nature chemical biology

BRIEF COMMUNICATION

https://doi.org/10.1038/s41589-019-0460-x

Check for updates

Structural basis of cycloaddition in biosynthesis of iboga and aspidosperma alkaloids

Lorenzo Caputi[1,5], Jakob Franke[2,5], Kate Bussey[3], Scott C. Farrow[3], Ivo Jose Curcino Vieira[4], Clare E. M. Stevenson[3], David M. Lawson[3]✉ and Sarah E. O'Connor[1]✉

Cycloaddition reactions generate chemical complexity in a single step. Here we report the crystal structures of three homologous plant-derived cyclases involved in the biosynthesis of iboga and aspidosperma alkaloids. These enzymes act on the same substrate, named angryline, to generate three distinct scaffolds. Mutational analysis reveals how these highly similar enzymes control regio- and stereo-selectivity.

Aspidosperma and iboga alkaloids are plant-derived natural products with diverse biological properties, including anticancer and antiaddiction activities (Supplementary Fig. 1; ref. [1]). Despite dissimilar structural architectures, these molecules are synthesized through nearly identical pathways[2]. Extensive evidence suggests that the structural divergence occurs as a result of formal Diels–Alder reactions, where 'Diels–Alderases' would act on the same substrate, dehydrosecodine, to generate the iboga and aspidosperma scaffolds[3–5]. The dihydropyridine group of dehydrosecodine could act as a diene and the methyl acrylate as a dienophile to form the iboga alkaloid (+)-catharanthine (**4**). Alternatively, the dihydropyridine could act as the dienophile and the vinyl indole could act as the diene to form the aspidosperma alkaloid (−)-tabersonine (**5**; Supplementary Fig. 2).

We recently demonstrated the enzymatic formation of both 4 and 5 by incubating the intermediate precondylocarpine acetate (**1**) with a reductase (DPAS) and either catharanthine synthase (CS) or tabersonine synthase (TS) from the plant *Catharanthus roseus* (Fig. 1; ref. [2]). Furthermore, we identified a reductase (dihydroprecondylocarpine acetate synthase, DPAS) and cyclase (coronaridine synthase, CorS) pair (from the plant *Tabernanthe iboga*) that react with precondylocarpine acetate to form the iboga alkaloid (−)-coronaridine (**6**), which is similar in structure to **4**, but has the opposite configuration as well as a reduced oxidation state (Fig. 1; ref. [6]). These three cyclases have high amino acid identity (Supplementary Fig. 3), yet distinct product profiles, providing a system in which to probe the basis for cycloaddition product selectivity.

We first identified the substrate of CS, TS and CorS, enzymes that had been previously characterized only in coupled assays with DPAS. Neither dihydroprecondylocarpine acetate (**2**) or dehydrosecodine (**3a**, **3b**) have been isolated or characterized due to their instability[3–5,7,8]. When **1** is reacted with DPAS, a compound at $m/z = 337.19$ is observed but rapidly decomposes into a variety of products (Supplementary Figs. 4 and 5). After extensive optimization, the enzyme product was isolated and shown to be an intramolecularly protected form of dehydrosecodine that we colloquially named 'angryline' (**3c**) due to its high reactivity (Fig. 1 and Supplementary Note). CD spectral analysis indicated that **3c** is racemic, suggesting that equilibration between achiral dehydrosecodine and chiral **3c** is nonenzymatic (Supplementary Fig. 6). **3c** was incubated with CS or TS to yield **4** or **5**, respectively (Supplementary Figs. 7 and 8) and with CorS and DPAS (*T. iboga*) to yield **6** (ref. [6]). The isolation of **3c** suggests that desacetoxylation of **2** to form **3a** occurs either in solution, through a Grob-type fragmentation[9], or in the active site of DPAS, and that the sole function of CS, TS and CorS is to catalyze cyclization.

Isolated **3c** did not undergo cyclization in the absence of a cyclase under optimal reaction conditions, indicating that these cycloaddition reactions are enzymatically catalyzed (Supplementary Fig. 8). **3c** is stable in mild acidic conditions, but under neutral/alkaline pH can undergo imine-enamine tautomerization (Supplementary Fig. 9) which, in the absence of a cyclase, leads to degradation (Supplementary Fig. 5 and 8). The relatively slow equilibration of **3c** to **3b** did not allow for the accurate measurement of kinetic constants.

We next solved the crystal structures (Supplementary Table 1) of CS (2.2-Å resolution, Protein Data Bank code 6RT8), TS (1.3-Å resolution, 6RS4) and CorS (1.4-Å resolution, 6RJ8) using molecular replacement (2O7R). These enzymes, which are all carboxylesterase homologs, were structurally similar (r.m.s. deviation of the dimers, 1.045–1.188 Å), showing an α/β-hydrolase fold (Supplementary Fig. 10). The hydrogen-bonding network of the canonical carboxylesterase catalytic triad Ser/Cys–His–Asp[10,11] is disrupted, most notably by replacement of His with Tyr in CS and TS, and with Phe in CorS (Fig. 2 and Supplementary Fig. 11). Moreover, the characteristic Gly–Gly oxyanion hole is replaced by Ala–Gly in each of the cyclases (Supplementary Fig. 12). CS and TS fail to turn over model substrates that are typically hydrolyzed by this class of enzymes (4-methylumbelliferyl acetate and 4-methylumbelliferyl butyrate)[11], suggesting that ancestral esterase/hydrolase activity has been lost (Supplementary Fig. 13). The attempt to reintroduce the hydrolysis function in TS by Y297H mutation failed, consistent with the observation that the entire active site is distorted compared to typical carboxylesterases (Supplementary Figs. 11 and 13).

We attempted to cocrystallize CS, TS and CorS with a range of ligands. Electron density corresponding to a biologically relevant ligand was only observed in a structure of CS cocrystallized with **4**. The omit map suggested that the ligand was not **4**, but 16-carbomethoxycleaviminium (**7**; Fig. 2 and Supplementary Figs. 14 and 15a). **7** has been indirectly shown to be formed under acidic conditions

[1]Max Planck Institute of Chemical Ecology, Department of Natural Product Biosynthesis, Jena, Germany. [2]Leibniz University Hannover, Centre for Biomolecular Drug Research, Hannover, Germany. [3]John Innes Centre, Department of Biological Chemistry, Norwich Research Park, Norwich, UK. [4]Laboratorio de Ciencias Quimicas-UENF-Campos dos Goytacazes-RJ, Campos dos Goytacazes, Brazil. [5]These authors contributed equally: Lorenzo Caputi, Jakob Franke. ✉e-mail: david.lawson@jic.ac.uk; oconnor@ice.mpg.de

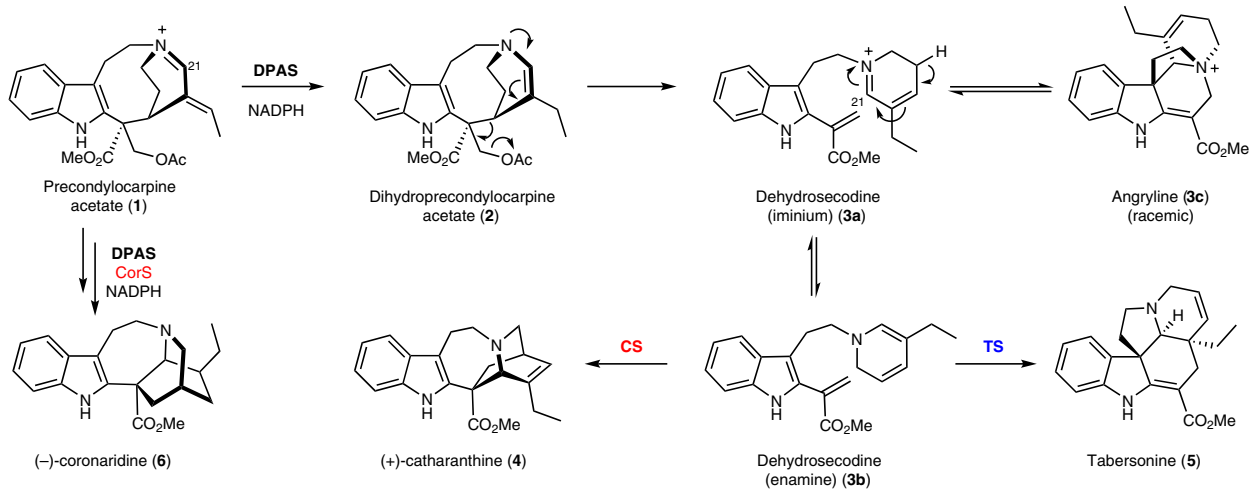

Fig. 1 | Biosynthesis of aspidosperma and iboga alkaloids. Biosynthesis begins with reduction of precondylocarpine acetate (**1**) to generate dehydrosecodine (**3b**), which can undergo one of two formal Diels–Alder reactions to form either (+)-catharanthine (**4**) or (−)-tabersonine (**5**). The biosynthesis of (−)-coronaridine **6** also starts from **1** but involves an additional reduction step.

via a retro-Mannich opening of **4** by the observation of 16-carbomethoxycleavamine (**8**) after chemical reduction[12,13]. A compound with the same mass and retention time was formed upon extended incubation of **4** with CS (Supplementary Fig. 16), indicating that CS is capable of catalysing this retro-Mannich reaction. This indirectly suggests that the forward reaction may also proceed via an ionic stepwise mechanism rather than a concerted Diels–Alder mechanism (Supplementary Fig. 15b). The negatively charged carboxyl group of a retro-Mannich intermediate (Supplementary Fig. 15b) could be stabilized by the oxyanion hole (Ala82–Gly83).

TS, in contrast, did not catalyze any detectable reaction with **5**, suggesting that the mechanisms of CS and TS are fundamentally different. We docked **5** into the active site of TS (Supplementary Fig. 12b). The enzyme active site may position the carbonyl of the methyl acrylate (Fig. 2) in plane with the indole, allowing the vinyl indole to function as the diene in a Diels–Alder mechanism. Additionally, due to a missing loop, the TS substrate is more exposed to the bulk water (Supplementary Fig. 17), which could facilitate protonation of the dihydropyridine amine. While **5** formation could also proceed via a stepwise mechanism, this protonated amine could serve as an electron withdrawing group to activate the dienophile in a Diels–Alder reaction (Supplementary Fig. 15c).

CorS reacts with **3c** to form an unstable product that is reduced by DPAS to form **6** (ref. [6]). We hypothesize that CorS converts **3c** to (−)-coronaridine iminium (**9**) through an aza-Diels–Alder reaction or a stepwise mechanism[14] (Supplementary Fig. 15d), although this proposed iminium is too unstable to isolate. Furthermore, while **4** and **5** appear to form directly from dehydrosecodine enamine **3b**, reaction in D_2O suggests that the formation of **6** requires tautomerization of the dihydropyridine moiety of **3b** to form 1,4-dehydrosecodine **3d** and 1,5-dehydrosecodine **3e** (ref. [6]; Supplementary Fig. 15d). Therefore, CorS likely promotes tautomerization of dehydrosecodine in addition to cyclization. Thus, the switch in enantioselectivity between CS and CorS is caused by tautomerization to a different isomer of dehydrosecodine (**3e**), which would cyclize via a different mechanism. Docking of **6** in CorS (Fig. 2, Supplementary Fig. 12c) demonstrates that this active site is shaped to accommodate the (−)-iboga scaffold.

This study provides a structural basis for design and assay of 47 mutants (Supplementary Figs. 12, 14 and 17). Five TS mutants (S170A, H78A, Y216D, Y216F, helix5 swap), five CS mutants (loop trim, loop trim + helix5 swap, loop trim + F218Y, helix5 swap, D273A) and three CorS mutants (S174A, S174C, P175T) showed qualitative changes in product profile (Supplementary Fig. 18–24 and Supplementary Table 2), providing the first clues to cyclization specificity.

The substrate conformation determines the cyclization selectivity between iboga and aspidosperma scaffolds, and the primary role of these enzymes is most likely to conformationally constrain dehydrosecodine. Most importantly, the methyl acrylate of dehydrosecodine must be in plane with the indole to form the aspidosperma scaffold, while these two moieties must be perpendicular to form the iboga scaffolds (Supplementary Fig. 15). Mutations of residues that appear to interact with the methyl acrylate or the indole had different effects in the three enzymes. Mutations S172A and S172C in CS, a residue that is close to the indole nitrogen and the ester group of dehydrosecodine, resulted in inactive enzyme. However, mutation S170C in TS did not affect activity, although S170A formed small amounts of **4**. When the same mutations were introduced into CorS, S174C produced **5** but no **9**, and CorS S174A produced a mixture of **9** and **4** (Supplementary Figs. 20, 22 and 24). Mutation of Pro175 in CorS, a residue that appears to contact the ester group of dehydrosecodine, resulted in the formation of all three cyclization products, while mutation of the corresponding residue in CS (Pro173) and in TS (Thr171) only led to a decrease of activity.

The residues of helix α5 also appear to interact with the methyl acrylate and dihydropyridine moieties, primarily through van der Waals interactions. Thus, this region of the active site may also be responsible for orienting the substrate. Swapping helix α5 (Asp217–Cys224) of CS with the corresponding TS sequence yielded a CS mutant that formed **5** (Supplementary Fig. 20). When this grafting was combined with the removal of the extended loop, we observed formation of small amounts of **6** in the presence of DPAS (Supplementary Fig. 25), suggesting that these residues may play a role in the tautomerization of the dihydropyridine ring required for **6** formation. However, the same swap in CorS led to a mutant that still formed small amounts of **6**, indicating that other residues are required for this function. When helix α5 from CS was introduced into TS, the enzyme lost **5**-forming activity and produced small amounts of an additional unknown product. Swapping the helix from TS into CorS resulted only in loss of activity. Finally, numerous other residues in the binding pocket were mutated; mutation of Tyr216 and His78 in TS (Supplementary Fig. 22), and of Asp273 in CS (Supplementary Fig. 20) led to changes in product profiles in the respective enzymes, but these mutations had minimal qualitative effects in the other enzymes.

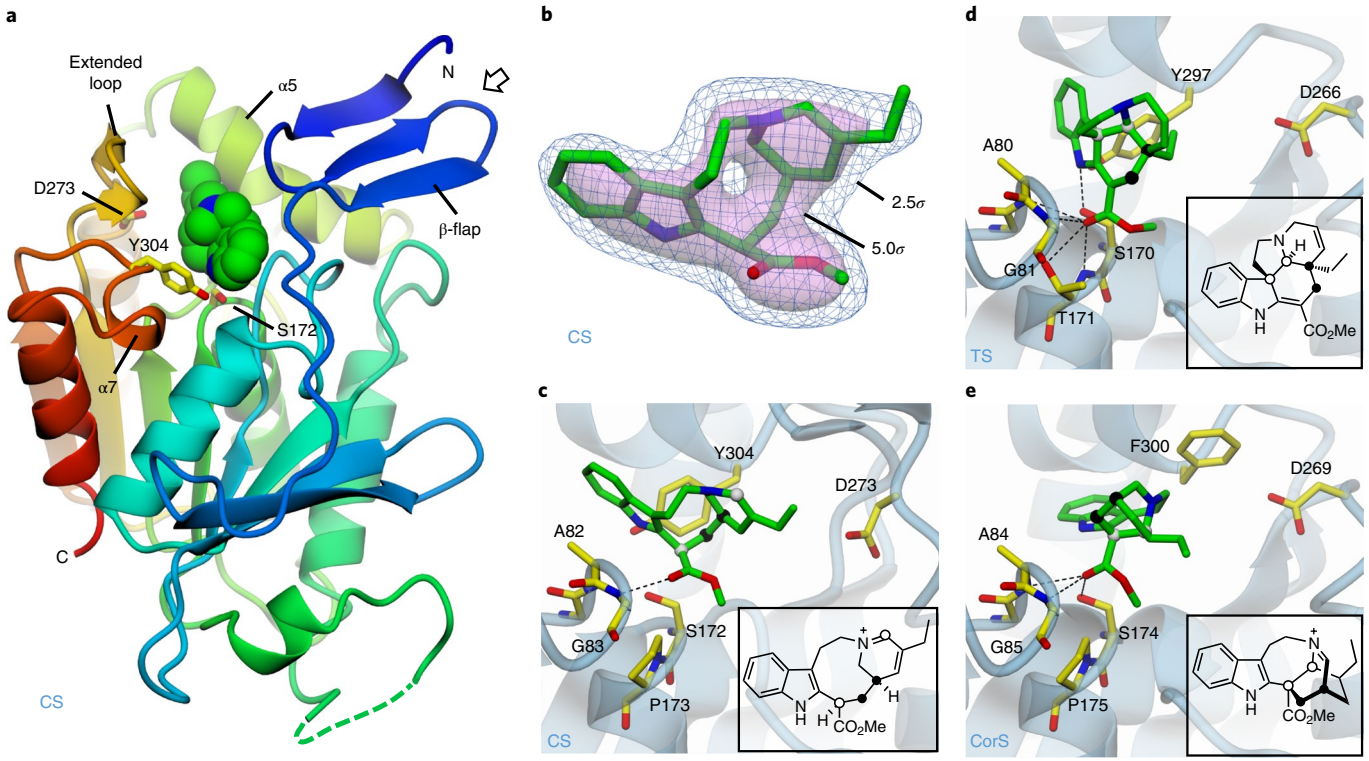

Fig. 2 | Crystal structures of CS, TS and CorS. a, A single CS subunit depicted in a cartoon representation with rainbow coloration from blue at the N terminus through to red at the C terminus. Also shown as van der Waals spheres with green carbons is the 16-carbomethoxycleaviminium (**7**) intermediate bound in the active site; the catalytic triad residues are in stick mode with yellow carbons. The sides of the active site cavity are delineated largely by helices α5 and α7, the N-terminal β-flap and, unique to CS, the extended loop; the base of the cavity is formed by the 'nucleophilic elbow' which bears the catalytic Ser, and the loop providing the oxyanion hole residues. The open arrow shows the approximate direction of view for panels **b**–**e**. **b**, Omit electron density at 2.2-Å resolution for the cleaviminium intermediate **7** bound to CS shown at two contour levels: 2.5σ in a blue mesh and 5.0σ as a semi-transparent pink surface. **c**, Detail of the CS active site, where residues from the catalytic triad and oxyanion hole are displayed with yellow carbons and the backbone trace is depicted as a semi-transparent pale blue cartoon. The cleaviminium intermediate **7** is shown with green carbons and is also depicted as a two-dimensional representation in the inset. In both representations, the two C–C bonds that form as a result of catalysis are indicated by spheres of the same color for each bond (that is, black or white). **d**. Detail of TS active site with docked **5** displayed as for CS in **c**. **e**. Detail of CorS active site with docked (−)-coronaridine iminium (**9**) displayed as for CS in **c**. In **c**–**e**, only hydrogen bonds between ligands and the active site residues are shown (dashed lines).

Given that no single mutation led to a complete switch in product selectivity, and since identical point mutations had different effects in the context of the three enzymes, all of the residues of the binding pockets likely work in concert to exert changes to the substrate conformation. Nevertheless, this study clearly demonstrates the plasticity of these enzymes with regard to product profile. More extensive analysis of CS, TS and CorS homologs found in other Apocynaceae plants, along with combinatorial mutagenesis, may be required to achieve complete changes in product selectivity. Although not characterized here, low levels of side products were observed (Supplementary Figs. 20, 22 and 25), suggesting that future enzyme engineering efforts could enable unnatural iboga and aspidosperma scaffolds to be accessed by additional variations of a cycloaddition reaction of dehydrosecodine isomers.

Here we report the enzymatic production and characterization of **3c**, a stabilized form of **3a** and **3b**, an alkaloid cycloaddition substrate proposed 60 years ago[3]. When reacted with CS, TS and CorS, this substrate yields (+)-iboga, (−)-aspidosperma and (−)-iboga scaffolds. Structural data suggest that these enzymes control substrate conformation, namely the orientation of the methyl acrylate relative to the indole, to yield these different scaffolds. Tautomerization of the dihydropyridine ring, mediated in part by helix α5 of CorS, allows an additional cyclization mode. These enzymes generate chemical diversity from a single substrate via different cycloaddition reactions[15–18] and provide a foundation for enzyme engineering for catalysis of additional alkaloids.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-019-0460-x.

Received: 19 July 2019; Accepted: 20 December 2019;
Published online: 17 February 2020

References

1. O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23**, 532–472 (2006).
2. Caputi, L. et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **60**, 1235–1239 (2018).
3. Wenkert, E. Biosynthesis of indole alkaloids. The aspidosperma and iboga bases. *J. Am. Chem. Soc.* **84**, 98–102 (1962).
4. Scott, A. I., Cherry, P. C. & Qureshi, A. A. Mechanisms of indole alkaloid biosynthesis. The Corynanthe-Strychnos relationship. *J. Am. Chem. Soc.* **91**, 4932–4936 (1969).
5. Kutney, J. P., Ehret, C., Nelson, V. R. & Wigfield, D. C. Studies on indole alkaloid biosynthesis. II. *J. Am. Chem. Soc.* **90**, 5929–5930 (1968).

6. Farrow, S. C. et al. Biosynthesis of an anti-addiction agent from the iboga plant. *JACS* **141**, 12979–12983 (2019).
7. Scott, A. I. et al. Biosynthesis of the indole alkaloids. *Acc. Chem. Res.* **3**, 151–157 (1970).
8. Qureshi, A. A. & Scott, A. I. Biogenetic-type synthesis of Iboga alkaloids: (±)-catharanthine. *J. Chem. Soc. Chem. Commun.* **16**, 947–948 (1968).
9. Kuehne, M. E., Roland, D. M. & Haften, R. J. Studies in biomimetic alkaloid syntheses. 2. Synthesis of vincadifformine from tetrahydro-beta-carboline through a secodine intermediate. *Org. Chem.* **43**, 3705–3710 (1978).
10. Marshall, S. D., Putterill, J. J., Plummer, K. M. & Newcomb, R. D. The carboxylesterase gene family from *Arabidopsis thaliana*. *J. Mol. Evol.* **57**, 487–500 (2003).
11. Ileperuma, N. R. et al. High-resolution crystal structure of plant carboxylesterase AeCXE1, from *Actinidia eriantha*, and its complex with a high-affinity inhibitor paraoxon. *Biochemistry* **46**, 1851–1859 (2007).
12. Andriamialisoa, R. Z., Langlois, N. & Langlois, Y. Preparation of 15-oxo-16-methoxycarbonyl-15, 20-dihydrocleavamine and coupling reaction with vindoline. *Heterocycles* **15**, 245–250 (1981).
13. Langlois, N., Guéritte, F., Langlois, Y. & Potier, P. Application of a modification of the Polonovski reaction to the synthesis of vinblastine-type alkaloids. *J. Am. Chem. Soc.* **98**, 7017–7024 (1976).
14. Buonora, P., Olsen, J.-C. & Oh, T. Recent developments in imino Diels–Alder reactions. *Tetrahedron* **57**, 6099–6138 (2001).
15. Lichman, B. R., O'Connor, S. E. & Kries, H. Biocatalytic strategies towards [4+2] cycloadditions. *Eur. J. Chem.* **25**, 6864–6877 (2019).
16. Tan, D. et al. Genome-mined diels–alderase catalyzes formation of the *cis*-octahydrodecalins of varicidin A and B. *J. Am. Chem. Soc.* **141**, 769–773 (2019).
17. Jeon, B. S. et al. Investigation of the mechanism of the SpnF-catalyzed [4+2]-cycloaddition reaction in the biosynthesis of spinosyn A. *Proc. Natl Acad. Sci.* **114**, 10408–10413 (2017).
18. Zhang, B. et al. Enzyme-catalysed [6+4] cycloadditions in the biosynthesis of natural products. *Nature* **568**, 122–126 (2019).

Methods

Materials and molecular biology kits. All solvents used in this study were either of HPLC or MS grade, depending on the application. All were purchased from Fisher Scientific. **4** was purchased from Sigma Aldrich, while **5** was obtained from Ava Chem Scientific. Stemmadenine acetate was synthesized as described previously[2]. NADPH was from Roche. Carbenicillin was from Formedium. All genes and fragment amplifications were performed using Platinum Superfi polymerase (Thermo Fisher) while colony PCRs were performed using Phire II master mix (Thermo Fisher). PCR product purifications were performed using the Macherey-Nagel PCR clean-up kit. Plasmids purifications were performed using Promega Wizard minipreps. Sequencing of the clones was performed by Eurofins Genomics (Germany).

Cloning and mutagenesis. Cloning of CrPAS (MH213134), CrDPAS (KU865331), CS (MF770512) and TS (MF770513) was reported in Caputi et al.[2]. Cloning of DPAS and CorS from *T. iboga* was reported in Farrow et al.[6]. CS, TS and CorS mutants were generated by overlap extension PCR. The codon(s) to be mutated was selected and two primers, one reverse and one forward (Supplementary Table 3), were designed to overlap and introduce the mutation. PCR products were gel purified, ligated into pOPINF expression vector[19] using the In-Fusion cloning kit (Clontech Takara) and transformed into competent *Escherichia coli* Stellar cells (Clontech Takara) according to the manufacturer's instructions. Mutant constructs were sequenced to verify the mutant gene sequence and correct insertion.

Protein expression and purification. CrPAS and CrDPAS were expressed and purified as reported previously[2]. All other proteins and mutants reported in this study were expressed in SoluBL21 *E. coli* cells (Amsbio). Chemically competent cells were transformed by heat shock at 42 °C. Transformed cells were selected on Luria-Bertani agar plates supplemented with carbenicillin (100 μg ml^{-1}).

For large-scale expression of proteins for crystallography studies, single colonies were used to inoculate starter cultures in 50 ml of 2× YT medium supplemented with carbenicillin (100 μg ml^{-1}) that were grown overnight at 37 °C. Starter culture (10 ml) was used to inoculate 1 l of 2× YT medium containing the antibiotic. The cultures were incubated at 37 °C until optical density OD_{600} reached 0.6 and then transferred to 18 °C for 30 min before induction of protein expression by addition of IPTG (0.2 mM). Protein expression was carried out for 16 h. Cells were harvested by centrifugation and resuspended in 50 ml of buffer A (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole,) with EDTA-free protease inhibitors (Roche Diagnostics). Cells were lysed by sonication for 4 min on ice. Cell debris was pelleted by centrifugation at 35,000*g* for 20 min. His_6-tagged enzymes were purified on an AKTA Pure system (GE Healthcare) using a HisTrap HP 5-ml column (GE Healthcare) equilibrated with buffer A. Samples were loaded at a flow rate of 2 ml min^{-1} and step-eluted using buffer B (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole). Eluted proteins were subjected to further purification on a Superdex Hiload 16/60 S200 gel filtration column (GE Healthcare) at a flow rate of 1 ml min^{-1} using buffer C (20 mM HEPES pH 7.5, 150 mM NaCl) and collected in 1.5 ml fractions. Fractions containing the protein of interest were collected and treated with His_6-tagged 3C protease overnight at 4 °C to cleave the His_6-tag. After filtration through low protein binding filters, the protein samples were passed through a HisTrap HP 1-ml column (GE Healthcare) equilibrated with buffer C to remove the protease and the uncut protein. The proteins were concentrated using centrifugal filters.

For small-scale expression of proteins for enzymatic activity, single colonies were used to inoculate starter cultures in 10 ml of 2× YT medium supplemented with carbenicillin (100 μg ml^{-1}) that were grown overnight at 37 °C. Starter culture (1 ml) was used to inoculate 100 ml of 2× YT medium containing the antibiotic. The cultures were incubated at 37 °C until OD_{600} reached 0.6 and then transferred to 18 °C for 30 min before induction of protein expression by addition of IPTG (0.2 mM). Protein expression was carried out for 16 h. Cells were harvested by centrifugation and resuspended in 10 ml of buffer A with EDTA-free protease inhibitors (Roche Diagnostics). Cells were lysed by sonication for 1 min on ice. Cell debris was pelleted by centrifugation at 35,000*g* for 20 min. Then 250 μl of Ni-nitrilotriacetic acid (NTA) agarose (Qiagen) were added to the lysates and incubated for 1 h at 4 °C on a shaker. The Ni-NTA beads were harvested by centrifugation at 1,000*g* for 1 min and washed three times with buffer A. Proteins were eluted by addition of 2 × 300 μl of buffer B. After filtration through low protein binding filters, the proteins were dialyzed in buffer C using centrifugal filters.

Protein crystallization. Crystallization screens were conducted by sitting-drop vapor diffusion in MRC2 96-well crystallization plates (Swissci) with a mixture of 0.3 μl of well solution from the PEGs (Qiagen) and PACT (Qiagen) screens and 0.3 μl of protein solution. Protein concentrations were adjusted to 20–30 mg ml^{-1} for all proteins. Ligands (**4**, **5** and stemmadenine acetate) were prepared at a concentration of 30 mM in methanol and added to a final concentration of 1 mM (1 μl in 30 μl of protein solution). Solutions were dispensed by an Oryx8 robot (Douglas Instruments).

Crystals were obtained in several screening conditions and tested for diffraction on the beamline. The best diffracting CS crystals were obtained from 0.2 M Na thiocyanate and 20% (w/v) PEG3350; the best TS crystals were obtained from 25% (w/v) PEG 1500 in 0.1 M MIB buffer pH 4.0 (where MIB is malonate dibasic monohydrate, imidazole and boric acid; Molecular Dimensions); and the best CorS crystals were obtained from 25% (w/v) PEG 1500 in 0.1 M SPG buffer pH 6.0 (where SPG is succinic acid, sodium dihydrogen phosphate and glycine; Molecular Dimensions).

All crystals were cryoprotected by soaking in crystallization solution containing 25% (v/v) ethylene glycol before flash cooling in liquid nitrogen.

X-ray data collection, processing and structure solution. All X-ray data were recorded on beamline I03 at the Diamond Light Source (Oxfordshire, UK) at a wavelength of either 0.976 Å (CS and CorS) or 0.980 Å (TS) using a Pilatus3 6 M hybrid photon counting detector (Dectris) with the crystal maintained at 100 K by a Cryojet cryocooler (Oxford Instruments). Diffraction data were integrated and scaled using either DIALS[20] or XDS[21] via the XIA2 expert system[22] then merged using AIMLESS[23]. Data collection statistics are summarized in Supplementary Table 1.

The majority of the downstream analysis was performed through the CCP4i2 graphical user interface[24]. Both TS and CorS were solved by molecular replacement using the structure of *Actinidia eriantha* carboxylesterase (Protein Data Bank entry 2O7R)[11] as a template, with which they share 29% and 32% sequence identity, respectively. Templates were prepared with reference to the appropriate sequences using SCULPTOR[25] and then all side chains were truncated to Cβ.The structures were solved using PHASER[26] which located two copies of the template in the $P2_1$ asymmetric unit of TS (with a corresponding solvent content of 53%), and a single copy of the template in the $P6_522$ asymmetric unit of CorS (with a corresponding solvent content of 63%). After editing these solutions to remove poorly fitting regions in COOT[27] and subsequent refinement with REFMAC5[28], the resultant models were completely rebuilt with BUCCANEER[29] to give much-improved models. These models were finalized through several iterations of manual editing in COOT and further refinement with REFMAC5. Anisotropic temperature factors were used for the latter in both cases.

The CS structure was solved in a similar fashion, but instead starting from a template derived from TS with which it shares 78% sequence identity. PHASER was successful in finding eight copies of the subunit template in the asymmetric unit, arranged as four homodimers (with a corresponding solvent content of 53%). From the initial electron density maps there was clear evidence for a bound ligand in each of the eight CS active sites. However, it could not be reconciled with the structure of **4**, which had been added to the crystallization. It was only after the rest of the model was essentially complete that a cleavaminium intermediate could be confidently assigned to the density. For the final refinement cycles, TLS refinement was used with a single TLS domain defined for each of the eight CS subunits.

The statistics of the final refined models are shown in Supplementary Table 4. The Ramachandran statistics (favored/allowed/outlier expressed as percentages), evaluated using MolProbity[30], were as follows: CS, 95.2/4.2/0.6; TS, 98.2/1.5/0.3; CorS, 96.1/3.3/0.6. All structural figures were prepared using CCP4mg[31].

Docking simulations. Ligands were docked into the active sites of the TS and CorS crystal structures using AutoDock Vina[32]. Ligand coordinate files were generated using the Lidia function within COOT[27]. To prevent the appearance of extended, nonproductive, conformations for the noncyclized ligands, the three rotatable bonds linking the indole to the dihydropyridine ring were manually adjusted to place the latter alongside the methyl acrylate and then they were fixed. Otherwise all the remaining rotatable bonds were allowed to rotate during the simulations. Both the protein and ligand coordinates were prepared for the docking calculations using AutoDockTools[33]. The side chains of several residues lining the active site pocket were allowed to rotate during the simulations, which are listed below for each protein. For the final simulations, an exhaustiveness value of 128 was used and, given the largely hydrophobic character of the active site pockets, the hydrogen-bonding strength was increased (with weight_hydrogen value of −1.2). Otherwise the default AutoDock Vina parameters were used. In assessing the results, we reasoned that the indole moiety would probably occupy roughly the same region of the active site pocket throughout the catalytic cycle and that the carbonyl oxygen of the methyl acrylate would most likely interact with the oxyanion hole, and possibly the catalytic Ser. Moreover, we favored results that were broadly consistent in placement and orientation with the **7** seen in the CS crystal structure. We thus selected plausible poses that satisfied these criteria for each of the ligands, which were not always the lowest energy solutions. These are displayed in Supplementary Fig. 10.

For TS, the best results were obtained using the protein chain B, possibly because it adopts a slightly more open conformation. During the simulations, the side chains of Tyr14, Ser170, Thr171, Tyr216, Tyr224 and Tyr297 were allowed to rotate, and a search space of 20 × 20 × 20 Å^3 encompassing the active site cavity was used.

For CorS the side chains of Tyr18, Ser174, Tyr205, Tyr219, Tyr224, Tyr227, Phe300, Phe301 and Phe304 were allowed to rotate, and a search space of 22 × 22 × 20 Å^3 encompassing the active site cavity was used. However, no suitable docking poses were obtained. Closer inspection of these in the context of a molecular surface indicated that the side chain of Phe300 was mostly directed into

the active site and likely to sterically inhibit ligand binding. Thus, for subsequent runs, Phe300 was manually rotated out of the active site and fixed.

NMR and CD spectroscopy. NMR spectra (one- and two-dimensional NMR) were acquired using a Bruker Neo 600 MHz NMR spectrometer equipped with a TCI cryoprobe. The solvent residual peaks of CD_3OD (with chemical shifts (δ) of 3.31 and 49.0, respectively), CD_3CN (with δ of 1.94 and 1.32, respectively) and $CDCl_3$ (with δ of 7.26 and 77.16, respectively) were used as internal standards in ^1H and ^{13}C NMR. The number of scans depended on sample concentration and are indicated in the supplementary figures and tables accordingly.

CD spectra were recorded in 1-nm steps with a 0.5 s averaging time on a Chirascan Plus spectropolarimeter (Applied Photophysics) at 20 °C in a 1-mm cuvette. Measurements were collected in triplicates, averaged and background subtracted with 100 mM MES buffer, pH 6.0.

Activity assays. Activity assays of CS and TS for pH optimum were performed in the following buffers, all at a concentration of 100 mM: MES (pH 5.5, pH 6.0 and pH 6.5), HEPES (pH 7.0 and pH 7.5), Tris-HCl (pH 8.0 and pH 8.5) and CHES (pH 9.0, pH 9.5 and pH 10.0). The total volume of the reactions was 100 µl and they contained 100 nM enzyme and 1 µM **3c**. Reactions were started by addition of the substrate and incubated at 37 °C for 20 min. Then 10 µl of the reaction mixtures were collected at time 0 and after 20 min and quenched in 90 µl of 90:9:1 MeOH:H_2O:FA.

Activity assays of the mutant proteins were performed in 100 mM Tris-HCl buffer, pH 8.5. Total volume of the reactions was 100 µl. Reactions contained 100 nM enzyme and 1 µM **3c**. Reactions involving the coupling of CorS and DPAS were performed as described in Farrow et al.[6]. Assays were performed at 37 °C for 20 min. Then 10 µl of the reaction mixtures were collected at time 0 and after 20 min and quenched in 90 µl of 90:9:1 MeOH:H_2O:FA. All reactions were performed in triplicates.

UPLC/MS and UPLC/MS–MS methods. For high-resolution MS analysis and MS–MS studies, **3c** was infused at 5–10 µl min^{-1} using a Harvard Apparatus syringe pump onto a Synapt G2 HDMS mass spectrometer (Waters) calibrated using a sodium formate solution. Samples were analyzed for 1 min with a scan time of 1 s in the mass range of 50–1,200 *m/z*. Capillary voltage was 3.5 kV, cone voltage was 40 V, source temperature was 120 °C, desolvation temperature was 350 °C and desolvation gas flow was 800 l h^{-1}. Leu-enkephaline peptide (1 ng $µl^{-1}$) was used to generate a dual lock-mass calibration with $[M+H]^+ = 556.2766$ and $m/z = 278.1135$ measured every 10 s. Spectra were generated in MassLynx 4.1 by combining a number of scans and peaks, and were centered using automatic peak detection with lock-mass correction.

UPLC/QqQ–MS analysis was carried out on a UPLC (Waters) equipped with an Acquity BEH C18 1.7 µm (2.1 × 50 mm^2) column connected to Xevo TQS triple quadrupole (Waters). Chromatographic separation was performed using 0.1% FA as mobile phase A and methanol as mobile phase B. A linear gradient from 30% to 35% B in 3 min was performed for separation of the compounds followed by 0.5 min isocratic at 35%. The column was then re-equilibrated at 30% B for 1.5 min. The column was kept at 35 °C throughout the analysis and the flow rate was 0.6 ml min^{-1}. MS detection was performed in positive ESI. Capillary voltage was 3.0 kV, the source was kept at 150 °C, desolvation temperature was 500 °C, the cone gas flow was 50 l h^{-1} and desolvation gas flow was 800 l h^{-1}. Unit resolution was applied to each quadrupole. The multiple reaction monitoring (MRM) transitions used to monitor the elution of the alkaloids of interest are reported in Supplementary Table 4.

Formation of 16-carbomethoxycleavamine from catharanthine. Chemical synthesis of **8** from **4** was performed using a modified method derived from Andriamialisoa et al.[12]. Then 2.5 mg of **4** were dissolved in 100 µl of TFA and stirred at room temperature for 3 h. Formation of **7** intermediate ($m/z = 337.19$) was monitored every hour by UPLC/QqQ–MS analysis. Then 1-µl aliquotes of the reaction mixture were diluted to 1 ml with 90:9:1 MeOH:H_2O:FA and injected. Before appearance of the by-products, the iminium intermediate was reduced by dropwise addition of an excess of sodium cyanoborohydride in MeOH (5 mg in 200 µl of MeOH). After stirring at room temperature for 30 min the final products, **8** isomers ($m/z = 339.19$), were analyzed by UPLC/QqQ–MS.

Enzymatic synthesis of **8** from **4** using CS as catalyst was performed in 50 mM HEPES buffer, pH 7.5. Then 3 mg of **4** were dissolved in 20 µl of DMSO and added to the reaction together with 1.4 mg of enzyme. Reactions were incubated at 37 °C for 3 h, and the progress of the reaction and formation of **7** intermediate was monitored by UPLC/QqQ–MS analysis. When the reactions had generated enough product, the enzyme was removed by solid phase extraction (SPE) on an OASIS BHL cartridge (30 mg) and the product was eluted with 200 µl of 90:9:1 MeOH:H_2O:FA. Then 500 µl of material were reacted with an excess of sodium cyanoborohydride in MeOH (1 mg in 200 µl of MeOH). After stirring at room temperature for 30 min, the final products were analyzed by UPLC/QqQ–MS.

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

Data availability

Structures have been deposited at the Protein Data Bank under the accession codes 6RJ8 (coronaridine synthase), 6RS4 (tabersonine synthase) and 6RT8 (catharanthine synthase).

References

19. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
20. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D. Biol. Crystallogr.* **74**, 85–97 (2018).
21. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
22. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
23. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–1214 (2013).
24. Potterton, L. et al. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D. Biol.* **74**, 68–84 (2018).
25. Bunkoczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 303–312 (2011).
26. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240–255 (1997).
29. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 1002–1011 (2006).
30. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nuc. Acids Res.* **35**, W375–W383 (2007).
31. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 386–394 (2011).
32. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
33. Morris, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).

Acknowledgements

S.E.O. acknowledges ERC (788301). J.F. acknowledges financial support by the SMART BIOTECS alliance between the Technische Universität Braunschweig and the Leibniz Universität Hannover, supported by the Ministry for Science and Culture (MWK) of Lower Saxony, Germany. We acknowledge Diamond Light Source for access to beamline I03 under proposal MX13467 with support from the European Community's Seventh Framework Program (No. FP7/2007–2013) under Grant Agreement No. 283570 (BioStruct-X).

Author contributions

L.C. and S.E.O. conceived the project. D.M.L. managed all crystallography experiments. D.M.L., L.C., S.C.F. and C.E.M.S solved the crystal structures. L.C. and S.C.F. performed all biochemical experiments. K.B. and I.J.C.V. isolated substrates and products. J.F. solved the structure of the enzymatic substrates and developed the enzymatic mechanisms.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-019-0460-x.

Correspondence and requests for materials should be addressed to D.M.L. or S.E.O.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Sarah E. O'Connor, David M. Lawson

Last updated by author(s): Dec 10, 2019

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	For crystallographic data: CCP4i2, XDS, XIA2, AIMLESS, SCULPTOR, PHASER, BUCCANEER, COOT, REFMAC5, MolProbity, CCP4mg, AutoDockVina v1.1.2.
Data analysis	Excel for LCMS data, Clustal X for sequence alignment, ESPript3.0 to generate the alignment figure. Protparam (ExPasy) was used to calculate MW and epsilon for determining protein concentrations.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Structures have been deposited under the accession codes 6RJ8 (coronaridine synthase), 6RS4 (tabersonine synthase) and 6RT8 (catharanthine synthase).

Field-specific reporting

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

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

Life sciences study design

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

Sample size	All enzyme reactions reported in the study were performed and analyzed in triplicates. Due to the high reproducibility of the analysis and to the high accuracy of the instrumentation used, and extensive literature precedent for how enzyme assays are performed, we reckoned this number was sufficient to provide the qualitative and quantitative data reported in the manuscript.
Data exclusions	No data were excluded from the analyses.
Replication	All attempts at replication were successful.
Randomization	All samples for enzyme activity were analysed in no particular order
Blinding	No blinding was required for enzyme assays or structural analysis.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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