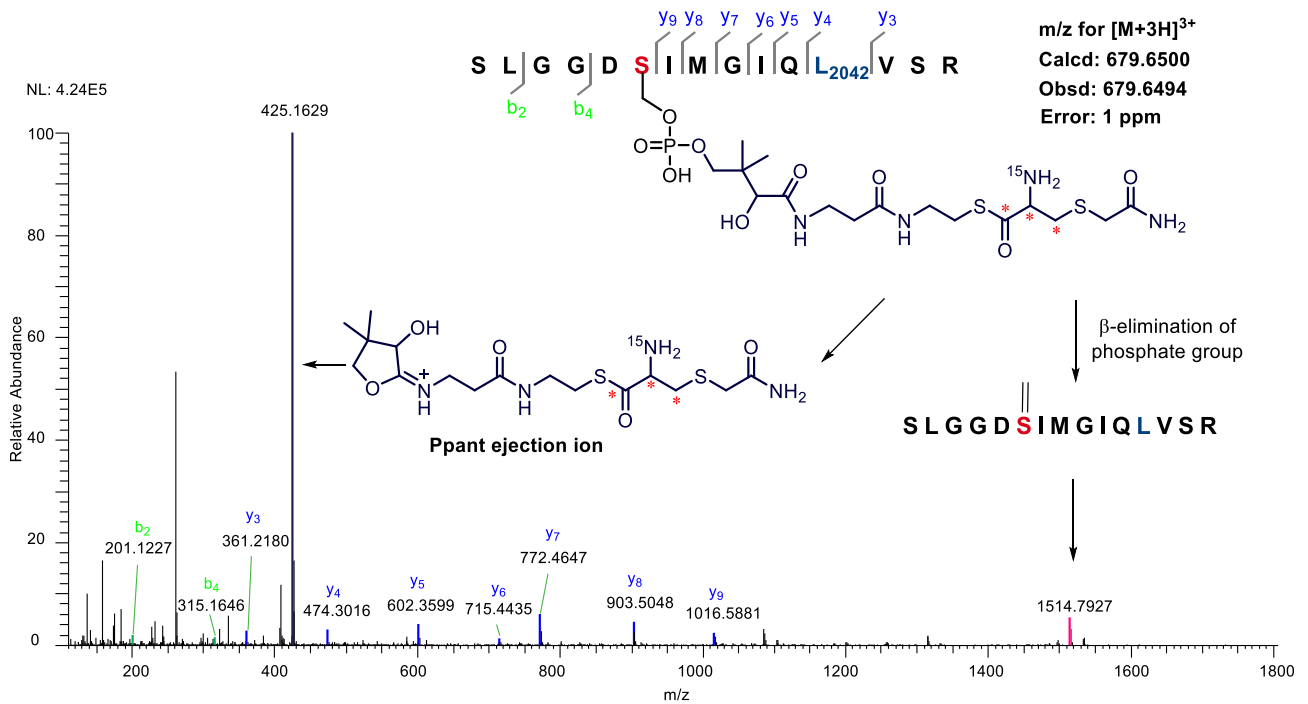


(b) Incubation of thiolated CaeA2^{F2042L}, CaeB1 and L-[1,2,3-¹³C₃,¹⁵N]cysteine. MS-detectable products include ¹⁵N and/or ¹³C labelled 4 (i) and 5 (ii).

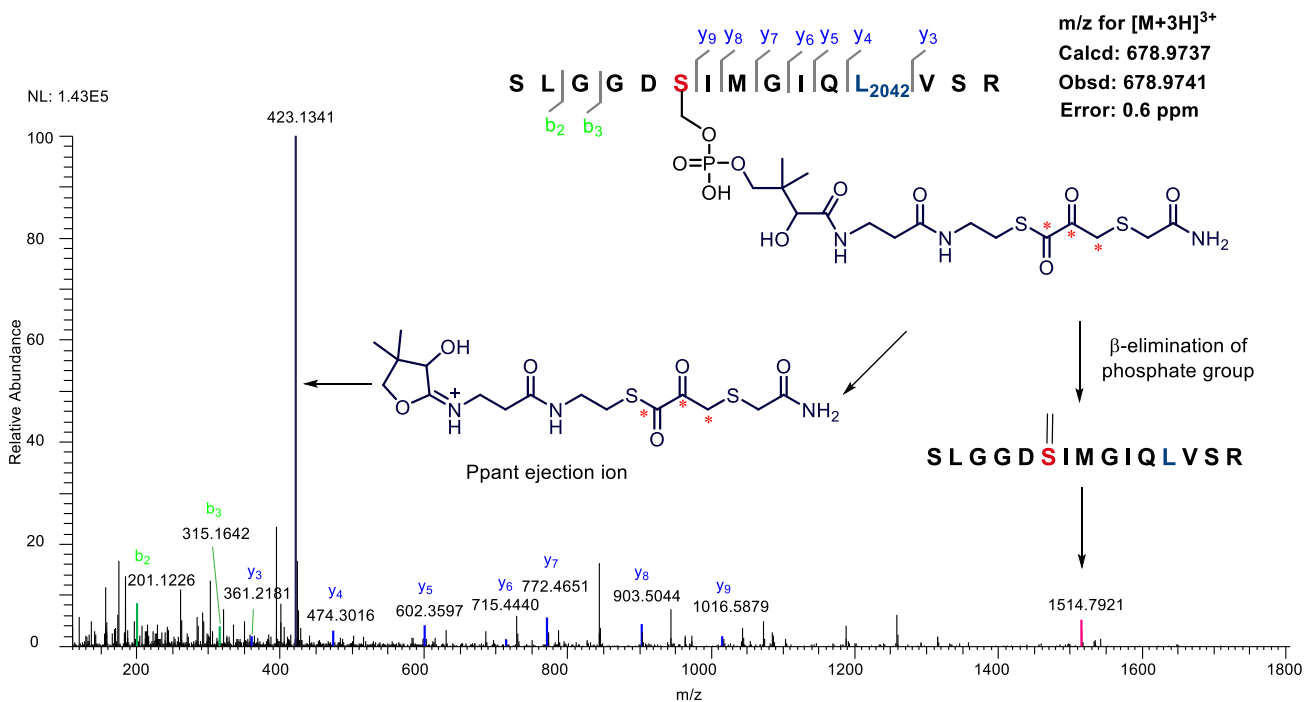
b



i



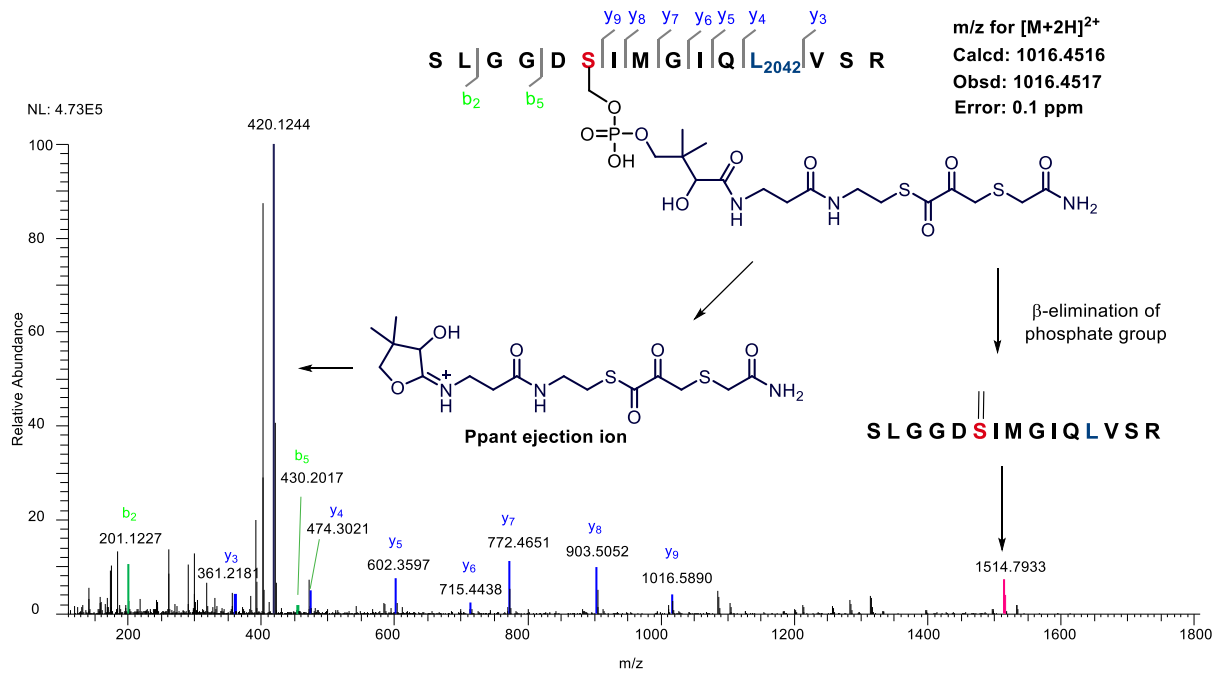
ii



(c) Incubation of thiolated CaeA2^{F2042L}, CaeB1 and L-[2,3,3-D₃]cysteine. MS-detectable products include labelled **4** (i) and unlabeled **5** (ii).



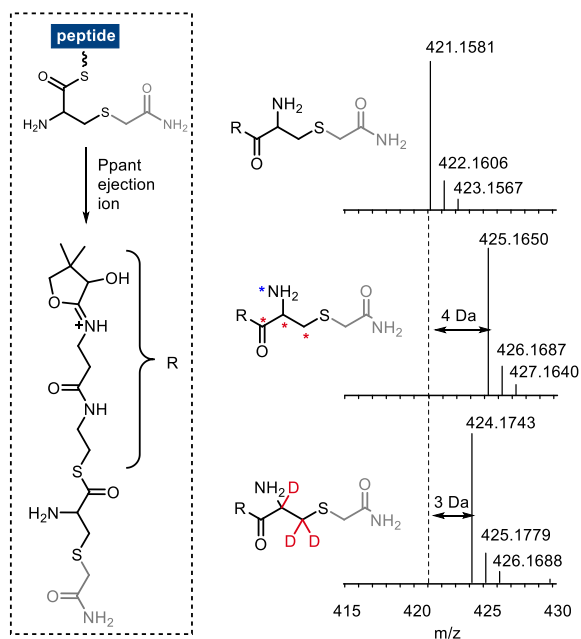
ii



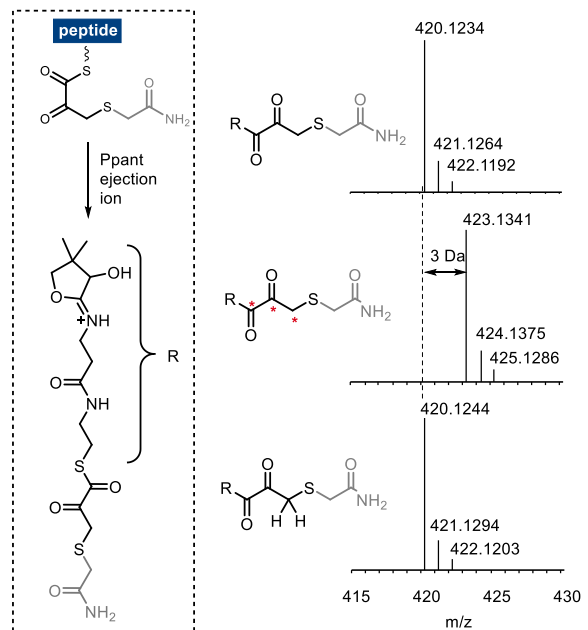
(d) MS comparison of **4** (i) and **5** (ii) in their associated Ppant ejection products, using L-cysteine (top), L-[1,2,3-¹³C₃, ¹⁵N]cysteine (middle) and L-[2,3,3-D₃]cysteine (down) as the substrates, respectively.

d

i

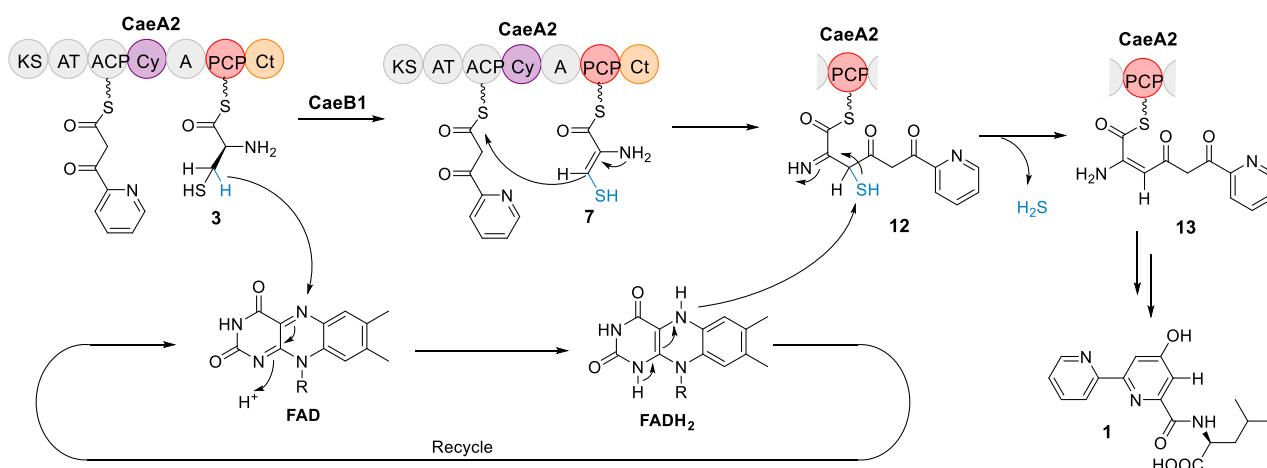


ii

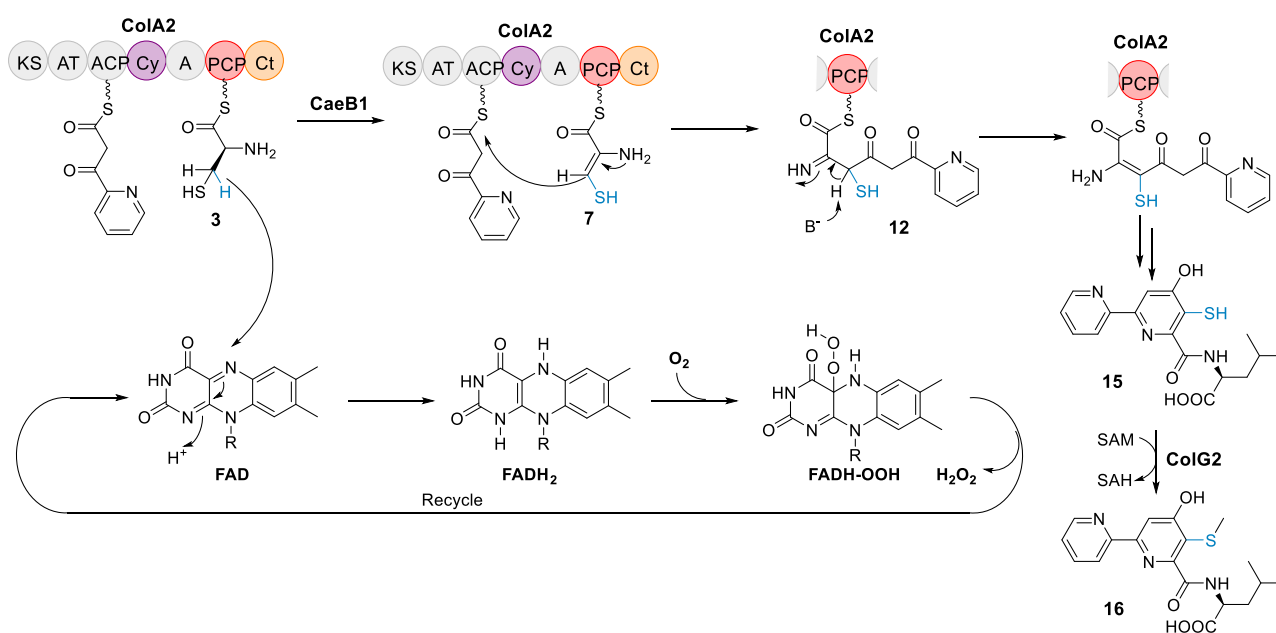


Supplementary Figure 7. Proposed mechanisms for recycling FAD in the CAEs (a) and COLs (b) biosynthetic pathways, respectively.

a



b



Supplementary Figure 8. Sequence analysis of the C domains of CaeA3 and ColA3, and the Cy and Ct domains of CaeA2 and ColA2. For comparison, the homologs include the epimerization (E) domain of the NRPS TycA (610-913 aa, AAC45928.1) in tyrocidine biosynthesis¹², the first C domain of the NRPS BacB (72-361 aa, AAC06347.1) in bacitracin biosynthesis¹³, the C domain of the NRPS MycC (855-1156 aa, AAF08797.1) in mycosubtilin biosynthesis¹⁴, and the Cy domain of the NRPS EpoB (71-364 aa, ADB12489.1) in epothilone biosynthesis¹⁵. The boundaries of each domain were determined by Pfam. The active site residue L-histidine of C domains is indicated by red star, and the active site residue L-glutamic acid of Cy domains is indicated by black dot.

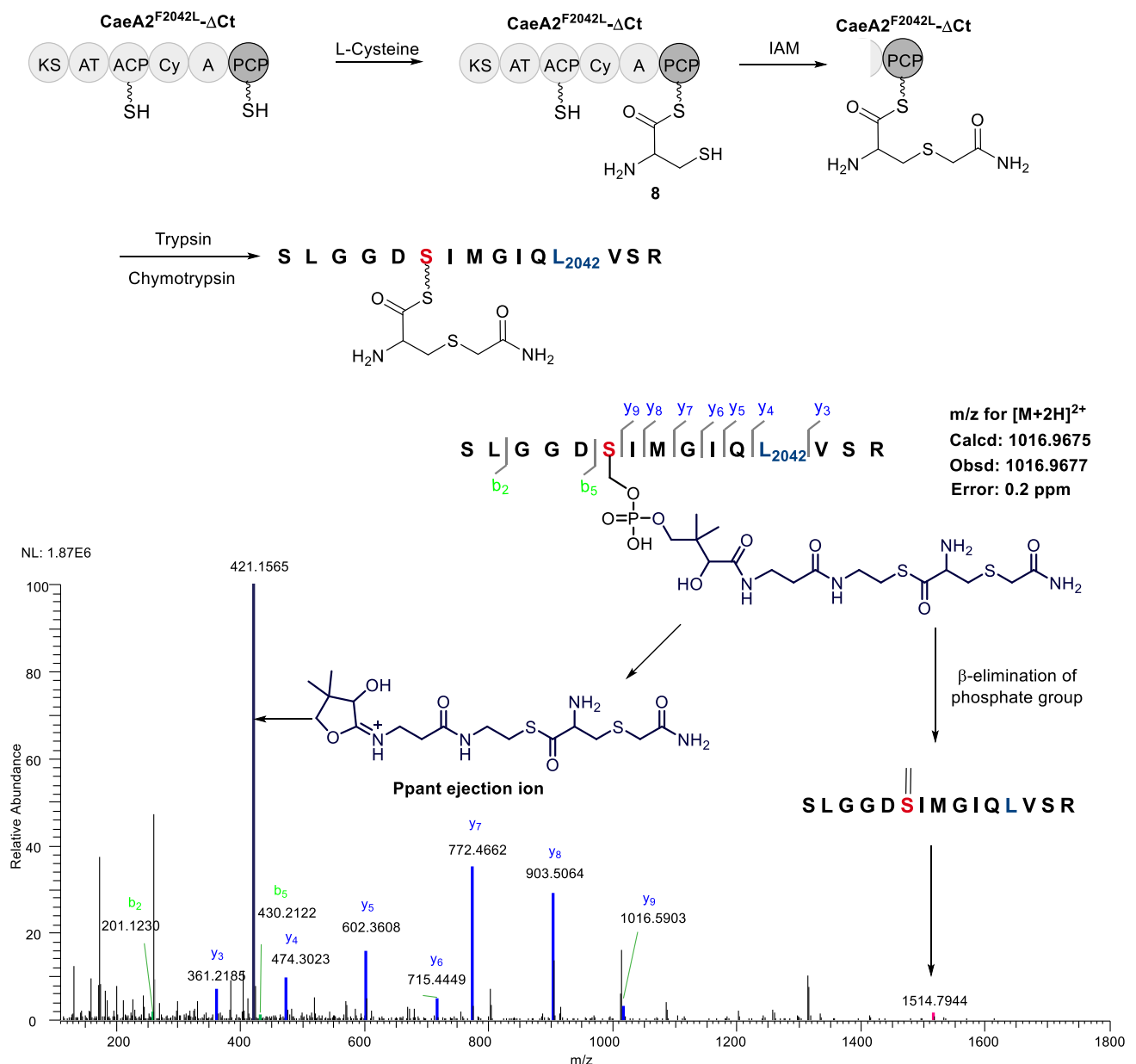
	1	10	20	30	40	50	60
CaeA3-C	...IEDVYE	LTPIQ	QGLLFEQLANGDAGVYIEQ	LLTFDGL	TLHPPEEFQ	RAWQMVVD	RHPILRTSFHWRP
ColA3-C	...IADIYE	LSPIQ	QGLLYEQLAQPGGLGIYVEQ	LRLEFSG	TMHPEHFE	RAWQLVVD	RHPILRTSFHWRK
CaeA2-Ct	ITVEAAASD	LDPALAE	RALREIA	DHPALRLRSGRD
ColA2-Ct	FDPVPA	ARALNALA	DHPALRLTRVREE
TycA-E	...IAGNV	LTPIQ	K...WFFGKNFTINTGHWNQSS	VLYRPE	GFDPKVI	QSVMDKI	IEHHDALRMVYQHEN
BacB-C1	...KSGTY	LSRE	QRMFILNQDDSKTAYNMP	LAVKING	EVQISRL	EQAWKALI	KRHESLRTSFVMLD
MycC-C	...IKGPAL	LTPIQ	H...WFFDQRYPDLLHHYNQAV	MLYWE	GLNVPM	LREVMRKIV	EHHDALRMVYVPAK
CaeA2-Cy	QFPF	LNEMQAQWIGRLGSFDMGGVAPH	LYLEFDS	DTLDAAR	LQRAWQQV	VRHEMLRIVVLADG
ColA2-Cy	EPFP	LTMEQ	QAQWIGRLSSFDMGGVAPH	LYFEFDS	RTIETAR	LERAWQRVQ
EpoB-Cy	PAERHVP	FPF	LTPIQ	GSYWLGRGTGAFVTPSGI	HAIREY	DCADLD	VARLSRAFRKVVARHDMRLAHTLPDM
	70	80	90	100	110	120	130
CaeA3-C	EGLPLQVVH	..QHAELPL	EVLDWRH	IPVAEQKEK	ITEWLDR	ERVEGFD	VTQAPVMRLSVI
ColA3-C	NGSAVQVVH	..GAARLPL	ETLDWRD	LDERTQEER	LRVSLDA	ERAEGFD	LTDTVPLMRSTLI
CaeA2-Ct	DGTPGQYD	GGSPGDGL	FS	IDLSGLAEDARDDA	LAEIVAEM	SEEVGLA	EREPVKLAVVRAGADDVRL
ColA2-Ct	GQRFVAVRP	G..PGEDFDV	PEIDLAA	LPDAVRAAA	VEEMIGEM	MARELD	IETGPAVKFAVFR
TycA-E	GNVVQHNR	GLGGQLYDF	FSYNL	TAQPD	..VQQA	IEAETQRL	HSSMNLQEGPLVKVAL
BacB-C1	GEPVQKIE	..QEAEFR	LEYSEL	GDQSI	QEKISRF	IKPFLEKAPLLRAE
MycC-C	HGYEARNR	EIDEGDLF	SLEVFSL	..LEENNVA	QTIETLSNE	IQQS	IQLAEGPLIKLGL
CaeA2-Cy	RQRILD	NVPDYEF	EELDLRG	VDPEEARAL	LAEVSDRM	STEVRPADTWPLWEV
ColA2-Cy	RQQILH	DAEPYRF	EVLDLRT	TDPEESERQ	LAGIRD	RMATEVRPADVWP
EpoB-Cy	MQVIE	PKVDADI	EIIDLRG	LDRSTREAR	LVSLRDAM	SHRIYDTERPPLYHVVA
	140	150	160	170	180	190	
CaeA3-C	WRLSHLLMD	GW	SFGL	AVSDFI	DCYRAQCLGYEPERR	PMR	..PYRDYV
ColA3-C	WRFSLHLLMD	GW	SFTL	AIQDFI	DHYRVLCRGGRPTV	PPGR	..PYRDYLSWWRD
CaeA2-Ct	WLAAG	AALDD	DT	SWRL	LLAEFG	RAYQQLAERGSIDW
ColA2-Ct	CTAAQGLMD	DA	SVLL	LCRELI	QAYDRL	AAGRPFVW
TycA-E	LAIIHHLVVD	GISWRI	..LFED	LATGYA	QALAGQAI	SLPEKTD	SFQSWSQWLQ
BacB-C1	VDMMHH	IIISD	GV	SIGI	LMKEFAD	CCEGK	...ELS..PLAV
MycC-C	IVAHHLVID	GV	SWRI	..LIED	IAAA	YEQLLNGE	AIQLPKKTD
CaeA2-Cy	ISFDL	LCAD	VASFF	FYQILP	SWREF	HENP	...ELPSEPP
ColA2-Cy	ISFDL	LVAD	VVSFF	FYQILP	QWREF	YHHP	...DRDPEPL
EpoB-Cy	LSIDL	INVD	LG	SLSI	IFKDWL	SFYEDP	...ETSLFVLELS
	200	210	220	230	240	250	260
CaeA3-C	AGFT	PPEQ	LRMG	GTPPAPGEP	SHG	FVELSMA	ELEAGLRD
ColA3-C	ADYR	PVEQ	VHLGG	TEIPEGE	PTHAF	EQVLG	DLAPRLTALARA
CaeA2-Ct	EGAA	PPPS	VETVA	FTGEL	RTVL	DADVTTAFAEAT
ColA2-Ct	RR	KPAHP	AGLAD	TPGTPGEL	PRRR	TVELDAT	RTADLFTAAAGSHH
TycA-E	SQAKN	VSLPK	DYEV	TDCKQK	SVRNMR	IRLHPE	ETEQLK
BacB-C1	RGDI	PVLN	MPLD	FPRPK	IRSFQGN	RTVVELD	QDTTKK
MycC-C	EHIPL	PKLP	KDNE	QEIGLA	EDRETI	IVQWTA	EETERL
CaeA2-Cy	AQLP	PAPE	LPTV	VAPGEL	VTPEF	VRRHGS	LDPRVWAG
ColA2-Cy	RELPA	APPE	LPTV	VQAGG	GERLGF	VRRHAR	LDAPWGR
EpoB-Cy	AELP	PPPM	LPMK	ADPSTL	KEIR	FRHTEQ	WLPSDSWSR

Supplementary Figure 9. Characterization of L-CysteinyI-S-CaeA2^{F2042L}ΔCt (8) by nanoLC-MS/MS

following treatment with IAM and subsequent complete digestion with trypsin and chymotrypsin.

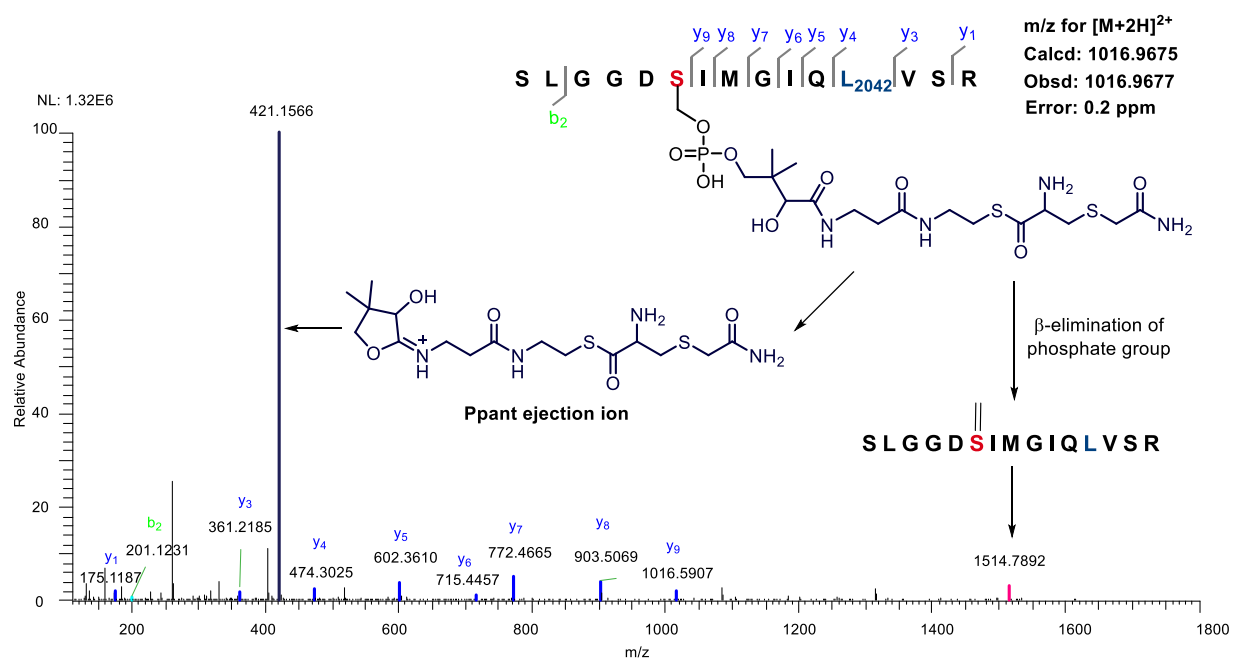
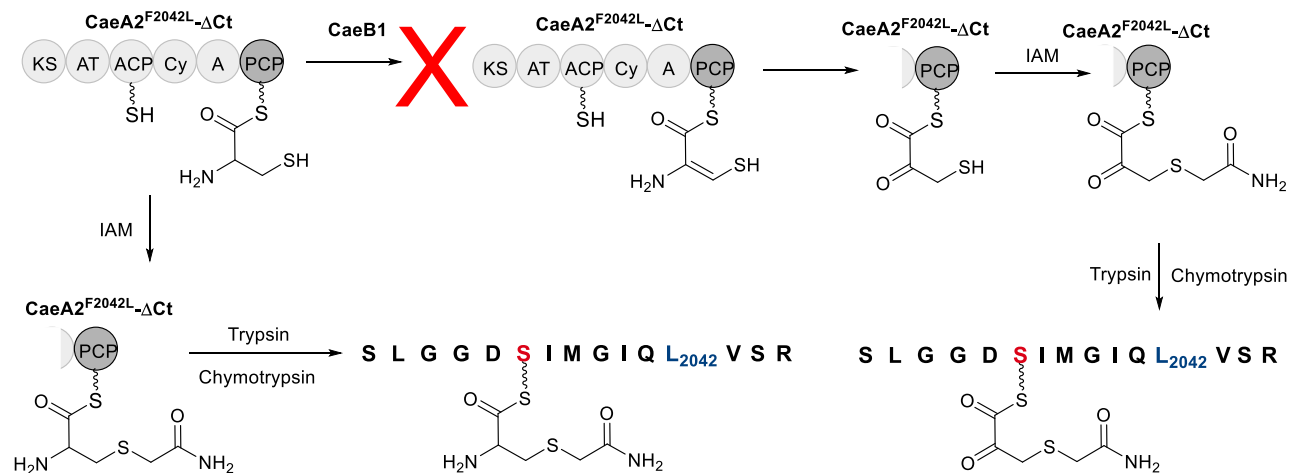
(a) Incubation of truncated CaeA2^{F2042L}ΔCt and L-cysteine.

a

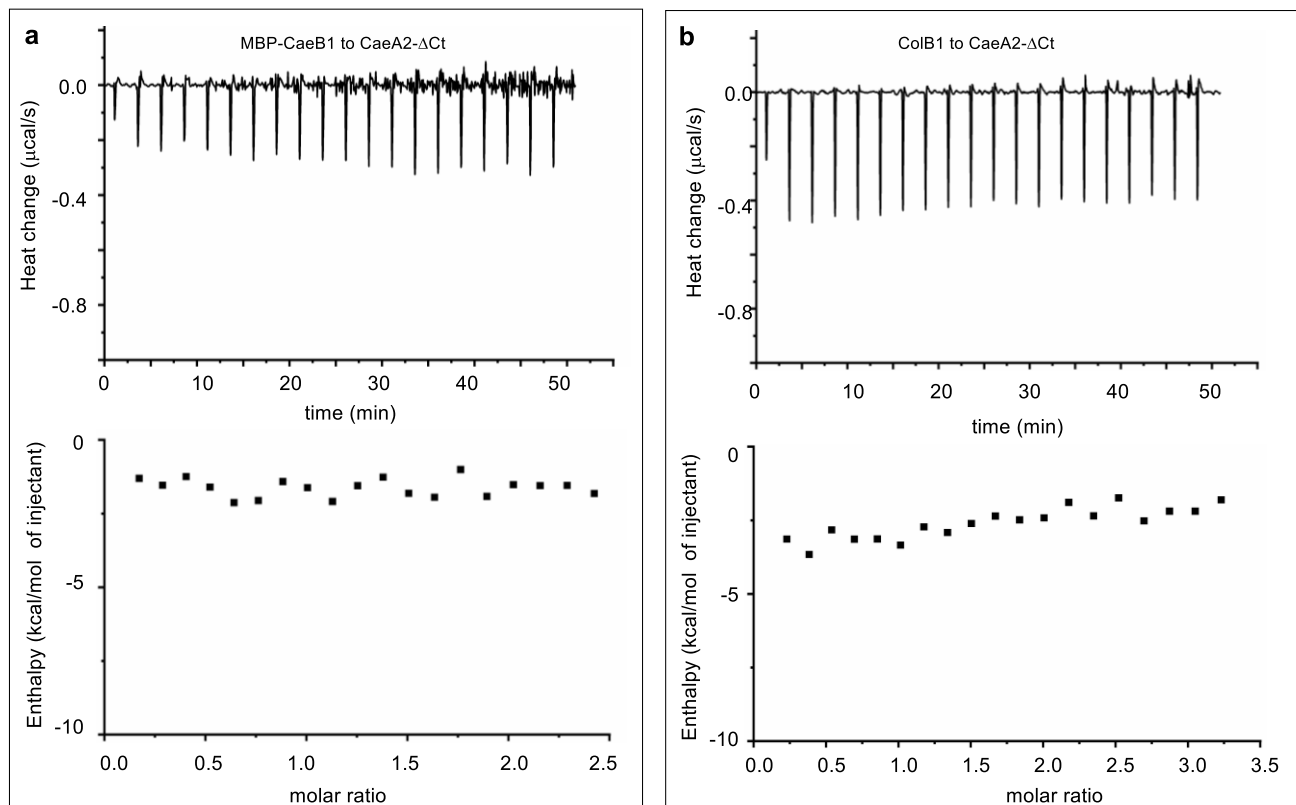


(b) Incubation of CaeB1 with truncated CaeA2^{F2042L}ΔCt and L-cysteine.

b



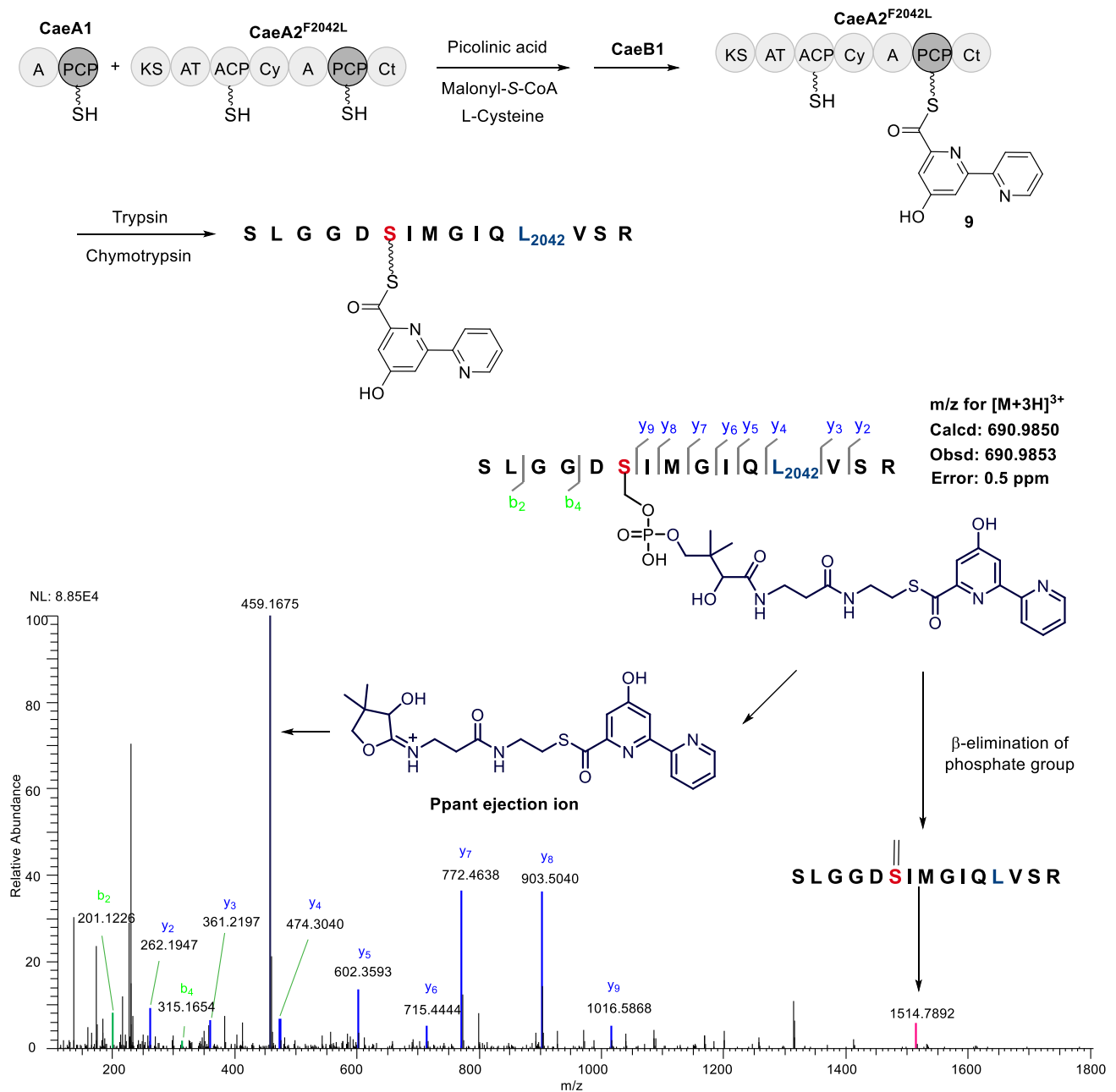
Supplementary Figure 10. Measurement of the interactions of truncated CaeA2- Δ Ct with related flavoproteins by ITC. Raw dates were shown on top, and the integrated curves containing experimental points and the best fitting line obtained from the single binding site model were shown on bottom. **(a)** Titrating MBP-fused CaeB1 to CaeA2- Δ Ct. **(b)** Titrating ColB1 to CaeA2- Δ Ct.



Supplementary Figure 11. Characterization of 2,2'-bipyridinyl-S-CaeA2^{F2042L} (**9**) by nanoLC-MS/MS following complete digestion with trypsin and chymotrypsin.

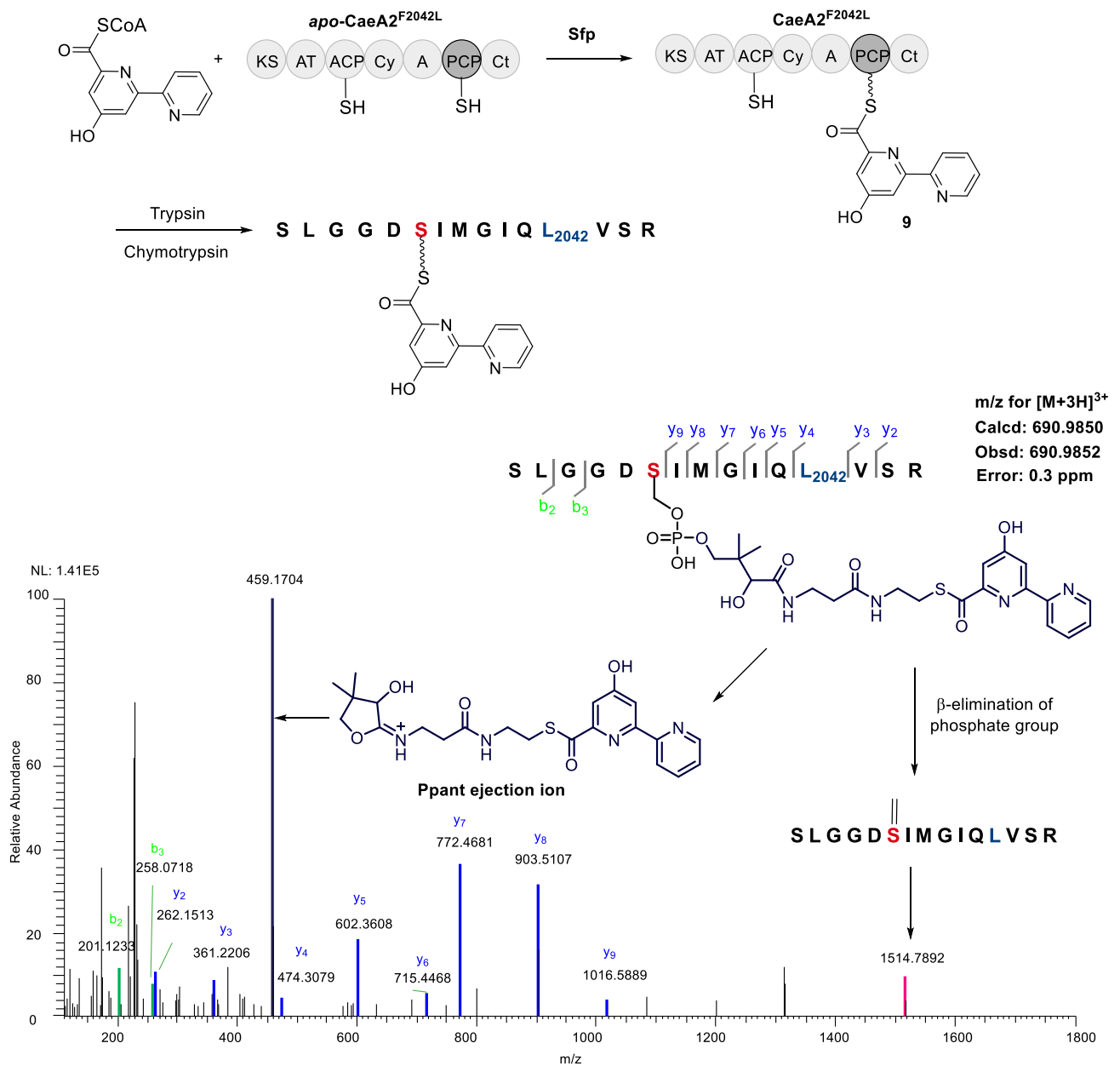
(a) Incubation of CaeA1, CaeA2^{F2042L} and CaeB1 with picolinic acid, malonyl-S-CoA and L-cysteine.

a

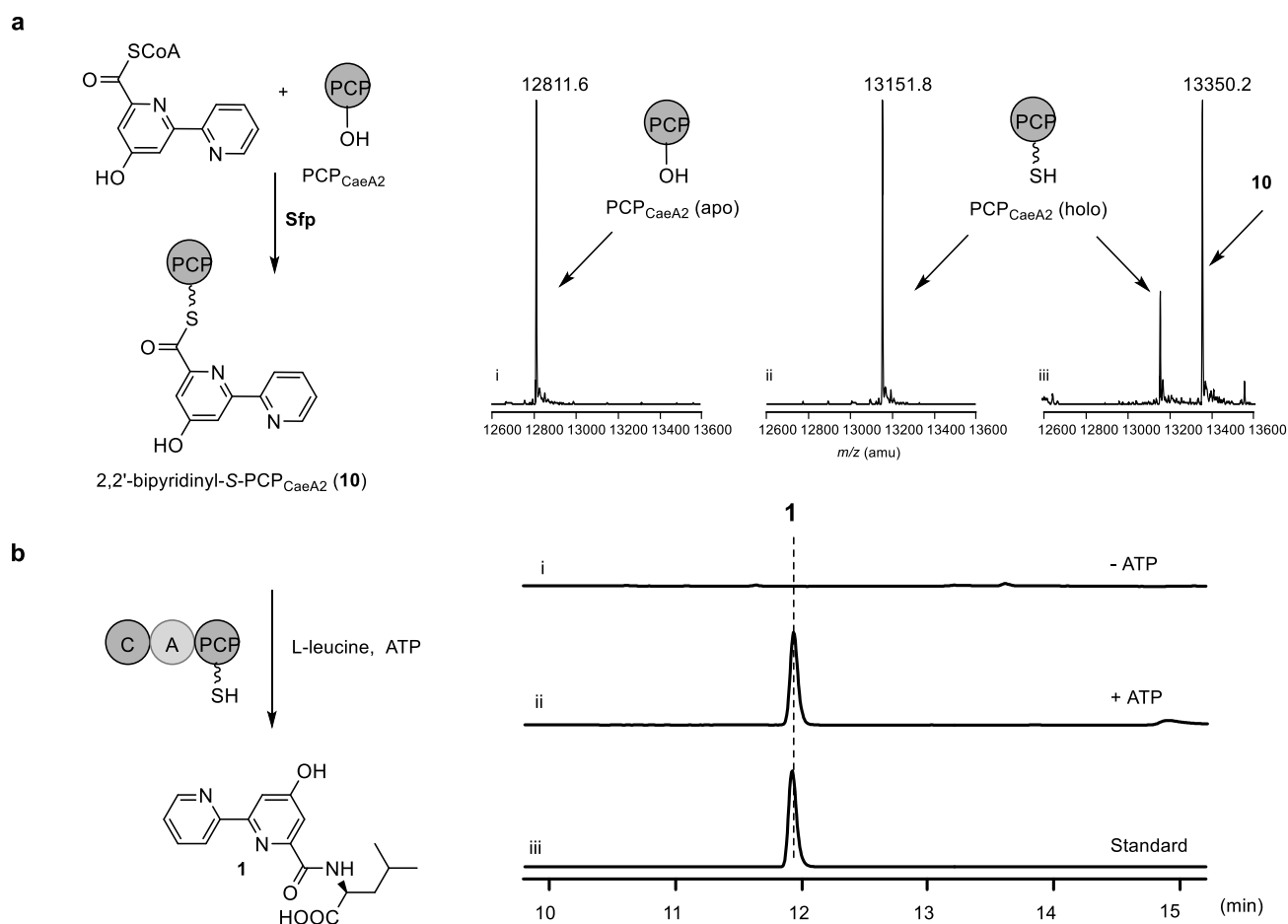


(b) Incubation of Ppant-unmodified CaeA2^{F2042L} (produced in *E. coli* BL21(DE3)) with synthesized 2,2'-bipyridinyl-*S*-CoA in the presence of Sfp.

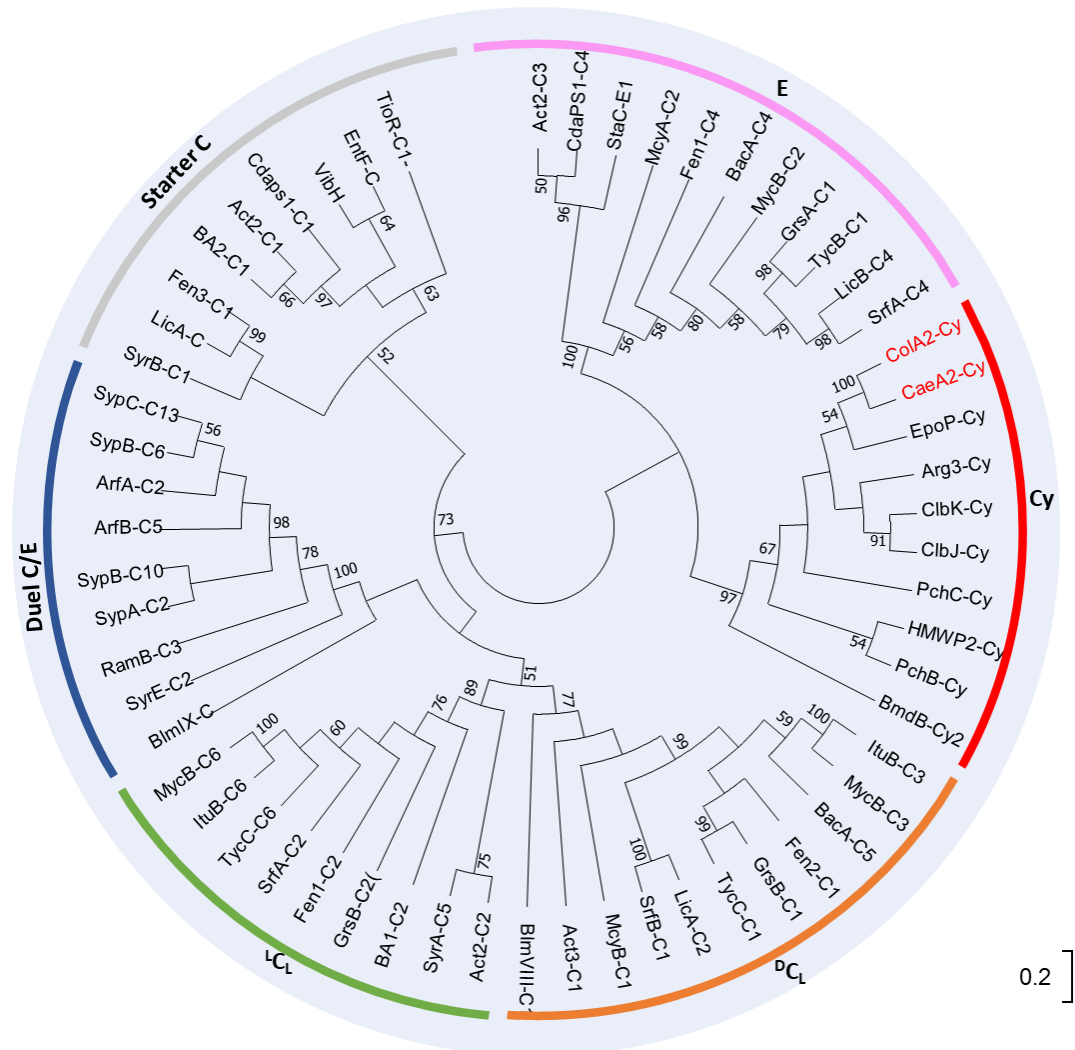
b



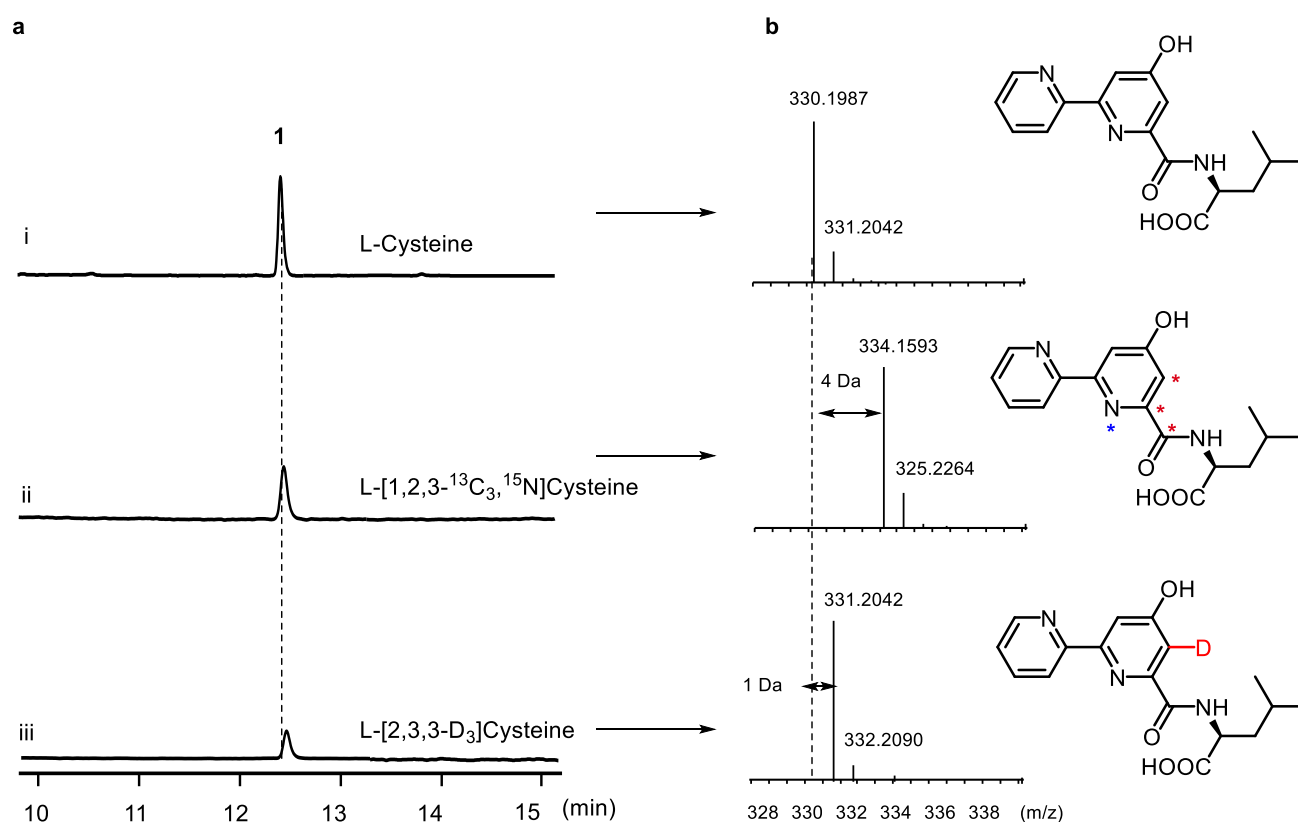
Supplementary Figure 12. Validation of 2,2'-bipyridinyl-*S*-Ppant as an intermediate in the production of **1**. **(a)** Preparation of 2,2'-bipyridinyl-*S*-PCP_{CaeA2} (**10**) by the incubation of PCP_{CaeA2} in apo form (produced in *E. coli* BL21(DE3)) with synthesized 2,2'-bipyridinyl-*S*-CoA in the presence of Sfp. PCP_{CaeA2} in both apo (i) and holo (ii) forms and 2,2'-bipyridinyl-*S*-PCP_{CaeA2} (iii) were examined by HR-MS. **(b)** Incubation of prepared 2,2'-bipyridinyl-*S*-PCP_{CaeA2} with CaeA3 and L-leucine in the absence (i) and presence (ii) of ATP, the standard of **1** (iii).



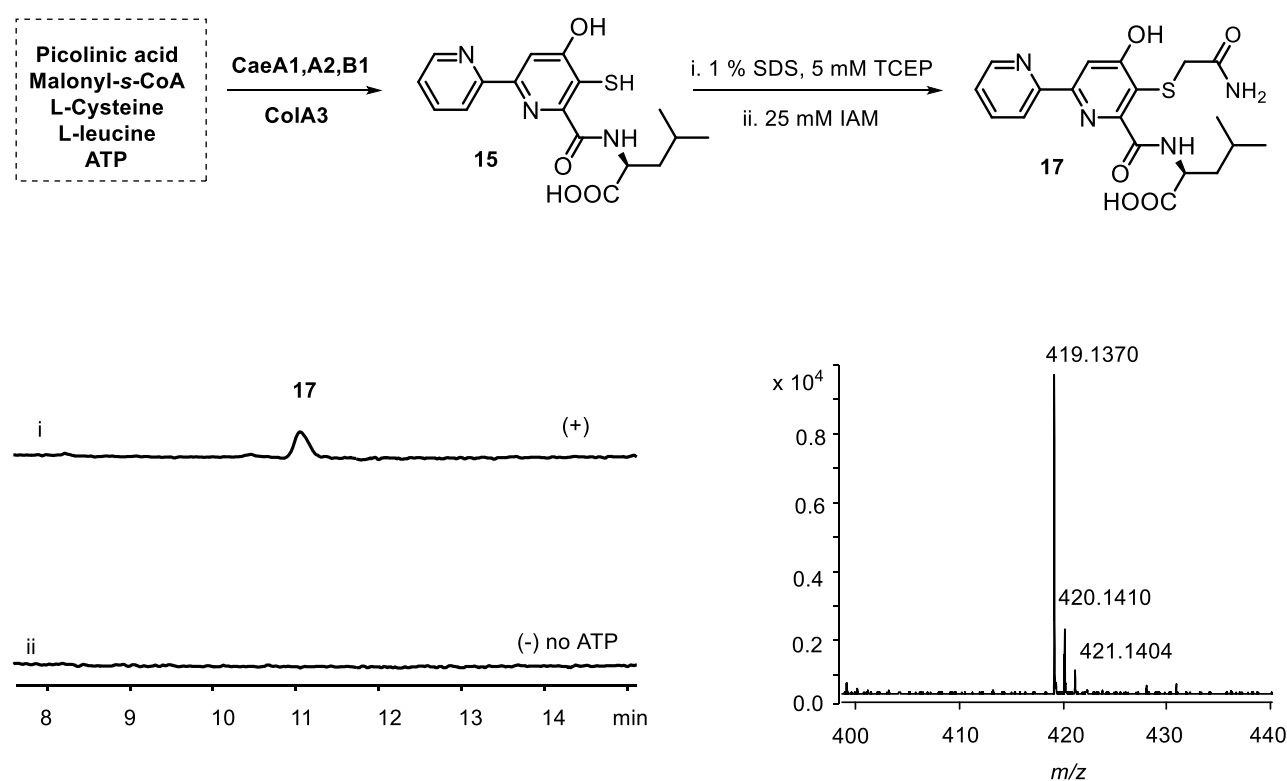
Supplementary Figure 14. Phylogenetic analysis of the Cy domains of the NRPSs CaeA2 and ColA2. The evolutionary distances were computed using the p-distance method. The support for grouping the C domain subtypes is indicated by bootstrap value. The Cy domains of CaeA2 and ColA2 are shown in red. Members of the C domain subtypes (^LCL, ^DCL, Starter C, Epimerization (E), Dual E/C and Heterocyclization (Cy) domains) include those of the NRPSs Act2 and Act3 in actinomycin biosynthesis²⁷; ArfA and ArfB in arthrofactin biosynthesis²⁸, Arg3 in argyrins biosynthesis²⁹; BA1 and BA2 in bacitracin biosynthesis¹³; BmdB in bacillamide biosynthesis³⁰; BlmVIII and BlmIV in bleomycin biosynthesis³¹; ClbJ and ClbK in colibactin biosynthesis³²; CdaPS1 in calcium-dependent antibiotic biosynthesis¹⁷; EpoP in epothilone biosynthesis¹⁵; EntF in enterobactin biosynthesis³³; Fen1, Fen2 and Fen3 in fengycin biosynthesis¹⁹; GrsA and GrsB in gramicidin biosynthesis²⁰; HMWP2 in yersiniabactin biosynthesis³⁴; ItuB in iturin biosynthesis²¹; LicA and LicB in lichenicin biosynthesis²²; McyA and McyB in microcystin biosynthesis³⁵; MycB in mycosubtilin biosynthesis²³; PchB and PchC in pyochelin biosynthesis³⁶; RamB in ramoplanin biosynthesis³⁷; SrfA and SrfB in surfactin biosynthesis³⁸; SypA, SypB and SypC in syringopeptin biosynthesis³⁹; StaC in A47934 biosynthesis⁴⁰; SyrA, SyrB and SyrE in syringomycin biosynthesis⁴¹; TycB and TycC in tyrocidin biosynthesis¹²; TioR in thiocoraline biosynthesis⁴²; VibH in vibriobactin biosynthesis⁴³. The sequences were downloaded from the database of NaPDoS, in which the biochemical function of these related C domains have been characterized^{25,26}.



Supplementary Figure 15. Comparison in the production of the CAE 2,2'-bipyridine intermediate **1** using the substrate L-cysteine (i), L-[1,2,3- $^{13}\text{C}_3$, ^{15}N]cysteine (ii) and L-[2,3,3- D_3]cysteine (iii), respectively. **(a)** HPLC analysis. **(b)** HR-MS analysis.



Supplementary Figure 16. Examination of thiol intermediate 5-sulfhydryl-2,2'-bipyridinyl-L-leucine (**15**) by IAM derivatization. **15** was produced in the reactions where CaeA1, CaeA2, CaeB1 and ColA3 were combined with picolinic acid, malonyl-S-CoA, L-cysteine and L-leucine in the presence (i) or absence (ii) of ATP. The reaction mixtures were treated with IAM to produce **17**, which were then analyzed by HPLC (left, $\lambda = 315$ nm) and HR-MS (right, $[M + H]^+$ m/z : calcd. 419.1389 for $C_{19}H_{22}O_5N_4S$, obs. 419.1370).



SUPPLEMENTARY TABLES

Table S1. Bacterial strains in this study.

Strain/plasmid	Characteristics	Source
<i>E. coli</i>		
DH5 α	Host for general cloning	Invitrogen
S17-1	Donor strain for conjugation between <i>E. coli</i> and <i>Streptomyces</i>	2
BL21 (DE3)	Host for protein expression	NEB
BAP1	Host for protein expression, contains <i>sfp</i> gene under the control of the T7 RNA polymerase promoter	44
<i>S. raseosporus</i>		
NRRL 11379	COL-producing wild type strain	NRRL
QL2003	<i>colA2</i> inframe deletion mutant, COL non-producing	7
QL2006	QL2003 derivative, in which chimeric gene <i>col/caeA2</i> was expressed <i>in trans</i>	This study
<i>A. cyanogriseus</i>		
NRRL B-2194	CAE-producing wild type strain	NRRL

Table S2. Plasmids in this study.

Plasmids	Characteristics	Source
pMD19-T	<i>E. coli</i> subcloning vector, ampicillin resistance	Takara
pSET152	<i>E. coli-Streptomyces</i> shuttle vector for gene complementation, apramycin resistance	2
pET-28a(+)	Protein expression vector used in <i>E.coli</i> , encoding N-terminal 6 × His-tag, kanamycin resistance	Novagen
pET-37b(+)	Protein expression vector used in <i>E.coli</i> , encoding C-terminal 8 × His-tag, kanamycin resistance	Novagen
pSJ5	Protein expression vector used in <i>E.coli</i> , encoding N-terminal His ₆ -Trx-tag, ampicillin resistance	Transformed from pET32a, adding Trx-tag and after N-terminal His
pQ8	Protein expression vector used in <i>E.coli</i> , encoding N-terminal His ₆ -MBP-His ₆ -tag, kanamycin resistance	⁴⁵
pQL1022	pSET152 derivative containing the 8.0 kb <i>ermE*</i> promoter + <i>colA2</i> gene	⁷
pQL1026	pMD19-T derivative, containing <i>caeA1</i>	This study
pQL1027	pET-28a(+) derivative, containing <i>caeA1</i>	This study
pQL1028	pMD19-T derivative, containing <i>caeA2</i>	This study
pQL1029	pET-37b(+) derivative, containing <i>caeA2</i>	This study
pQL1030	pMD19-T derivative, containing <i>caeA3</i>	This study
PQL1031	pET-28a(+) derivative, containing <i>caeA3</i>	This study
pQL1032	pMD19-T derivative, containing <i>caeB1</i>	This study
pQL1033	pET-28a(+) derivative, containing <i>caeB1</i>	This study
pQL1034	pMD19-T derivative, containing <i>colB1</i>	This study
PQL1035	pET-28a(+) derivative, containing <i>colB1</i>	This study
pQL1036	pMD19-T derivative, containing <i>colA3</i>	This study

pQL1037	pET-28a(+) derivative, containing <i>colA3</i>	This study
pQL1038	pMD19-T derivative, containing <i>caeA4</i>	This study
pQL1039	pET-28a(+) derivative, containing <i>ceaA4</i>	This study
pQL1040	pMD19-T derivative, containing <i>caeA2-ΔCt</i>	This study
PQL1041	pET-28a(+) derivative, containing <i>caeA2-ΔCt</i>	This study
pQL1042	pMD19-T derivative, containing <i>colG2</i>	This study
pQL1043	pET-28a(+) derivative, containing <i>colG2</i>	This study
pQL1044	pMD19-T derivative, containing the gene encoding the PCP domain of CaeA2	This study
pQL1045	pET-28a(+) derivative, containing the gene encoding the PCP domain of CaeA2	This study
pQL1046	pMD19-T derivative, the gene encoding the C domain of CaeA3	This study
pQL1047	pMD19-T derivative, containing the gene encoding the A-PCP didomain of ColA3	This study
pQL1048	pET-28a(+) derivative, containing the gene encoding chimeric NRPS Cae/ColA3	This study
pQL1049	pET-37b(+) derivative, containing the gene encoding CaeA2-F2042L	This study
pQL1050	pET-37b(+) derivative, containing the gene encoding CaeA2-F2042I	This study
pQL1051	pET-37b(+) derivative, containing the gene encoding CaeA2-F2042V	This study
pQL1052	pET-37b(+) derivative, containing the gene encoding CaeA2-F2042L- Δ Ct	
pQL1053	pMD19-T derivative, containing the gene encoding the NRPS module of CaeA2	This study

pQL1054	pQL1022 derivative, containing the chimeric gene <i>col/caeA2</i> coding for the hybrid protein that harbors the PKS module from ColA2 and the NRPS module from CaeA2	This study
pQL1055	pSJ5 derivative, containing the gene encoding the Ct domain of CaeA2	This study
pQL1055	pQ8 derivative, containing the gene encoding CeaB1	This study

Table S3. Primers used in this study. The sequences of restriction enzymes are underlined.

Primer	Sequence
CaeA1-For	5'-GCGCGAATTCC <u>CATATG</u> GGAAGAACTGCGAAGGATT-3' (NdeI)
CaeA1-Rev	5'-ATATA <u>AAGCTT</u> ACTCGAGGGTTCGGGAAGCGTCC-3' (HindIII)
CaeA2-For	5'-AGCTGAATTCC <u>CATATG</u> ACCGGTCATGACACCGT-3' (NdeI)
CaeA2-Rev	5'-AGCTAAGCTT <u>ACTCGAGCTT</u> CTCAAGGGTGGACAGC-3' (XhoI)
CaeA3-For	5'-GCGCGAATTCC <u>CATATG</u> AGCCAGTTGAAAAAGGGCATCG-3' (NdeI)
CaeA3-Rev	5'-ATATA <u>AAGCTT</u> ACTCGAGCGCCCGCTCCCTCTCCAC-3' (HindIII)
CaeA4-For	5'-GCGCGAATTCC <u>CATATG</u> TTCAGGTCGTGGTTCGGCGG-3' (NdeI)
CaeA4-Rev	5'-ATATA <u>AAGCTT</u> TTCAGGGGGACGACCGCTGCGGCG-3' (HindIII)
CaeB1-For	5'-ATGCGAATTCC <u>CATATG</u> AGCGCACCGACGGTGACCAC-3' (NdeI)
CaeB1-Rev	5'-ATGCA <u>AAGCTT</u> ACTCGAGCAGCGACCCCGCGATGCTCC-3' (HindIII)
ColB1-For	5'-GCGCGAATTCC <u>CATATG</u> AATACGAACCCCGGACAGAGC-3' (NdeI)
ColB1-Rev	5'-ATATA <u>AAGCTT</u> ACTCGAGGTCACCGGGGCTCCCT-3' (HindIII)
ColA3-For	5'-TTGAATTCC <u>CATATG</u> GCGACAACGGCGGACGGGACT-3' (NdeI)

ColA3-Rev	5'-TTAAGCTT <u>ACTCGAGCCGTGCCCCCTCCCCCGCGACC</u> -3' (HindIII)
CaeA2-ΔCt-For	5'-TTGAATTCC <u>CATATG</u> ACCGGTCATGACACCGTCGAGTC-3' (NdeI)
CaeA2-ΔCt-Rev	5'-TTAAGCTT <u>ACTCGAG</u> GGCCGCCTGGTGCGGGGTGAAC-3' (XhoI)
ColG2-For	5'-TTGAATTCC <u>CATATG</u> AGCGGCGCGATATCGACCGAG-3' (NdeI)
ColG2-Rev	5'-TTAAGCTTCTACCCCGGTACGACGGTCTCGA-3' (HindIII)
CaeA2-PCP-For	5'-TTGAATTCC <u>CATATG</u> CCCCGCGCCGGACCGCGTCGG-3' (NdeI)
CaeA2-PCP-Rev	5'-TTAAGCTT <u>ACTCGAGCGCGACCGTCCCGTCCGCCTC</u> -3' (HindIII)
CaeA3-C-For	5'- TTGAATTCC <u>CATATG</u> AAAAAGGGCATCGAGGATGTCTACGAACTC AC-3' (NdeI)
CaeA3-C-Rev	5'-TTGCTAGCGTCGGGGTTGGCGGCGGCCAGGTC-3' (NheI)
ColA3-APCP-For	5'-TTGCTAGCGCCGTGCGCTGCGGAGGCCG-3' (NheI)
ColA3-APCP-Rev	5'-TTAAGCTT <u>ACTCGAGCCGTGCCCCCTCCCCCGCGAC</u> -3' (HindIII)
CaeA2-F2042L-For	5'-TGGGTATCCAGTTGGTCAGCCGTGCCG-3'
CaeA2-F2042L- Rev	5'-GGCTGACCAACTGGATAACCATGATCGA-3'
CaeA2-F2042I- For	5'-TGGGTATCCAGATCGTCAGCCGTGCCG-3'
CaeA2-F2042I-Rev	5'-GGCTGACGATCTGGATAACCATGATCGA-3'
CaeA2-F2042V-For	5'-TGGGTATCCAGGTCGTCAGCCGTGCCG-3'
CaeA2-F2042V- Rev	5'-GGCTGACGACCTGGATAACCATGATCGA-3'
Cae-CO-for	5'-TTAGATCTCAGCCGTTCCCGCTGAACGAGAT-3' (BglII)
Cae-Ct-rev	5'-TTGATATCTTACTTCTCAAGGGTGGACAGCAGGGC-3' (EcoRV)

Table S4. MS/MS data for PCP-aminoacylation on CaeA2.

Detail MS/MS data for **Figure S4A**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1220	9 ppm
y3	361.2198	361.2183	4 ppm
y4	508.2882	508.2856	5 ppm
y5	636.3467	636.3420	7 ppm
y7	806.4521	806.4467	6 ppm
y8	937.4925	937.4888	4 ppm
y9	1050.576	1050.5739	2 ppm
Ppant ejection ion	261.1267	261.1252	5 ppm

Detail MS/MS data for **Figure S4B**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1220	9 ppm
y3	361.2198	361.2178	5 ppm
y4	474.3038	474.3040	0.5 ppm
y5	602.3623	602.3628	1 ppm
y6	715.4463	715.4470	1 ppm
y7	772.4677	772.4686	1 ppm
y8	903.5081	903.5049	3 ppm
y9	1016.592	1016.5834	8 ppm
Ppant ejection ion	261.1267	261.1269	1 ppm

Detail MS/MS data for **Figure S5A**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1233	3 ppm
y2	262.1514	262.1507	3 ppm
y4	474.3038	474.3027	2 ppm
y5	602.3623	602.3621	3 ppm
y6	715.4463	715.4460	4 ppm
y7	772.4677	772.4669	1 ppm
y8	903.5081	903.5062	2 ppm
y9	1016.592	1016.5920	0 ppm
Ppant ejection ion	364.1359	364.1356	1 ppm

Detail MS/MS data for **Figure S5B**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1225	7 ppm
b4	430.1936	430.1879	13 ppm
y3	361.2198	361.2177	6 ppm
y4	474.3038	474.3012	5 ppm
y5	602.3623	602.3595	5 ppm
y6	715.4463	715.4433	4 ppm
y7	772.4677	772.4642	4 ppm
y8	903.5081	903.5040	4 ppm
y9	1016.592	1016.5881	4 ppm
Ppant ejection ion	421.1574	421.1555	4 ppm

Detail MS/MS data for **Figure S6A, i**

	Calcd m/z	Obsd m/z	Error
y3	361.2198	361.2189	2 ppm
y4	474.3038	474.3031	1 ppm
y5	602.3623	602.3622	0.2 ppm
y6	715.4463	715.4452	1 ppm
y7	772.4677	772.4666	1 ppm
y8	903.5081	903.5055	3 ppm
y9	1016.592	1016.5910	1 ppm
Ppant ejection ion	421.1574	421.1570	1 ppm

Detail MS/MS data for **Figure S6A, ii**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1225	7 ppm
b3	258.1453	258.1435	7 ppm
y3	361.2198	361.2168	8 ppm
y4	474.3038	474.3021	3 ppm
y5	602.3623	602.3593	5 ppm
y6	715.4463	715.4423	5 ppm
y7	772.4677	772.4636	5 ppm
y8	903.5081	903.5048	4 ppm
y9	1016.592	1016.6003	8 ppm
Ppant ejection ion	420.1257	420.1234	5 ppm

Detail MS/MS data for **Figure S6B, i**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1227	6 ppm
b3	315.1667	315.1646	7 ppm
y3	361.2198	361.2180	5 ppm
y4	474.3038	474.3016	4 ppm
y5	602.3623	602.3599	4 ppm
y6	715.4463	715.4435	4 ppm
y7	772.4677	772.4647	4 ppm
y8	903.5081	903.5048	4 ppm
y9	1016.592	1016.5881	4 ppm
Ppant ejection ion	425.1646	425.1629	4 ppm

Detail MS/MS data for **Figure S6B, ii**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1226	6 ppm
b3	315.1667	315.1642	8 ppm
y3	361.2198	361.2181	5 ppm
y4	474.3038	474.3016	5 ppm
y5	602.3623	602.3597	4 ppm
y6	715.4463	715.4440	3 ppm
y7	772.4677	772.4651	3 ppm
y8	903.5081	903.5044	4 ppm
y9	1016.592	1016.5879	4 ppm
Ppant ejection ion	423.1360	423.1341	4 ppm

Detail MS/MS data for **Figure S6C, i**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1227	6 ppm
y3	361.2198	361.2181	5 ppm
y4	474.3038	474.3016	5 ppm
y5	602.3623	602.3602	3 ppm
y6	715.4463	715.4440	3 ppm
y7	772.4677	772.4651	3 ppm
y8	903.5081	903.5044	4 ppm
y9	1016.592	1016.5883	3 ppm
Ppant ejection ion	424.1762	424.1747	3 ppm

Detail MS/MS data for **Figure S6C, ii**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1227	6 ppm
b5	430.1936	430.2017	18 ppm
y3	361.2198	361.2181	5 ppm
y4	474.3038	474.3021	3 ppm
y5	602.3623	602.3597	4 ppm
y6	715.4463	715.4438	3 ppm
y7	772.4677	772.4651	3 ppm
y8	903.5081	903.5052	3 ppm
y9	1016.592	1016.5890	3 ppm
Ppant ejection ion	420.1257	420.1244	3 ppm

Detail MS/MS data for **Figure S8A**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1230	5 ppm
b5	430.1936	430.2122	40 ppm
y3	361.2198	361.2185	4 ppm
y4	474.3038	474.3023	3 ppm
y5	602.3623	602.3608	2 ppm
y6	715.4463	715.4449	2 ppm
y7	772.4677	772.5064	50 ppm
y8	903.5081	903.5052	3 ppm
y9	1016.592	1016.5903	2 ppm
Ppant ejection ion	421.1574	421.1565	2 ppm

Detail MS/MS data for **Figure S8B**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1231	4 ppm
y1	175.1194	175.1187	4 ppm
y3	361.2198	361.2185	3 ppm
y4	474.3038	474.3025	2 ppm
y5	602.3623	602.3610	2 ppm
y6	715.4463	715.4457	1 ppm
y7	772.4677	772.4665	1 ppm
y8	903.5081	903.5069	1 ppm
y9	1016.592	1016.5907	1 ppm
Ppant ejection ion	421.1574	421.1566	2 ppm

Detail MS/MS data for **Figure S9A**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1226	6 ppm
b4	315.1667	315.1654	4 ppm
y3	361.2198	361.2197	0.5 ppm
y4	474.3038	474.3040	0.5 ppm
y5	602.3623	602.3593	5 ppm
y6	715.4463	715.4444	3 ppm
y7	772.4677	772.4638	5 ppm
y8	903.5081	903.5040	5 ppm
y9	1016.592	1016.5868	5 ppm
Ppant ejection ion	459.1697	459.1675	5 ppm

Detail MS/MS data for **Figure S9B**

	Calcd m/z	Obsd m/z	Error
b2	201.1239	201.1233	3 ppm
Y2	262.1514	262.1513	0.5 ppm
y3	361.2198	361.2206	2 ppm
y4	474.3038	474.3079	9 ppm
y5	602.3623	602.3608	2 ppm
y6	715.4463	715.4468	1 ppm
y7	772.4677	772.4681	0.5 ppm
y8	903.5081	903.5107	3 ppm
y9	1016.592	1016.5889	3 ppm
Ppant ejection ion	459.1697	459.1704	2 ppm

SUPPLEMENTARY REFERENCES

- 1 Green, M. R. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2012).
- 2 Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Streptomyces Genetics*. (John Innes Foundation, 2000).
- 3 Li, D. *et al.* pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics* **21**, 3049-3050 (2005).
- 4 Chen, M., Pang, B., Du, Y. N., Zhang, Y. P. & Liu, W. Characterization of the metallo-dependent amidohydrolases responsible for "auxiliary" leuciny removal in the biosynthesis of 2,2'-bipyridine antibiotics. *Synth Syst Biotechnol* **2**, 137-146 (2017).
- 5 Wei, C. *et al.* Dual-Reactable Fluorescent Probes for Highly Selective and Sensitive Detection of Biological H₂S. *Chem. Asian J.* **11**, 1376-1381 (2016).
- 6 Bumpus, S. B. & Kelleher, N. L. Accessing natural product biosynthetic processes by mass spectrometry. *Curr. Opin. Chem. Biol.* **12**, 475-482 (2008).
- 7 Qu, X. *et al.* Caerulomycins and collismycins share a common paradigm for 2,2'-bipyridine biosynthesis via an unusual hybrid polyketide-peptide assembly Logic. *J. Am. Chem. Soc.* **134**, 9038-9041 (2012).
- 8 Watanabe, K., Khosla, C., Stroud, R. M. & Tsai, S.-C. Crystal Structure of an Acyl-ACP Dehydrogenase from the FK520 Polyketide Biosynthetic Pathway: Insights into Extender Unit Biosynthesis. *J. Mol. Biol.* **334**, 435-444 (2003).
- 9 Toogood, H. S. *et al.* Extensive domain motion and electron transfer in the human electron transferring flavoprotein.medium chain Acyl-CoA dehydrogenase complex. *J. Biol. Chem.* **279**, 32904-32912 (2004).
- 10 Battaile, K. P. *et al.* Crystal structure of rat short chain acyl-CoA dehydrogenase complexed with acetoacetyl-CoA: comparison with other acyl-CoA dehydrogenases. *J. Biol. Chem.* **277**, 12200-12207 (2002).
- 11 Kung, J. W., Seifert, J., von Bergen, M. & Boll, M. Cyclohexanecarboxyl-coenzyme A (CoA) and cyclohex-1-ene-1-carboxyl-CoA dehydrogenases, two enzymes involved in the fermentation of benzoate and crotonate in *Syntrophus aciditrophicus*. *J. Bacteriol.* **195**, 3193-3200 (2013).
- 12 Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843-6850 (1997).
- 13 Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* **4**, 927-937 (1997).
- 14 Duitman, E. H. *et al.* The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13294-13299 (1999).
- 15 Molnár, I. *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97-109 (2000).
- 16 Gaudelli, N. M., Long, D. H. & Townsend, C. A. beta-Lactam formation by a non-ribosomal

- peptide synthetase during antibiotic biosynthesis. *Nature* **520**, 383-387 (2015).
- 17 Hojati, Z. *et al.* Structure, Biosynthetic Origin, and Engineered Biosynthesis of Calcium-Dependent Antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**, 1175-1187 (2002).
- 18 Chiu, H. T. *et al.* Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8548-8553 (2001).
- 19 Steller, S. *et al.* Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chem. Biol.* **6**, 31-41 (1999).
- 20 Turgay, K., Krause, M. & Marahiel, M. A. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6**, 529-546 (1992).
- 21 Tsuge, K., Akiyama, T. & Shoda, M. Cloning, sequencing, and characterization of the iturin A operon. *J. Bacteriol.* **183**, 6265-6273 (2001).
- 22 Konz, D., Doekel, S. & Marahiel, M. A. Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J. Bacteriol.* **181**, 133-140 (1999).
- 23 Bamas-Jacques, N., Lorenzon, S., Lacroix, P., de Swetschin, C. & Crouzet, J. Cluster organization of the genes of *Streptomyces pristinaespiralis* involved in pristinamycin biosynthesis and resistance elucidated by pulsed-field gel electrophoresis. *J. Appl. Microbiol.* **87**, 939-948 (1999).
- 24 Mast, Y. *et al.* Characterization of the 'pristinamycin supercluster' of *Streptomyces pristinaespiralis*. *Microb Biotechnol* **4**, 192-206 (2011).
- 25 Ziemert, N. *et al.* The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, e34064 (2012).
- 26 Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78-92 (2007).
- 27 Schauwecker, F., Pfennig, F., Schroder, W. & Keller, U. Molecular cloning of the actinomycin synthetase gene cluster from *Streptomyces chrysomallus* and functional heterologous expression of the gene encoding actinomycin synthetase II. *J. Bacteriol.* **180**, 2468-2474 (1998).
- 28 Roongsawang, N. *et al.* Cloning and characterization of the gene cluster encoding arthrofacin synthetase from *Pseudomonas* sp MIS38. *Chem. Biol.* **10**, 869-880 (2003).
- 29 Pogorevc, D. *et al.* Biosynthesis and Heterologous Production of Argyrins. *Acs Synth Biol* **8**, 1121-1133 (2019).
- 30 Yuwen, L. *et al.* The role of aromatic L-amino acid decarboxylase in bacillamide C biosynthesis by *Bacillus atrophaeus* C89. *Sci. Rep.* **3**, 1753-1762 (2013).
- 31 Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7**, 623-642 (2000).
- 32 Homburg, S., Oswald, E., Hacker, J. & Dobrindt, U. Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*. *FEMS Microbiol. Lett.* **275**, 255-262 (2007).
- 33 Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648-2659 (1998).

- 34 Gehring, A. M. *et al.* Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem. Biol.* **5**, 573-586 (1998).
- 35 Nishizawa, T. *et al.* Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin. *J. Biochem.* **127**, 779-789 (2000).
- 36 Nowak-Thompson, B., Chaney, N., Wing, J. S., Gould, S. J. & Loper, J. E. Characterization of the pyoluteorin biosynthetic gene cluster of *Pseudomonas fluorescens* Pf-5. *J. Bacteriol.* **181**, 2166-2174 (1999).
- 37 Chen, J.-S. *et al.* Functional identification of the gene encoding the enzyme involved in mannosylation in ramoplanin biosynthesis in *Actinoplanes* sp. *Biotechnol. Lett* **35**, 1501-1508 (2013).
- 38 Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301-310 (2002).
- 39 Scholz-Schroeder, B. K., Soule, J. D. & Gross, D. C. The *sypA*, *sypB* and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptin by *Pseudomonas syringae* pv. *syringae* B301D. *Mol. Plant. Microbe Interact.* **16**, 271-280 (2003).
- 40 Pootoolal, J. *et al.* Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL 15009. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8962-8967 (2002).
- 41 Guenzi, E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. Characterization of the syringomycin synthetase gene cluster - A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**, 32857-32863 (1998).
- 42 Lombo, F. *et al.* Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *Streptomyces* species. *Chembiochem* **7**, 366-376 (2006).
- 43 Wyckoff, E. E., Stoeckner, J. A., Reed, K. E. & Payne, S. M. Cloning of a *Vibrio cholerae* vibriobactin gene cluster: Identification of genes required for early steps in siderophore biosynthesis. *J. Bacteriol.* **179**, 7055-7062 (1997).
- 44 Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790-1792 (2001).
- 45 Fang, J. *et al.* Cloning and characterization of the tetrocarcin A gene cluster from *Micromonospora chalcone* NRRL 11289 reveals a highly conserved strategy for tetronate biosynthesis in spirotetronate antibiotics. *J. Bacteriol.* **190**, 6014-6025 (2008).