LanD-like Flavoprotein-Catalyzed Aminovinyl-Cysteine Formation through Oxidative Decarboxylation and Cyclization of a Peptide at the C-Terminus

Jingyu Liu[1,#], Yanping Qiu[1,#], Tao Fu[1,#], Miao Li[1], Yuqing Li[1,2], Qian Yang[1], Zhijun Tang[1], Haoyu Tang[1], Guangyu Li[4], Lifeng Pan[1,*] and Wen Liu[1,2.3,*]

[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] Department of Chemistry, Shanghai Normal University, 100 Guiling Road, Shanghai 200032, China;

[3] Huzhou Center of Bio-Synthetic Innovation, 1366 Hongfeng Road, Huzhou 313000, China.

[4] Nuclear Magnetic Resonance Facility, Shanghai Institute of Organic Chemistry, University of Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, China

* To whom correspondence should be addressed: Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Rd., Shanghai 200032, China. Lifeng Pan, Email: panlf@sioc.ac.cn, Tel: 86-21-54925561; Wen Liu, Email: wliu@sioc.ac.cn, Tel: 86-21-54925111, Fax: 86-21-64166128.

[#] These authors contributed to this work equally.

ABSTRACT:

Aminovinyl-cysteine residues arise from processing the C-terminal L-Cys and an internal L-Ser/L-Thr or L-Cys of a peptide. Formation of these nonproteinogenic amino acids, which occur in a macrocyclic ring of diverse ribosomally synthesized lanthipeptides and non-lanthipeptides, remains poorly understood. Here, we report that LanD-like flavoproteins in the biosynthesis of distinct non-lanthipeptides share an unexpected dual activity for aminovinyl-cysteine formation. Each flavoprotein catalyzes oxidative decarboxylation of the C-terminal L-Cys and couples the resulting enethiol nucleophile with the internal residue to afford a thioether linkage for peptide cyclization. The cyclization step, which largely depends on proximity effect by positioning the enethiol intermediate with a bent conformation at the active site, can be substrate-dependent, proceeding inefficiently through nucleophilic substitution for an unmodified peptide or efficiently through Michael addition for a dehydrated/dethiolated peptide. Uncovering this unusual flavin-dependent paradigm for thioether residue formation advances the understanding in the biosynthesis of aminovinyl-cysteine-containing RiPPs and renews interest in flavoproteins, particularly those involved in non-redox transformations. LanD-like flavoproteins activity, which is flexible in peptide substrate and amenable for evolution by engineering, can be combined with different post-translational modifications for structural diversity, thereby holding promise for peptide macrocyclization/functionalization in drug development by chemoenzymatic or synthetic biology approaches.

INTRODUCTION

Flavin-dependent proteins are ubiquitous in living organisms and participate in numerous redox-related biochemical processes in all domains of life[1,2]. Increasing evidence indicates their critical roles in the biosynthesis of structurally-complex bioactive natural products, many of which are known to be ribosomally synthesized and post-translationally modified peptides (RiPPs)[3,4]. Each RiPP arises from post-translational modifications (PTMs) of a precursor peptide, which is typically composed of an N-terminal leader peptide (LP) and a C-terminal core peptide (CP). Only CP is transformed into mature product(s), usually in a manner dependent on LP. Using 20 proteinogenic amino acids as building blocks, nature creates stunningly diverse RiPPs largely because of unique enzymes for PTMs.

Aminovinyl-cysteine residues[3], i.e., S-[(Z)-2-aminovinyl]-cysteine (AviCys) and S-[(Z)-2-aminovinyl]-3-methyl-cysteine (AviMeCys), are unusual thioether amino acids that occur in an array of RiPPs, including the lanthipeptide epidermin (EPI) and the antitumor non-lanthipeptides thioviridamides (TVAs) and cypemycin (CYP) (**Fig. 1**). These RiPPs share Avi(Me)Cys as part of a macrocyclic ring system that contains the C-terminal 4 or 6 residues of a precursor peptide, despite differences in their overall architectures and other residue characteristics. Knowledge concerning Avi(Me)Cys formation is poor and has gained primarily from investigations into related lanthipeptides[5-9]. Typically, it relies on processing two residues that are flanked by 2 or 4-aa, i.e., dehydration of an internal L-Ser/L-Thr, oxidative decarboxylation of the C-terminal L-Cys and the formation of a thioether linkage via a Michael addition between the resulting 2,3-didehydroalanine (Dha)/2,3-didehydrobutyrine (Dhb) and C-terminal enethiol[10]. In fact, lanthipeptides are characterized by the distinct thioether amino acids lanthionine (Lan) or 3-methyl-lanthionine (MeLan) residues (**Fig. 1**), each of which arises from dehydration of an L-Ser/L-Thr residue and subsequent cyclization that adds the side chain of an L-Cys residue onto the newly formed dehydroamino acid for

3

macrocyclization[11]. The two reactions that occur in tandem can be attributed to the dedicated activities of a LanB dehydratase and a LanC cyclase for Class I lanthipeptides or to the combined activity of a bifunctional (Me)Lan synthetase, i.e., LanM, LanKC or LanL for Class II, III or IV lanthipeptides, respectively[11].

In the biosynthesis of lanthipeptides containing an Avi(Me)Cys residue, it has long been believed that the dehydration activity involved in (Me)Lan formation plays an additional role, by producing the Dha/Dhb residue lying 2 or 4-aa upstream of the C-terminal L-Cys. Transforming this L-Cys to an enethiol nucleophile depends on the activity of a LanD flavoprotein, which belongs to the family of homo-oligomeric flavin-containing cysteine decarboxylases (HFCDs)[10]. LanD activity for oxidative decarboxylation of the C-terminal L-Cys residue of a peptide was known over two decades ago[12, 13]; however, questions regarding Avi(Me)Cys formation remain in terms of timing this reaction and L-Ser/L-Thr dehydration and determining whether the following Michael addition occurs non-enzymatically or, if enzymatically, which enzyme acts as the catalyst, e.g., a LanD flavoprotein, a monofunctional LanC cyclase, or a LanM, LanKC or LanL-represented bifunctional (Me)Lan synthetase[9,14]. Here, we report the unexpected dual activity of LanD-like flavoproteins in the biosynthesis of the non-lanthipeptides TVAs and CYP, where these flavoproteins can oxidatively decarboxylate and cyclize a peptide, either unprocessed or dehydrated/dethiolated, through the formation of an Avi(Me)Cys residue at the C-terminus. This study advance the understanding in the biosynthesis of different Avi(Me)Cys-containing RiPPs, where associated PTM logics and enzymatic mechanisms have not been established since they were discovered.

RESULTS AND DISCUSSION

TvaF$_{S-87}$ can oxidatively decarboxylate and cyclize a peptide

We began investigating Avi(Me)Cys formation in the biosynthesis of the poorly understood non-lanthipeptides TVAs, a class of Avi(Me)Cys-containing thioamide RiPPs (**Fig. 1**)[15,16]. Recent survey of the published genome sequences revealed a number of new biosynthetic gene clusters (BGCs) that share nearly head-to-tail homology with the prototypal TVA BGC (*tva*) characterized in *Streptomyces olivoviridis*[17], and subsequent product mining in selected microorganisms led to a rapid enrichment of the TVA family[18-22], in which several members appear to be promising for cancer treatment. All these BGCs do not encode any homologs of LanC, LanM, LanKC or LanL proteins that possess cyclase activity, consistent with the absence of (Me)Lan in the structures of TVAs. Comparative analysis of the biosynthetic pathways of different Avi(Me)Cys-containing RiPPs revealed that only the LanD-like, HFCD-fold flavoprotein is common (**Supplementary Fig. 1**), indicating the importance of this type protein in the formation of the shared characteristic residue Avi(Me)Cys. To determine the function of this LanD-like flavoprotein in TVA biosynthesis, we overexpressed $TvaF_{S-87}$ (201-aa, a homolog of TvaF from *Streptomyces sp.* NRRL S-87) in *Escherichia coli* (**Supplementary Fig. 2**).

The recombinant $TvaF_{S-87}$ flavoprotein, which proved to non-covalently bind with flavin mononucleotide (FMN), was used to *in vitro* process the N-terminally truncated substrate **1**, a 21-aa-long, C-terminal mimic of the TVA precursor peptide $TvaA_{S-87}$ (**Fig. 2** and **Supplementary Figs. 3** and **4**). This synthetic substrate contains the intact CP sequence of $TvaA_{S-87}$, $S_1VMAAAAT_8VAFHC_{13}$ (numbering relates to residue position in CP, as below), and the 8-aa sequence of LP adjacent to CP. The $TvaF_{S-87}$-catalyzed conversion of **1** produced a -46 Da derivative, **1-I**. Analysis by high-performance liquid chromatography with high-resolution mass spectrometry (HPLC-HR-MS) confirmed that oxidative decarboxylation occurred on the C-terminal ʟ-Cys to generate an enethiol, as previously observed in LanD catalysis for Avi(Me)Cys-containing lanthipeptide biosynthesis[9,12,13]. **1-I** is unstable and tends to

be dethiolated to yield a dominant aldehyde derivative (**1-II**) via hydrolysis. Similar results were recently observed when coexpressing the TvaF and TvaA homologs in *E. coli*[23]. Different from this *in vivo* transformation, we observed an additional -64 Da product, **1-III**, in TvaF$_{S-87}$ catalysis. The observation of **1-III** is unexpected, given the loss of molecular weight (MW) that indicates the occurrence of an additional H_2O elimination (-18 Da). Likely, TvaF$_{S-87}$ possesses more than the activity for oxidative decarboxylation of a peptide.

Based on HR-MS/MS analysis, the difference of **1-III** from **1**, **1-I** and **1-II** was narrowed to the C-terminal sequence **MAAAAT$_8$VAFHC$_{13}$**, in which only L-Thr8 can be the target residue subjected to H_2O elimination (**Supplementary Fig. 4** and **Table 4**). Specifically, no fragment ions were observed from the C-terminal **T$_8$VAFHC$_{13}$** sequence of **1-III**, and corresponding fragment ions were indeed found when analyzing **1**, **1-I** and **1-II**, all of which are linear. Furthermore, **1-III** was resistant to thiol derivatization with N-ehtylmaleimide (NEM); in contrast, **1** and **1-I**, both of which bear a free (ene)thiol group, were sensitive to this treatment, yielding the derivatives conjugated by one NEM molecule (**Supplementary Fig. 5**). These findings supported the notion that **1-III** is a cyclized product resulting from the dual activity of TvaF$_{S-87}$, which can selectively process both L-Cys13 and L-Thr8 and form an AviMeCys linkage. During this process, **1-I** can be released from TvaF$_{S-87}$ as a linear enethiol intermediate.

To further determine the tolerance of TvaF$_{S-87}$ catalysis to changes of the target residues, **1** was subjected to site-specific mutagenesis, resulting in the substrate variants **1-T8A**, **1-T8S**, **1-T8C**, **1-C13A**, **1-C13S** and **1-C13T** (**Fig. 2** and **Supplementary Fig.3** and **Table 4**). The mutations C13A, C13S and C13T caused no substrate conversion. The mutations T8A, T8S and T8C did not affect L-Cys13 processing, leading to the production of the enethiol intermediates **1-T8A-I**, **1-T8S-I** and **1-T8C-I**, respectively. Only **1-T8S-I** can be further transformed to the cyclized product, **1-T8S-III**, and form a distinct AviCys residue. Clearly, TvaF$_{S-87}$ activity tolerates the change of L-Thr8 to

the chemically similar residue L-Ser. We then assayed TvaF$_{S-87}$ activity using a series of further shortened mimics of TvaA$_{S-87}$ CP, i.e., **2** (12-aa), **3** (11-aa), **4** (10-aa), **5** (8-aa) and **6** (6-aa, **T$_8$VAFHC$_{13}$**), all of which lack the LP sequence and vary in truncation extent at the N-terminus of CP (**Supplementary Fig. 3** and **Table 4**). Consequently, both the -46 Da enethiol intermediate (**2-I** or **3-I**) and the -64 Da cyclized product (**2-III** or **3-III**) were observed in the transformation of **2** or **3**; however, the yields were much lowered. Although a trace of enethiol product (-46 Da, **4-I**, **5-I** or **6-I**) was detected when **4**, **5** or **6** was used as the substrate, coupling L-Thr8 via H$_2$O elimination-associated cyclization appears to occur only in the presence of a > 4-aa upstream flanking sequence. Therefore, the upstream residues of Avi(Me)Cys, either from LP or CP, appear to be less dependent but do have impact on the efficiency and completeness of TvaF$_{S-87}$ catalysis.

CypD shares the ability for AviCys formation

To examine the generality of LanD-like flavoprotein activity for distinct non-lanthipeptides, we *in vitro* assayed the activity of CypD (190-aa) involved in the biosynthesis of the CYP. This RiPP is an AviCys-containing linaridin-type RiPP that possesses several L-Thr-derived Dhbs but lacks MeLan residues characteristic of lanthipeptides (**Fig. 1**)[24,25]. Similar to *tva*, the CYP BGC (*cyp*) does not contain any genes encoding the homologs of proteins with (Me)Lan cyclase activity (**Supplementary Fig. 1**). The biosynthesis of this non-lanthipeptide is particularly interesting because two L-Cys residues of the precursor peptide are processed into an AviCys residue. We overexpressed CypD in *E. coli* (**Supplementary Fig. 2**). Different from TvaF$_{S-87}$, purified CypD was observed to bind non-covalently with the cofactor flavin adenine dinucleotide (FAD). Using the synthetic substrate **7** (**GSTIC$_{19}$LVC$_{22}$**), an N-terminally truncated 8-aa CP sequence of the precursor peptide CypA, CypD catalysis produced a -46 Da enethiol intermediate **7-I** (along with its dethiolated derivative **7-II**) and, particularly, a -80 Da cyclized product, **7-III**, in a manner similar

to that found in TvaF$_{S-87}$ catalysis (**Figs. 2** and **3** and **Supplementary Figs. 3，6a** and **Table 4**). The structural

differences between **7-I** and **7-III** were further distinguished by using NEM for thiol derivatization (**Supplementary**

Fig. 7). Similar to the substrate **7**, the intermediate **7-I** contains two (ene)thiol moieties and can be conjugated by two

NEM molecules. The product **7-III** cannot be modified and thus has no free thiol group. The formation of an AviCys

residue was first indicated by HR-MS/MS analysis, which revealed a specific ion fragment ([M+H]$^+$ m/z: calcd.

258.1271, obs. 258. 1271) containing this residue (**Fig. 3** and **Supplementary Table 4**). Then, the CypD-catalyzed

transformation of **7** was scaled up, leading to the accumulation of **7-III** in ~1000 mL of the reaction mixture for

structural elucidation. Consequently, selected Nuclear Magnetic Resonance (NMR) analyses of purified **7-III**

supported the presence of an AviCys residue (**Fig. 3** and **Supplementary Fig. 8** and **Table 6**). Together, the above

results indicate that resembling TvaF$_{S-87}$, CypD catalyzes oxidative decarboxylation and cyclization of a peptide to

form an AviCys linkage along with H$_2$S elimination.

To examine the tolerance of the target residues in CypD catalysis, we synthesized **7** variants, i.e., **7-C19A**, **7-C19S**

and **7-C19T**, by mutating the internal L-Cys19, and **7-C22A**, **7-C22S** and **7-C22T**, by mutating the C-terminal L-

Cys22, respectively (**Supplementary Fig. 3** and **Table 4**). In the presence of CypD, both **7-C19A** and **7-C19T**

underwent oxidative decarboxylation to produce -46 Da enethiols (i.e., **7-C19A-I** and **7-C19T-I**). Only **7-C19S** can

be transformed further into the cyclized product **7-III** (-64 Da, in a yield ~50-fold lower than that in the transformation

of **7**) under the same reaction conditions, while a distinct enethiol intermediate, **7-C19S-I** (-46 Da), accumulated.

Clearly, L-Ser can substitute for the internal L-Cys19 during AviCys formation, indicating that in this CypD catalysis,

L-Ser shares with L-Cys a common mechanism for H$_2$O or H$_2$S elimination. CypD did not transform the variants **7-**

C22A, **7-C22S** and **7-C22T**, again validating the necessity of the C-terminal L-Cys.

8

TvaF$_{S-87}$ and CypD are dodecameric Rossmann-fold flavoproteins

To gain structural insights into the flavoproteins TvaF$_{S-87}$ and CypD for non-lanthipeptide biosynthesis, we determined their apo-form crystal structures for comparative analysis (to 2.27 Å and 2.38 Å resolutions for TvaF$_{S-87}$ and CypD, respectively) (**Fig. 4** and **Supplementary Table 7**). Despite low homology observed in sequence, TvaF$_{S-87}$ and CypD are highly similar in overall structure to each other and to characterized LanD proteins (**Figs. 4a-e** and **Supplementary Fig. 9a-e**) (e.g., EpiD and MrsD in the biosynthesis of the lanthipeptides EPI and mersacidin[26,27], respectively). Each protein forms a stable dodecamer composed of four trimeric units that are located at the vertices of a tetrahedron. In a trimer unit, each Rossmann-fold monomer associates with a flavin cofactor (**Supplementary Figs. 10** and **11**). This factor is located in a cavity at the interface of two monomers, of which one plays a predominant role for hydrophobic contacts and polar interactions, facilitating packing with the third monomer. Compared to CypD, which binds FAD, TvaF$_{S-87}$ is more similar to the archetypal LanD protein EpiD largely because they share a manner to accommodate the FMN cofactor (**Supplementary Fig. S9b, c**). Recently, two research groups reported similar observations based on the crystallization of CypD and a different TvaF homolog, respectively[23,28]. Notably, we observed two slightly different conformations of the monomers in the crystal structure of TvaF$_{S-87}$ (**Supplementary Fig. S9f**). The region connecting β5 and α9 is disordered and unsolved in some monomers. Whereas in others, it forms a short anitparallel β-sheet resembling that as part of the substrate-binding clamp observed in the EpiD complex with a peptidyl substrate mimic (PDB ID: 1G5Q) (**Supplementary Fig. 9g**), suggesting that TvaF$_{S-87}$ and EpiD adopt a similar mode for substrate-binding in the first oxidative decarboxylation step.

CypD and TvaF$_{S-87}$ share a great majority of the residues responsible for interaction with the flavin cofactor, either FMN or FAD, through hydrophobic contacts and hydrogen bonds with the common isoalloxazine ring

9

(**Supplementary Figs. 10-12**). In their structures, interaction between trimeric units are attributed to residues within the α1, α2, and α3 regions of the monomer (**Supplementary Figs. 12-14**). We selectively mutated the key residues V28 and M62, respectively, to L-Asp in these regions to destabilize this interaction in TvaF$_{S-87}$. The two mutations exclusively led to the decomposition of the active TvaF$_{S-87}$ dodecamer to stable trimer, which retains the ability to bind FMN but losses the activity to transform the peptide substrate **1** (**Supplementary Figs. 4** and **13c**). Thus, the formation of a dodecamer is necessary for the activity of LanD-like flavoproteins. This structural nature cannot be applied to 4'-phosphopantothenoylcysteine (PPC) decarboxylases, which fall into the same HFCD family but function in an active trimer form during the biosynthesis of coenzyme A[29].

Capture of a cyclization transition state in the CypD complex

After numerous attempts, we eventually resolved the co-crystal structure of CypD in complex with **8** (**GSTIS$_{19}$LVS$_{22}$**) to 2.3 Å resolution (**Fig. 4g-i** and **Supplementary Table 8**). **8** is a synthetic derivative of the peptide substrate **7** and fused with CypD by a GS-repeat linker. Compared to **7**, this derivative, in which the two residues involved in AviCys formation (i.e., L-Cys19 and L-Cys22) are replaced with L-Ser, cannot be transformed by CypD. The complex of CypD highly resembles its apo-form, as twelve FAD-associated monomers adopt a similar conformation for assembly to a homododecamer accommodating twelve **8** through the formation of four trimeric units (**Fig. 4d-i** and **Supplementary Fig. 15a**). For each bound octapeptide, while the N-terminal tripeptide **GST** cannot be modeled due to a lack of electron density, the remaining pentapeptide **IS$_{19}$LVS$_{22}$** can be clearly defined. The **IS$_{19}$LVS$_{22}$** sequence adopts an obvious bent conformation and packs extensively with a solvent-exposed, highly charged pocket formed between two trimeric units (**Fig. 5a** and **Supplementary Fig. 16**). This pocket, which is close to a bound FAD cofactor, overlaps the putative substrate binding region revealed by superimposition of CypD and the mutant EpiD

(H67N) complex with a pentapeptide substrate mimic (**Supplementary Fig. 17**)[26].

In the CypD complex (**Fig. 5**), the hydrophobic side chains of the residues L-Val21, L-Leu20, and L-Ile18 of **8** occupy a hydrophobic groove formed by the residues I19, L23, T128, P130, V168 and F170 of one monomer and the residue L60 of another monomer within the neighboring trimeric unit. The negatively charged, C-terminal carboxyl group of **8** forms a weak hydrogen bond with the monomer residue H59, while the backbone carboxyl group of L-Leu20 interacts with the polar side chains of the monomer residues S18 and S20 via two highly specific hydrogen bonds. Theses interactions allow approaching between the polar side chains of L-Ser22 and L-Ser19 to form an intramolecular hydrogen bond in **8**, where the bent conformation is further stabilized (**Fig. 5b**). This conformation leads to a hypothesis that during the transformation of **7/7-C19S**, cyclization to form an AviCys residue in CypD catalysis proceeds via nucleophilic substitution between L-Cys19/L-Ser19 and the C-terminal enethiol resulting from oxidative decarboxylation. Distinct from the mutant EpiD (H67N) complex, the CypD complex with **8** unlikely restores a transition state for L-Cys22 oxidative decarboxylation given their significant differences in substrate binding. Instead, enethiol production might be bypassed (assumedly due to replacing the C-terminal L-Cys with L-Ser) in this CypD complex, where a transition state for the following cyclization reaction was captured. The mechanism for CypD-catalyzed, L-Cys22 oxidative decarboxylation was recently proposed to be similar to that in EpiD catalysis[30]; thus, repositioning the enethiol intermediate at the active site of CypD appears to be necessary for nucleophilic substitution. In the CypD complex with **8**, the backbone carboxyl group of the substrate residue L-Val21 approaches N5 of the isoalloxazine ring of a bound FAD within a distance of 3.5 Å (**Fig. 5b**). The reduction of FAD to $FADH_2$ during L-Cys22 processing can protonate N5 and facilitate the formation of a hydrogen bond with the backbone carboxyl group of L-Val21. This polar interaction might compensate the loss of the weak hydrogen bond between the monomer residue H59 and the C-terminal L-Ser22 of **8** and thus can stabilize the bent conformation of

11

the decarboxylated enethiol intermediate.

To validate this hypothesis, we assayed CypD activity using the substrate **7-C19S** in which L-Ser19 was deuterium-labeled (i.e., L-[2,3,3-D$_3$]Ser). As expected, -46 Da intermediate **7-C19S-I** and -64 Da product **7-III** were observed (**Supplementary Fig. 18**). They both share +3 Da in MW compared to those produced in the control reaction where unlabeled L-Ser was used. The observation of +3 Da **7-III** strongly supported that following C-terminal L-Cys22 processing, coupling L-Ser19 with the newly formed enethiol proceeds via nucleophilic substitution because all deuterium atoms were retained. Most likely, TvaF$_{S-87}$ shares the mechanism for Avi(Me)Cys formation in the transformation of **1/1-T8S**.

Determination of the nature of CypD substrate in CYP biosynthesis

Although the dual activity of TvaF$_{S-87}$ and CypD is experimentally and structurally evidenced, it is unlikely that these flavoproteins act on an unmodified peptide substrate for Avi(Me)Cys formation in the biosynthesis of TVAs or CYP because of the low yield of cyclized products versus the relatively high yield of enethiol intermediates (for calculation, their aldehyde derivative were included). To examine PTM dependence, we assayed CypD activity using the variants of **7/7-C19S** (i.e., **7-C19S-Ac**, **7-C19S-Glu**, **7-C19S-P** and **7-Dha19**) in which L-Cys19/L-Ser19 was replaced with acetylated L-Ser, glutamylated L-Ser, phosphorylated L-Ser and Dha, respectively (**Fig. 2**). Acetylation, glutamylation and phosphorylation can form a leaving group, which might facilitate cyclization. While CypD failed to convert **7-C19S-Glu** and **7-C19S-P** likely due to the steric or electronic effects that arise from glutamyl or phosphoryl substitution, this flavoprotein tolerated the substrate **7-C19S-Ac** and produced **7-III** with a yield ~2-fold higher than that in the conversion of the unmodified substrate **7-C19S**. In both conversions, the yields of the enethiol

12

intermediates were comparable (**Supplementary Fig. 6b**), indicating that L-Ser acetylation does motivate enethiol coupling.

In contrast, the CypD-catalyzed conversion of **7-Dha19** produced **7-III** with a yield much higher than those in the conversions of unmodified **7** (by ~10-fold) and **7-C19S** (by ~500-fold). Notably, this conversion resulted in an additional product, **7-IV**, which was even ~2.5-fold higher than **7-III** in yield (**Fig. 2**). **7-IV** and **7-III** were validated to be identical to each other in MW, MS/MS analysis, AviCys-containing and chemical derivatization (**Fig. 2** and **Supplementary Fig. 19**), suggesting that they are diastereomers different from each other in the Cα configuration of the AviCys residue. **7-III** is believed to have an L-AviCys residue because nucleophilic substitution does not change the configuration at Cα; thereby, **7-IV** might have a D-AviCys residue. This prediction was further supported by the success in converting the substrate **7-d-C19** (in which L-Cys19 was replaced with D-Cys) to produce **7-IV** only (**Fig. 2** and **Supplementary Fig. 6b**). Remarkably, enethiol intermediate (i.e., **7-Dha19-I**) did not accumulated during the CypD-catalyzed conversion of **7-Dha19**, indicating that Michael addition to the Dha19 residue immediately follows oxidative decarboxylation of the C-terminal L-Cys22 and occurs efficiently for cyclization. Together, CypD activity prefers a dethiolated/dehydrated substrate (rather than hijacking the L-Cys19/L-Ser19-functionalized intermediate) and tends to produce a D-AviCys-containing product. In fact, D-AviCys residue was found exclusively in related natural products that were characterized in stereochemistry[31,32]. The above results argue against the recent report by Zhang *et al.* who failed to observe AviCys formation in the conversion of a Dha-containing peptide and thus concluded that CypD has no cyclase activity[33]. In this study, two L-Lys residues were added at the N-terminus of the 7-aa-long, C-terminal CP mimic of CypA, yielding **KKSTI-Dha$_{19}$-LVC$_{22}$**, which CypD can oxidatively decarboxylate but cannot cyclize. Instead, this finding is consistent with that observed in aforementioned TvaF$_{S-87}$ catalysis, i.e., cyclization is an enzymatic process and the sequence upstream of those

13

involved in Avi(Me)Cys formation are necessary for the completeness of CypD catalysis.

Examination of the active site residues in CypD

Based on the crystal structure of CypD in complex with **8**, we first utilized **7-Dha19** as the substrate to examine the involvement of the active site residues in CypD catalysis (**Fig. 5**). While the mutation S20D completely abolished CypD activity, the mutations L23A, L23Q, F170Q, F170A, H59D and H29R, which change the physiochemical and/or steric properties of residues, significantly lowered the yields but did not alter the ratio between the diastereomers **7-III** and **7-IV** (~1:2.5) (**Fig. 5c**). No enethiol intermediates (and their aldehyde derivatives) accumulated in these assays. Given the fact that CypD employs a same pocket for accommodating **7-Dha19** for oxidative decarboxylation and the enethiol intermediate for subsequent Michael addition, the above residues might contribute to the both reactions for substrate/enethiol intermediate binding during AviCys formation. Processing the C-terminal L-Cys22 may serve as the rate-limiting step, for which CypD and TvaF$_{S-87}$ might share a flavin-dependent mechanism with the archetypal LanD protein EpiD[10]. At the active site, the peptide substrate approaches flavin via the thiol group of the C-terminal L-Cys, allowing oxidation occurring to initiate decarboxylation and form a reactive enethiolate ion. During this process, the residues L-His67 in EpiD could act as a base and participate in C-terminal L-Cys activation. Consistently, the TvaF$_{S-87}$ mutant in which the corresponding residue H85 is changed to L-Ala completely lost activity (**Supplementary Fig. 4**). In contract, CypD possesses a distinct N80 according to sequence alignment (**Supplementary Fig. 12**). Whereas the mutation N80D nearly completely abolished **7-III**/**7-IV** production, the mutation N80H resulted in a ~4-fold increase of the both cyclized products, indicating that a stronger base facilitates the deprotonation of the C-terminal L-Cys (**Fig. 5c**). The mutations H59A and S20A also improved the yields of **7-III** and **7-IV** (by ~1-fold for each), supporting that the hydrophobicity of the binding pocket can affect

14

the interaction with the substrate/ennethiol intermediate (**Fig. 5**). We then assayed the activities of the above CypD mutants by using the unprocessed substrate **7**. Overall, these mutants produced **7-III** and the enethiol intermediate/aldehyde derivative (**7-I/7-II**) with a similar yield trend (**Supplementary Fig. 20**), indicating that the cyclization activity of CypD or its variants can be examined using the readily prepared, unmodified peptide substrates (albeit lower in yields). Therefore, the proximity effect resulting from the bent conformation at the active site play a critical role in coupling L-Cys19 and L-Cys22 for AviCys formation.

CONCLUSION

Following an unexpected finding that LanD-like flavoproteins can function alone to cyclize a peptide, we uncovered a new flavin-dependent PTM paradigm for the formation of thioether amino acids occurring in non-lanthipeptides (**Fig. 6**). We discovered this paradigm by assaying TvaF$_{S-87}$ activity in the biosynthesis of the thioamide RiPPs TVAs, and subsequently validated its generality upon more detailed biochemical and structural investigations into CypD in the biosynthesis of the linaridin RiPP CYP. LanD-like flavoprotein first catalyzes oxidative decarboxylation of the invariable C-terminal L-Cys residue; then, it adds the resulting enethiol nucleophile to the 2 or 4-aa upstream internal residue, which is variable and can be L-Ser, L-Cys or L-Thr, and forms an Avi(Me)Cys linkage for peptide cyclization. This cyclization is an enzymatic process and occurs largely due to proximity effect at the active site of the enzyme by repositioning and forcing the enethiol intermediate to a bent conformation. Mechanistically, it can be substrate-dependent. To process an unmodified peptide, the cyclization goes inefficiently through nucleophilic substitution that causes the accumulation of the enethiol intermediate, and the upstream LP and CP sequences can affect the catalytic efficiency. In contrast, cyclizing a Dha-containing peptide, as shown in CypD catalysis where reduced flavin cofactor appears to be necessary to stabilize the bent conformation of the enethiol intermediate, proceeds efficiently via

15

Michael addition to produce two diastereomers, of which the product with a D-AviCys residue is dominant. Consistent with the finding that CypD catalysis favors a dethiolated substrate, a Dha19-containing peptide intermediate was observed by heterologous overexpression of the *cyp* gene cluster in which *cypD* was inactivated[25]. TvaF and CypD are capable of oxidatively processing the C-terminal L-Cys and non-oxidatively coupling the internal residue, raising a question of whether LanD proteins share this dual activity in the biosynthesis of related lanthipeptides, where (Me)Lan cyclase activity has long been suggested to participate in Avi(Me)Cys formation. Recently, the combined activities of LanKC and LanD homologs proved to be necessary to form an AviCys-like avionin motif in the biosynthesis of lipolanthines; however, it remains to determine the individual roles of the either homologs in this process[34]. The study presented here renews great interest in flavoproteins, as their catalytic mechanisms have not been fully appreciated in various biochemical processes, particularly those involving additional non-redox reactions[35-37]. The LanD-like flavoproteins tolerate a wide variety of peptide substrates, is evolvable by engineering, and can be combined with different PTM enzymes for structural diversity. Using chemoenzymatic or synthetic biology strategies, enzymatic Avi(Me)Cys formation has potential for peptide drug development by application in peptide macrocyclization and functionalization, which remain a challenge to current chemical synthesis approaches.

REFERENCES:

1. R. L. P. Fagan, B. A. Palfey, *Flavin-dependent enzymes.* In Ch.3 in Comprehensive Natural Products II vol 7. (Edited by T. P. Begley. Elsevier; 2010).

2. Walsh CT, Wencewicz TA. Flavoenzymes: versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* **30,** 175-200 (2013).

3. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160 (2013).

16

4.   Ortega MA, van der Donk WA. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **23**, 31-44 (2016).

5.   Schnell N, Entian KD, Schneider U, Gotz F, Zahner H, Kellner R*, et al.* Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* **333**, 276-278 (1988).

6.   Schnell N, Engelke G, Augustin J, Rosenstein R, Ungermann V, Gotz F*, et al.* Analysis of genes involved in the biosynthesis of lantibiotic epidermin. *Eur. J. Biochem.* **204**, 57-68 (1992).

7.   Augustin J, Rosenstein R, Wieland B, Schneider U, Schnell N, Engelke G*, et al.* Genetic analysis of epidermin biosynthetic genes and epidermin-negative mutants of Staphylococcus epidermidis. *Eur. J. Biochem.* **204**, 1149-1154 (1992).

8.   Majer F, Schmid DG, Altena K, Bierbaum G, Kupke T. The Flavoprotein MrsD Catalyzes the Oxidative Decarboxylation Reaction Involved in Formation of the Peptidoglycan Biosynthesis Inhibitor Mersacidin. *J. Bacteriol.* **184**, 1234-1243 (2002).

9.   Ortega MA, Cogan DP, Mukherjee S, Garg N, Li B, Thibodeaux GN*, et al.* Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine during NAI-107 Biosynthesis. *ACS Chem. Biol*. **12**, 548-557 (2017).

10.   Sit CS, Yoganathan S, Vederas JC. Biosynthesis of aminovinyl-cysteine-containing peptides and its application in the production of potential drug candidates. *Acc. Chem. Res*. **44**, 261-268 (2011).

11.   Repka LM, Chekan JR, Nair SK, van der Donk WA. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **117**, 5457-5520 (2017).

12.   Kupke T, Stevanovic S, Sahl HG, Gotz F. Purification and characterization of EpiD, a flavoprotein involved in the biosynthesis of the lantibiotic epidermin. *J. Bacteriol*. **174**, 5354-5361 (1992).

13.   Kupke T, Kempter C, Jung G, Gotz F. Oxidative decarboxylation of peptides catalyzed by flavoprotein EpiD. Determination of substrate specificity using peptide libraries and neutral loss mass spectrometry. *J. Biol. Chem*. **270**, 11282-11289 (1995).

14.   Gotz F, Perconti S, Popella P, Werner R, Schlag M. Epidermin and gallidermin: Staphylococcal lantibiotics. *Int. J. Med. Microbiol.* **304**, 63-71 (2014).

15.   Hayakawa Y, Sasaki K, Nagai K, Shin-ya K, Furihata K. Structure of thioviridamide, a novel apoptosis inducer from Streptomyces olivoviridis. *J. Antibiot. (Tokyo)* **59**, 6-10 (2006).

16.   Izumikawa M, Kozone I, Hashimoto J, Kagaya N, Takagi M, Koiwai H*, et al.* Novel thioviridamide derivative--JBIR-140: heterologous expression of the gene cluster for thioviridamide biosynthesis. *J.*

17

Antibiot. (Tokyo) **68**, 533-536 (2015).

17. Izawa M, Kawasaki T, Hayakawa Y. Cloning and heterologous expression of the thioviridamide biosynthesis gene cluster from Streptomyces olivoviridis. *Appl. Environ. Microbiol*. **79**, 7110-7113 (2013).

18. Kjaerulff L, Sikandar A, Zaburannyi N, Adam S, Herrmann J, Koehnke J, *et al.* Thioholgamides: Thioamide-Containing Cytotoxic RiPP Natural Products. *ACS Chem. Biol*. **12**, 2837-2841 (2017).

19. Frattaruolo L, Lacret R, Cappello AR, Truman AW. A Genomics-Based Approach Identifies a Thioviridamide-Like Compound with Selective Anticancer Activity. *ACS Chem. Biol.* 2017, **12**, 2815-2822 (2017).

20. Santos-Aberturas J, Chandra G, Frattaruolo L, Lacret R, Pham TH, Vior NM, *et al.* Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Res.* **47**, 4624-4637 (2019).

21. Frattaruolo L, Fiorillo M, Brindisi M, Curcio R, Dolce V, Lacret R, *et al.* Thioalbamide, A Thioamidated Peptide from Amycolatopsis alba, Affects Tumor Growth and Stemness by Inducing Metabolic Dysfunction and Oxidative Stress. *Cells* **8**, (2019). DOI: 10.3390/cells8111408.

22. Li Y, Liu J, Tang H, Qiu Y, Chen D, Liu W. Discovery of New Thioviridamide-Like Compounds with Antitumor Activities. *Chinese J. Chem.* **37**, 1015-1020 (2019).

23. Lu J, Li J, Wu Y, Fang X, Zhu J, Wang H. Characterization of the FMN-Dependent Cysteine Decarboxylase from Thioviridamide Biosynthesis. *Org. Lett*. **21**, 4676-4679 (2019).

24. Minami Y, Yoshida K, Azuma R, Urakawa A, Kawauchi T, Otani T, *et al.* Structure of Cypemycin, a New Peptide Antibiotic. *Tetrahedron Lett.* **35**, 8001-8004 (1994).

25. Claesen J, Bibb M. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proc. Natl. Acad. Sci. U. S. A*. **107**, 16297-16302 (2010).

26. Blaesse M, Kupke T, Huber R, Steinbacher S. Crystal structure of the peptidyl-cysteine decarboxylase EpiD complexed with a pentapeptide substrate. *EMBO J*. **19**, 6299-6310 (2000).

27. Blaesse M, Kupke T, Huber R, Steinbacher S. Structure of MrsD, an FAD-binding protein of the HFCD family. *Acta. Crystallogr. D Biol. Crystallogr.* **59**, 1414-1421 (2003).

28. Mo T, Yuan H, Wang F, Ma S, Wang J, Li T, *et al.* Convergent evolution of the Cys decarboxylases involved in aminovinyl-cysteine (AviCys) biosynthesis. *FEBS Lett.* **593**, 573-580 (2019).

29. Stanitzek S, Augustin MA, Huber R, Kupke T, Steinbacher S. Structural basis of CTP-dependent peptide bond formation in coenzyme A biosynthesis catalyzed by Escherichia coli PPC synthetase. *Structure* **12**,

1977-1988 (2004).

30.   Liu L, Chan S, Mo T, Ding W, Yu S, Zhang Q*, et al.* Movements of the Substrate-Binding Clamp of Cypemycin Decarboxylase CypD. *J. Chem. Inf. Model.* **59**, 2924-2929 (2019).

31.   Prasch T, Naumann T, Markert RL, Sattler M, Schubert W, Schaal S*, et al.* Constitution and solution conformation of the antibiotic mersacidin determined by NMR and molecular dynamics. *Eur. J. Biochem.* **244**, 501-512 (1997).

32.   Shang Z, Winter JM, Kauffman CA, Yang I, Fenical W. Salinipeptins: Integrated genomic and chemical approaches reveal unusual D-amino acid-containing ribosomally synthesized and post-translationally modified peptides (RiPPs) from a Great Salt Lake Streptomyces sp. *ACS Chem. Biol.* **14,** 415-425 (2019).

33.   Ding W, Yuan N, Mandalapu D, Mo T, Dong S, Zhang Q. Cypemycin Decarboxylase CypD Is Not Responsible for Aminovinyl-Cysteine (AviCys) Ring Formation. *Org. Lett.* **20,** 7670-7673 (2018).

34.   Wiebach V, Mainz A, Siegert MJ, Jungmann NA, Lesquame G, Tirat S*, et al.* The anti-staphylococcal lipolanthines are ribosomally synthesized lipopeptides. *Nat. Chem. Biol.* **14**, 652-654 (2018).

35.   Bornemann S. Flavoenzymes that catalyse reactions with no net redox change. *Nat. Prod. Rep.* **19**, 761-772 (2002).

36.   Tian Z, Sun P, Yan Y, Wu Z, Zheng Q, Zhou S, *et al*. An enzymatic [4+2] cyclization cascade creates the pentacyclic core of pyrroindomycins. *Nat. Chem. Biol*. **11**, 259-265 (2015).

37.   Zheng Q, Gong Y, Guo Y, Zhao Z, Wu Z, Zhou Z, *et al*. Structural insights into a flavin-dependent [4+2] cyclase that catalyzes *trans*-decarlin formation in pyrroindomycin biosynthesis. *Cell Chem. Biol.* **25**, 718-727 (2018).

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

Data availability Atomic coordinates and structural factors for the reported crystal structures of TvaF$_{S-87}$ and CypD are deposited in the Protein Data Bank under the accession numbers 6KTP, 6KTT, 6KTI, and 6KT9.

FIGURE LEGENDS

Figure 1. Precursor peptides and structures of Avi(Me)Cys-containing RiPPs EPI, TVA-YJ-2 and CYP. In each precursor peptide, where the LP and CP sequences are indicated, the C-terminal L-Cys and the internal L-Ser/L-Thr or L-Cys that are involved in Avi(Me)Cys formation are numbered and colored in red and blue, respectively. The Avi(Me)Cys residues that arise from processing the above two target residues are highlighted by color in dashed rectangle. Lan and MeLan residues in EPI are indicated. For other unusual residues: Dha, 2,3-didehydroalanine; Dhb, 2,3-didehydrobutyrine; alle, L-*alle*-isoleucine; and Abu, α-aminobutyric acid.

Figure 2. LanD-type flavoprotein-catalyzed Avi(Me)Cys formation. Left, HPLC-MS traces; and right, substrates and their associated enethiol intermediates and cyclized products. The two residues involved in Avi(Me)Cys formation, i.e., the C-terminal L-Cys (red) and its upstream L-Thr/L-Ser or L-Cys (blue), are shown in balls. Conversions were conducted at 30°C for 2 hr in the presence of TvaF$_{S-87}$ or CypD. (**a**) *In vitro* assays of TvaF$_{S-87}$ activity using the 21-aa-long, C-terminal CP sequence of the precursor peptide TvaA$_{S-87}$ and its variants. i, **1**; ii, **1-T8A**; iii, **1-T8C**; iv, **1-T8S**; v, **1-C13A**; vi, **1-C13S**; and vii, **1-C13T**. (**b**) *In vitro* assays of CypD activity using the 8-aa-long, C-terminal CP sequence of the precursor peptide CypA and its variants. i, **7-Dha19**; ii, **7**; iii, **7-d-C19**; iv, **7-C19S**; v, **7-C19S-Ac**; vi, **7-C19S-Glu**; and vii, **7-C19S-P**. For details of the HR-MS and HR-MS/MS data, **Supplementary Fig. 6b** and **Table 4**.

Figure 3. Spectral analysis of the cyclized AviCys-containing products **7-III** and **7-IV**. (**a**) Structures of **7-III** and **7-IV**. Fragmentations by MS/MS are indicated. The carbon atom of which **7-III** and **7-IV** are different in chirality is labeled with asterisk. (**b**) HPLC-MS. (**c**) HR-MS. (**d**) MS/MS. (**e**) Selected ^{1}H NMR signals for **7-III**. 2D TOCSY f1

21

slice at f2 chemical shifts (600 MHz, DMSO-d_6) of H-1 of **7-III** and the key of spin-spin coupling constant ($J_{H1/H2}$ = 7.3 Hz) was highlighted. For details of the spectral data, see **Supplementary Fig. 8** and **Tables 4** and **6**.

Figure 4. Crystal structures of LanD-like flavoproteins. For TvaF$_{S-87}$ (**a-c**) or CypD (**d-f**) in apo-form, the dodecamer is shown by viewing along a 3-fold axis and in a ribbon-stick diagram (left) or surface representation (middle), while the monomer in complex with FMN or FAD (stick model) is shown by ribbon diagram (right). For CypD in complex with **IS$_{19}$LVS$_{22}$** (the C-terminal pentapeptide sequence of **8**, **g-i**), the dodecamer complex is shown in a ribbon-stick diagram (left) and surface representation (middle), and the monomer in complex with FAD and **IS$_{19}$LVS$_{22}$** (stick model) is shown by ribbon diagram (right).

Figure 5. Identification of the residues involved in CypD catalysis at its active site. (**a**) The combined surface representation and the stick-ball model showing the hydrophobic binding interface between CypD and **IS$_{19}$LVS$_{22}$**. The FAD cofactor and bound **IS$_{19}$LVS$_{22}$** are displayed in the stick-ball model, and the CypD dodecamer in the surface view is colored by amino acid types. In the surface model of CypD, the hydrophobic amino acid residues are drawn in yellow, the positively charged residues in blue, the negatively charged residues in red, and the uncharged polar residues in gray. (**b**) Stereoviews of the ribbon-stick models showing the detailed interactions of CypD with **IS$_{19}$LVS$_{22}$** in the complex. The hydrogen bonds involved in the interaction are shown as dotted lines. The N5 atom on the isoalloxazine ring of FAD and the putative hydrogen bond between the backbone carboxyl group of V$_{21}$ and reduced N5 atom are also indicated. (**c**) Comparison of the enzymatic activities of CypD with its variants based on examination of the production of **7-III** and **7-IV**.

Figure 6. Mechanistic comparison in the formation of the thioether amino acids. (**a**) Formation of (Me)Lan residues.

(**b**) Formation of Avi(Me)Cys residues. The cyclization activity of bifunctional LanD-like flavoproteins in the biosynthesis of non-lanthipeptides can be substrate-dependent, i.e., nucleophilic substitution for processing an unmodified peptide (*route a*) and Michael addition for processing a Dha-containing peptide (*route b*).
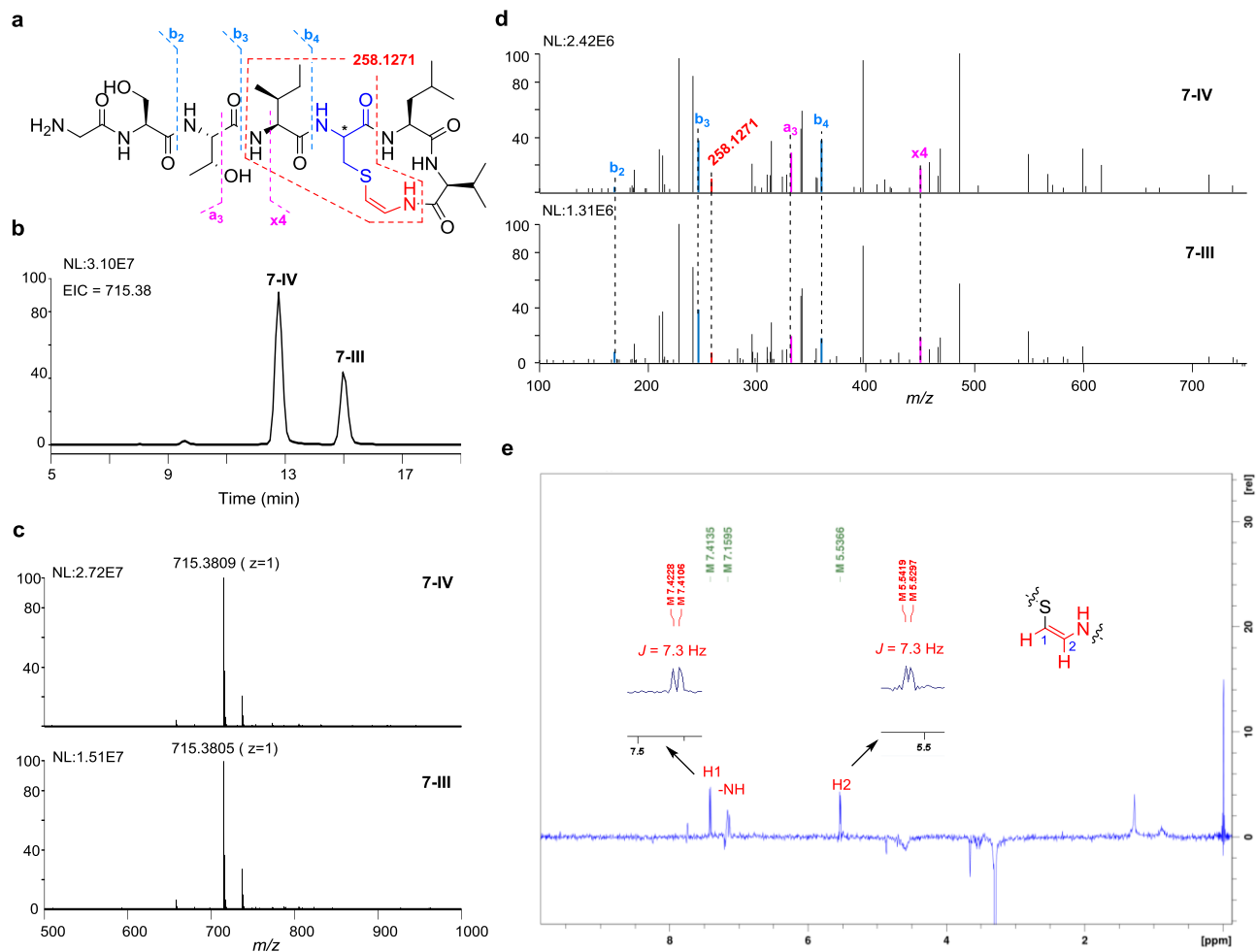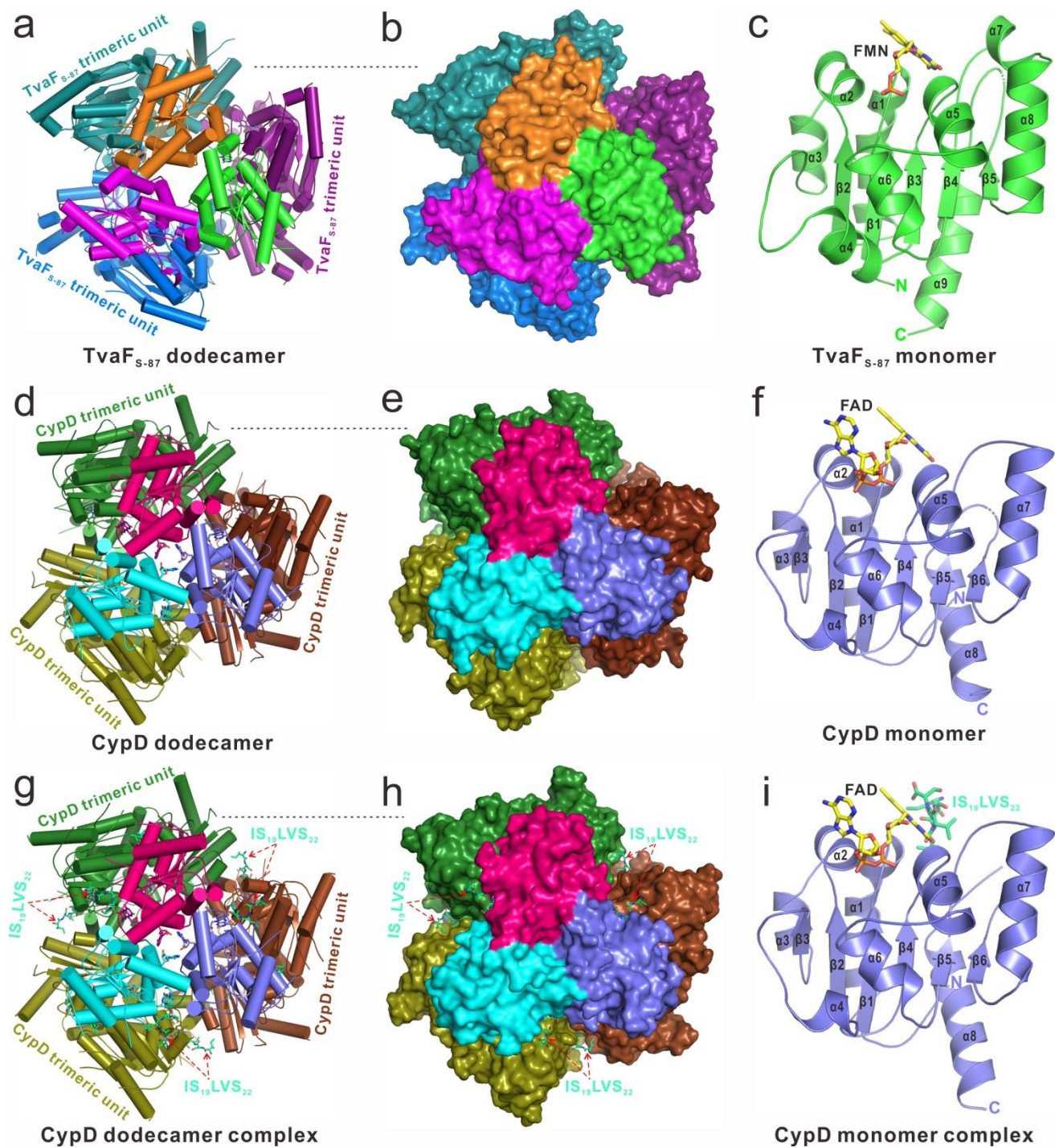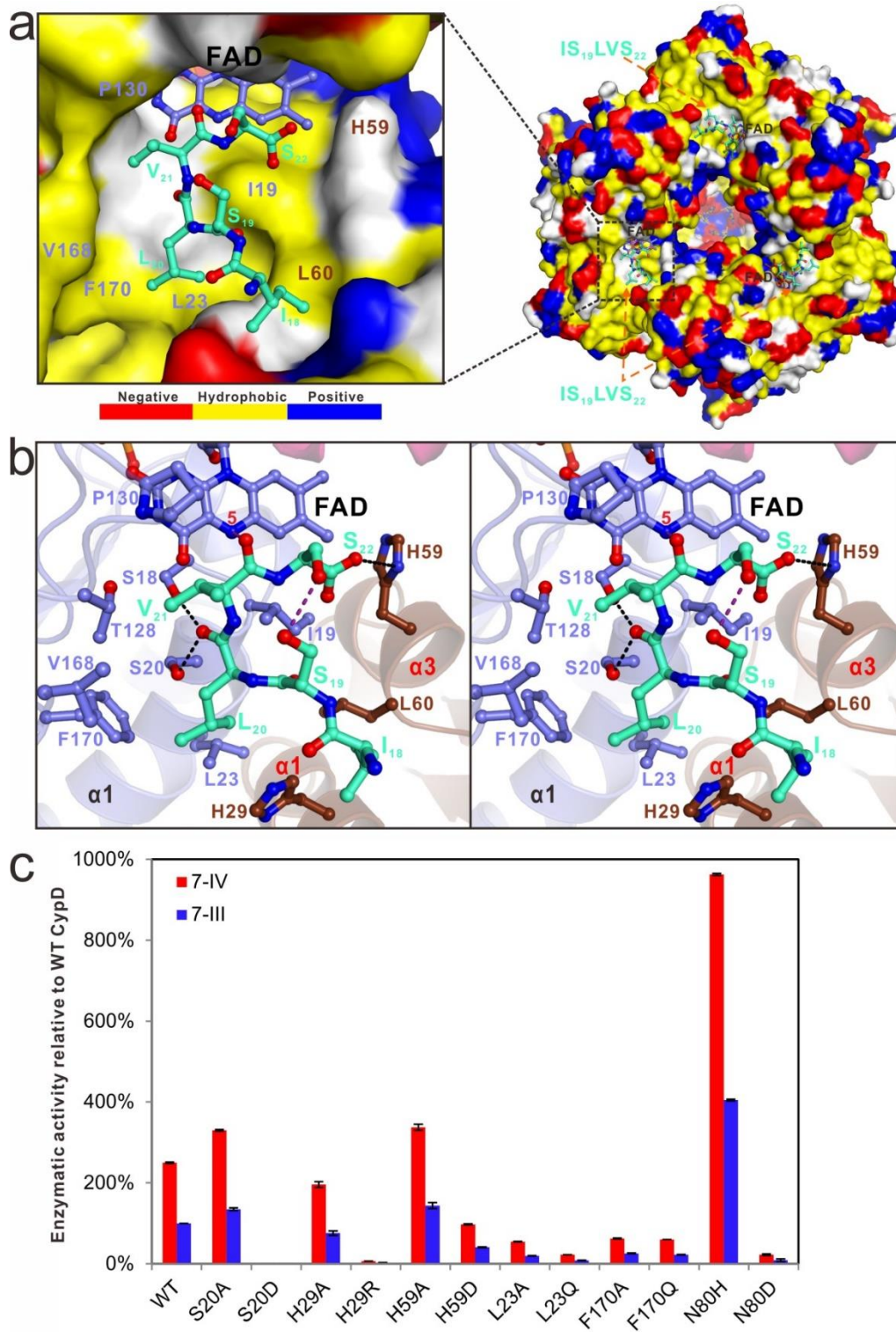
Figure 1.
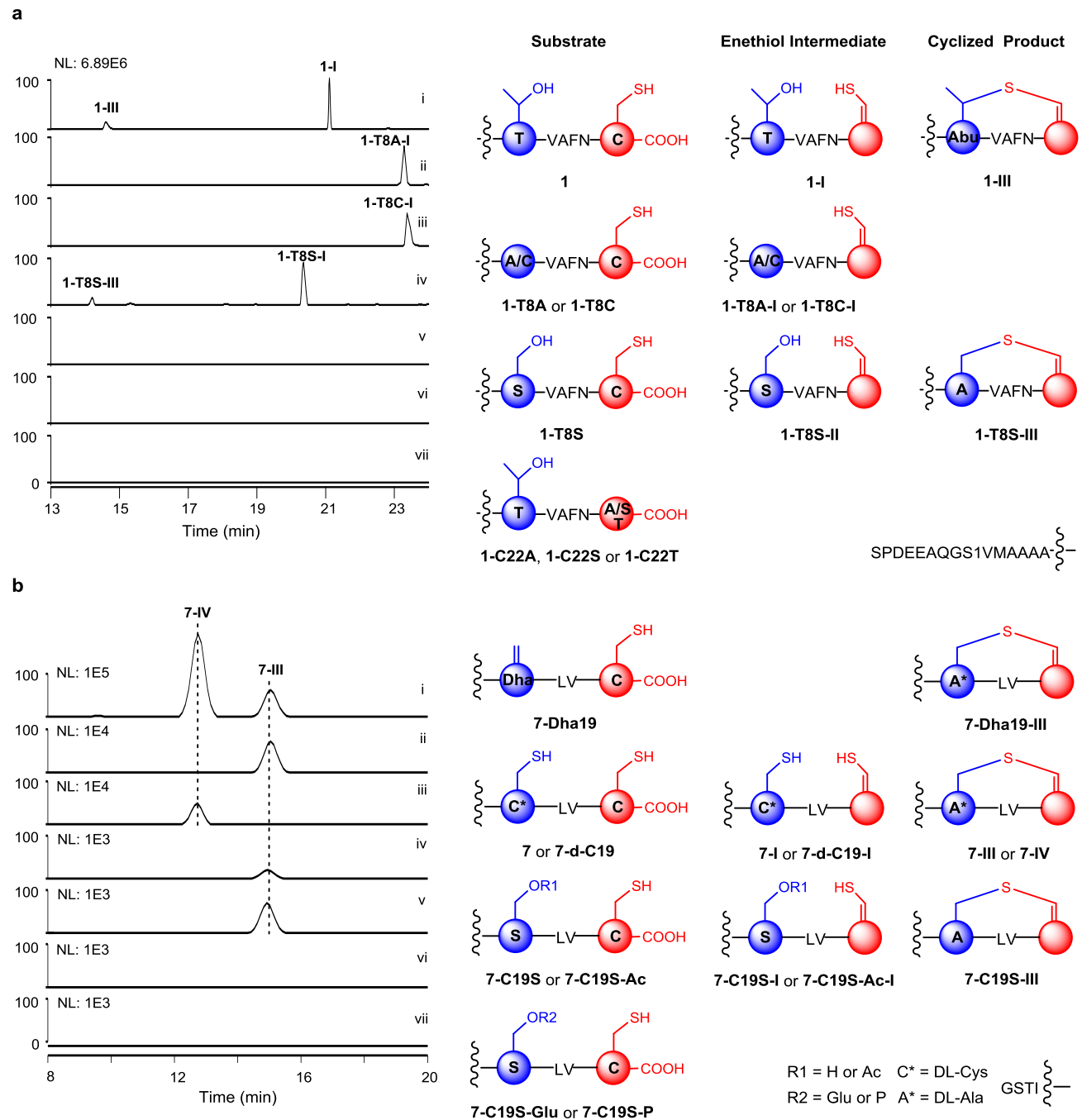
Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.