UCLA
UCLA Previously Published Works

Title

Catalytic mechanism and endo-to-exo selectivity reversion of an octalin-forming natural Diels–Alderase

Permalink

https://escholarship.org/uc/item/18q0h26x

Journal

Nature Catalysis, 4(3)

ISSN

2520-1158

Authors

Sato, Michio
Kishimoto, Shinji
Yokoyama, Mamoru
et al.

Publication Date

2021-03-01

DOI

10.1038/s41929-021-00577-2

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Catal. Author manuscript; available in PMC 2022 July 21.

Published in final edited form as:

Nat Catal. 2021 March ; 4(3): 223–232. doi:10.1038/s41929-021-00577-2.

Catalytic mechanism and *endo*-to-*exo* selectivity reversion of an octalin-forming natural Diels–Alderase

Michio Sato[1], **Shinji Kishimoto**[1], **Mamoru Yokoyama**[1], **Cooper S. Jamieson**[2], **Kazuto Narita**[1], **Naoya Maeda**[1], **Kodai Hara**[1], **Hiroshi Hashimoto**[1], **Yuta Tsunematsu**[1], **Kendall N. Houk**[2,3], **Yi Tang**[2,3], **Kenji Watanabe**[1,✉]

[1]Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan.

[2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA.

[3]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, USA.

Abstract

We have previously reported the identification of CghA, a proposed Diels–Alderase responsible for the formation of the bicyclic octalin core of the fungal secondary metabolite Sch210972. Here we show the crystal structure of the CghA–product complex at a resolution of 2.0 Å. Our result provides the second structural determination of eukaryotic Diels–Alderases and adds yet another fold to the family of proteins reported to catalyse [4 + 2] cycloaddition reactions. Site-directed mutagenesis-coupled kinetic characterization and computational analyses allowed us to identify key catalytic residues and propose a possible catalytic mechanism. Most interestingly, we were able to rationally engineer CghA such that the mutant was able to catalyse preferentially the formation of the energetically disfavoured *exo* adduct. This work expands our knowledge and understanding of the emerging and potentially widespread class of natural enzymes capable of catalysing stereoselective Diels–Alder reactions and paves the way towards developing enzymes potentially useful in various bio/synthetic applications.

The Diels–Alder (DA) reaction is of great value in synthesis, yet it was relatively recently that enzymes capable of catalysing DA reactions, that is, Diels–Alderases (DAases), began to surface[1]. For example, the bicyclic octalin (octahydronaphthalene) motifs that are

---

under exclusive licence to Springer Nature Limited 2021**Reprints and permissions information** is available at www.nature.com/reprints.

✉**Correspondence and requests for materials** should be addressed to K.W., kenji55@u-shizuoka-ken.ac.jp.

Author contributions

M.S., H.H. and K.W. conceived and designed the study. S.K. and K.N. chemically synthesized substrate analogues. Y. Tsunematsu and K.W. designed and performed molecular cloning. M.Y. performed the heterologous expression and purification as well as in vitro characterization of the enzymes. N.M. prepared the crystal of the CghA–**1** complex. M.Y. and Y. Tsunematsu prepared and analysed SeMet-substituted CghA. M.S. and K.N.H. elucidated the chemical structures. C.S.J. and K.N.H. performed the computational analysis. K.H. and H.H. performed the crystallographic studies and structural analysis. Y. Tang and K.W. analysed sequence and structure comparison data. All authors analysed and discussed the results. K.W. and K.N.H. prepared the manuscript.

Competing interests

The authors declare no competing interests.

Extended data is available for this paper at https://doi.org/10.1038/s41929-021-00577-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41929-021-00577-2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

widely present in fungal natural products are thought to be formed by enzyme-catalysed intramolecular Diels–Alder (IMDA) reactions[2,3]. Previously, CghA was shown to perform an *endo* IMDA reaction to form the octalin core of Sch210972 (**1**)[4], a potential anti-HIV agent from the filamentous fungus *Chaetomium globosum*[5]. Other octalin-forming stand-alone DAases include Eqx3 (ref.[6]) and Fsa2 (ref.[7]) for equisetin (**2**) biosynthesis and MycB (ref.[8]) for myceliothermophin E (**3**) biosynthesis. More recent studies have uncovered enzymes Phm7, PvhB, IccD and PyiF that catalyse DA reactions for the biosyntheses of phomasetin[9], varicidins[10], ilicicolin H[11] and pyrichalasin H[12], respectively (Fig. 1a, Extended Data Fig. 1). Apart from the above octalin-forming DAases, an oxadecalin-forming DAase, Tsn11, from the tetronasin biosynthetic pathway that was proposed to work in tandem with another DAase, Tsn15, was also reported recently[13]. These enzymes are proposed to catalyse IMDA reactions and to afford the stereospecific synthesis of multicyclic scaffolds[14].

To deepen our general understanding of how a DAase catalyses an IMDA reaction and exerts stereoselective control over the cycloaddition process, we determined the crystal structure of CghA in complex with the *endo* product **1** and performed detailed in vitro assays of CghA and its mutants. We also performed computational analysis of the DA reaction pathways to infer the mechanism of how the DAase CghA exerts diastereoselectivity over the [4 + 2] cycloaddition reaction and propose how it may evade product inhibition. Furthermore, to test our understanding of the mechanism underlying the observed diastereoselectivity CghA imposes on the DA reaction it catalyses, we attempted to engineer CghA by rationally reversing the stereoselectivity by site-directed mutagenesis of the active-site amino acid residues to favour the formation of the disfavoured *exo* adduct.

Results

Structural characterization of CghA.

Our previous characterizations of the *C. globosum cghA*-knockout strain and engineered *Aspergillus nidulans* A1145 strains carrying the *cgh* genes indicated that CghA controls the stereochemical course of the [4 + 2] cycloaddition reaction leading to the sole formation of the *endo* product **1** (Fig. 1b)[4]. To investigate how this potential DAase operates, we determined the crystal structures of CghA at 2.0 Å resolution, both apo and in complex with **1** (Supplementary Table 1). CghA is comprised of orthogonally packed N- and C-terminal β-barrel domains, each assuming a flattened lipocalin-like fold (Fig. 2a). This is a second example of AttH-like superfamily proteins whose structure has been determined[15]. None of the natural DAases that have been structurally characterized to date belong to the AttH-like superfamily, attesting to the wide diversity of DAases. The DAases PyrI4 (ref.[16]), AbyU (ref.[17]) and Tsn11 (ref.[13]) involved in the biosynthesis of tetronate/tetramate-containing compounds share a single-domain lipocalin-type fold, while SpnF (ref.[18]) adopts a characteristic SAM-dependent methyltransferase fold. More recently, IdmH, the DAase for the biosynthesis of the polyketide–nonribosomal peptide natural product indanomycin, was determined to assume the NTF2-like fold[19]. Therefore, our result adds yet another fold to the class of enzymes capable of catalysing [4 + 2] cycloaddition reactions. Unlike typical lipocalin-type β-barrel proteins[20], CghA forms its ligand-binding pocket at the domain

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interface. The pocket is deep and spacious, approximately 14.4 Å in depth and 850 $Å^3$ in volume (Methods) and predominantly hydrophobic (Fig. 2a, Extended Data Fig. 2).

The CghA–**1** complex structure shows that the bound ligand is actually the geometric diastereomer of **1** (Fig. 4a). We will discuss the importance of this finding later. The tetramic acid-containing side chain forms hydrogen bonds extensively with residues S65, N82, H94, K352 and N364 (Fig. 2b). Hydrogen-bonding donors can withdraw electron density and lower the lowest unoccupied molecular orbital (LUMO) energy of the dienophile to promote the DA cyclization reaction[21]. Therefore, the observed direct hydrogen-bonding interactions between S65 and N364 and the carbonyl groups of the tetramic acid, as well as the water-mediated hydrogen-bonding interaction N82 forms with S65 that can elevate the electron-withdrawing capacity of S65, may act to anchor the substrate into a catalytically relevant conformation within the binding pocket. More importantly, they may also lower the energetic barrier of the cyclization reaction (Supplementary Note 1 and Extended Data Fig. 3b). In a related work, we have also shown that tetramic acids are ionized in solution and that protonation, as would occur in the proton-rich binding site of CghA, enhances the dienophile reactivity for a similar octalin-forming IMDA reaction[10].

The polyketide-based octalin portion of **1** is surrounded almost entirely by hydrophobic side chain groups of the pocket-lining residues, including W183, W235, A242, M257 and V391 (Fig. 2b). The octalin core is flanked on each side by the indole side chains of W235 and W183. While the W183 side chain only forms a loose and angled edge-to-face interaction with the octalin core, the W235 side chain engages in a face-to-face stacking interaction with the octalin ring system (Fig. 2c). This type of stacking interaction between the cyclizing part of the substrate and the planar aromatic side chain groups has also been observed among natural and artificial DAases[16,22]. The packing interaction between the pocket and the bound octalin portion of the product appears to be tight, suggesting that the close shape complementarity is likely to play a key role in how CghA controls the diastereoselectivity of the DA reaction (Fig. 2c). Comparison of the apo and complex structures shows that the active site undergoes a minimal structural rearrangement upon ligand binding. The root-mean-square deviation between the two structures was 0.213 Å for the 379 matching α-carbons spanning the entire molecule. The major conformational changes are found with the key active site residue W235 and the loop on which it is located. These conformational changes open a pocket to create the space for the active site residue W235, which stacks against the octalin ring system (Fig. 2c). Similarly, the K352 side chain, which points away from the pocket opening in the apo structure, turns towards the pocket to form a specific hydrogen bond with the bound ligand in the complex structure (Fig. 2c).

Mechanistic studies of CghA and its mutants.

To investigate the mechanism of the enzymatic DA reaction, we designed a simplified substrate analogue **8** for in vitro assay of CghA (Fig. 3a, Supplementary Figs. 7–21, Supplementary Note 2). The crystal structure of CghA shows that the methyl group at C9 and the methyl and hydroxyl groups at the γ-position of the L-glutamic acid residue had minimal engagements with the active site-lining residues (Supplementary Fig. 7). Therefore, a simplified analogue **8** lacking these substituents was designed. CghA accepted **8** as a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

substrate and formed the DA adduct **9** as a product (Fig. 3a and Supplementary Note 3). The activity of CghA against **8** as a substrate was examined by single-point kinetic analyses. The apparent turnover number at the substrate concentration of 150 μM (hereafter the apparent k_{cat} or $k_{cat(app)}$) of the wild-type CghA was 30.4 ± 1.0 min^{-1}, similar to those determined for MycB (ref.[8]) and IccD (ref.[11]) (Methods, Fig. 3b and Supplementary Fig. 54 in Supplementary Note 5). We also designed CghA mutants based on the crystal structures and produced them recombinantly in *Escherichia coli* to identify residues that confer the *endo* selectivity (Methods). The mutants were assayed for the formation of the *endo* **9** and *exo* **10** adducts (Fig. 3b,c, Extended Data Figs. 3a,5 and 6 and Supplementary Notes 4 and 5) to determine the effect of the mutations on the stereocontrol of the cycloaddition reaction catalysed by CghA. First, S65, N82, H94, K352 and N364 were mutagenized to examine their contributions towards catalysis. Mutagenesis of S65, N82 and N364, which act as direct (S65 and N364) and indirect (N82) hydrogen-bonding donors to the carbonyl oxygens of the tetramic acid moiety (Fig. 2b, Extended Data Fig. 3b), resulted in substantial loss of activity, supporting the postulate that hydrogen bonds promote the DA reaction. However, in all cases the *endo* adduct was the dominant product after subtracting the background formation of the *endo* and *exo* products (Extended Data Figs. 3a and 4 and Supplementary Note 4). These findings indicate that protonation and hydrogen-bonding interactions between the tetramic acid moiety of **1** and CghA drive substrate binding and/or catalysis but do not contribute to the diastereoselectivity achieved by the enzyme. Similarly, when the octalin-core-packing W183 and W235 were mutagenized, the activities of the W235A and W183A–W235A double mutants were diminished, but did not show any increase in the *exo* selectivity (Fig. 3b,c, Extended Data Fig. 5 and Supplementary Fig. 45 in Supplementary Note 4 and Supplementary Figs. 50 and 52 in Supplementary Note 5). Surprisingly, the W183A mutation alone had almost no effect on the activity. The loss of the W183 side chain may be augmented by repositioning of the M257 side chain. Mutagenesis of W235 results in substantial loss of activity, with no residue to compensate for the missing octalin-stacking aromatic group.

Comparison of the amino acid sequences of the octalin-forming DAases CghA, Eqx3, Fsa2, MycB, Phm7 and PvhB indicates that these enzymes share less than 38% sequence identity with each other (Methods, Supplementary Fig. 5). Moreover, only 1 out of 22 binding pocket-lining residues (highlighted in yellow in Supplementary Fig. 5a) is conserved among the six DAases. Instead, the majority of the conserved residues occur in clusters in the second shell of the active site, many engaged in hydrogen-bonding and van der Waals interactions among themselves. Thus, the sequence conservation seems to function primarily in keeping the general structure of the binding pockets constant across these octalin-forming DAases. Construction of homology models of Eqx3 and MycB, members with the highest and lowest sequence homology to CghA, respectively and computational docking of corresponding ligands with the models (Methods, Supplementary Fig. 6) indicated that the lack of sequence conservation among binding pocket-lining residues is a consequence of refining the shape complementarity of the binding pockets to structurally variable substrates. The active site structure of CghA suggests that catalysis is achieved mainly by binding and restricting the movement of the straight-chain substrate into a conformation that facilitates a stereoselective cycloaddition reaction. Except for some

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

occurrences of hydrogen-bonding residues that can function to lower the reaction energy barrier, precise positioning of specific catalytic residues against bound substrate is secondary in catalysing DA reactions of tetronic and tetramic acids. Hence, there is less evolutionary drive to conserve active sites or key catalytic residues among enzymes that catalyse DA reactions. This may be one of the reasons why there is no unified fold among the proposed DAases identified to date[2,3]. Regardless, this lack of necessity to achieve precise positioning of the substrate and catalytic residues within the binding pocket, along with our results from the mutagenesis studies, hints to the feasibility of engineering AttH-like superfamily DAases, including CHGG_01241 (ref.[23]) for chaetoglobosin biosynthesis and CcsF (ref.[24]) for cytochalasin biosynthesis, to tune their activities towards new substrates or disfavoured reaction pathways.

Reversing the *endo*–*exo* diastereoselectivity of CghA.

Next, we examined the role of shape complementarity in conferring to CghA the ability to exert diastereoselectivity over the DA reaction it catalyses. The main structural difference between the *endo* and *exo* transition states (TSs) of the reaction is the conformation of the C3–C4 segment of the reactant (Fig. 1b). Examination of the crystal structure indicated that residues A242, M257 and V391 are in close proximity to the C3–C4 portion of the bound ligand. We reasoned that replacing A242 and V391 with slightly bulkier residues and M257 with a smaller residue might shift the pocket shape to be more complementary to the *exo* TS than the *endo* TS, thereby converting CghA into a disfavoured *exo* adduct-forming DAase. Initially, we prepared an A242S/M257V double mutant, but in vitro assay with **8** showed that the mutations resulted in loss of activity ($k_{cat(app)}$ of 8.53 ± 0.59 min^{-1}) and reductions in *endo* selectivity (Fig. 3c and Supplementary Fig. 53 in Supplementary Note 5). Next, a V391L mutation was introduced to the A242S/M257V double mutant to eliminate potential void space between the *exo* TS and the pocket (Fig. 2d, Methods). Single mutants for each of the three residues did not exhibit any *exo* selectivity, but the triple mutant dominantly produced the *exo* adduct, reversing the diastereoselectivity to preferentially form the disfavoured product (Fig. 3c and Supplementary Figs. 45–49 in Supplementary Note 4). While the overall $k_{cat(app)}$ of the triple mutant decreased 6.4-fold, the $k_{cat(app)}$ values for *endo* and *exo* formation were 1.92 ± 0.13 min^{-1} and 2.86 ± 0.19 min^{-1}, respectively (Extended Data Figs. 5 and 6 and Supplementary Fig. 56 in Supplementary Note 5). This study provides an example of rational engineering of a natural DAase to alter the stereoselectivity to favour the formation of a disfavoured adduct. We also investigated combinations of mutations at positions A242, M257 and V391 for increased *exo* selectivity in CghA (Fig. 3c and Supplementary Figs. 45–49 in Supplementary Note 4). This identified the A242N/M257V/V391L triple mutant that exhibited an even greater *exo* selectivity. While the wild-type exhibited 2.5% diastereoselectivity towards the *exo* adduct, the same diastereoselectivities of the A242S/M257V/V391L and A242N/M257V/V391L mutants were 57.9 and 72.7%, respectively. It appears that the bulkier side chain of asparagine than serine was able to fill the active site space better for improved packing against the *exo*-associated reactants. However, the spatially more constrained active site may have resulted in the reduced catalytic activity as indicated by the greater loss of activity with the A242N/M257V/V391L triple mutant over the A242S/M257V/V391L triple mutant (Extended Data Figs. 5 and 6 and Supplementary Fig. 56 in Supplementary Note 5). While

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the activity of these *exo* selective triple mutants has been reduced significantly, they could serve as productive seeds for directed evolution to highly active mutants.

Computational studies of the [4 + 2] cycloaddition reaction.

We performed density functional theory and wave function theory calculations on a simple protonated model to evaluate the energetics of the CghA catalysed *endo* and *exo* cycloadditions (Supplementary Note 6). Despite extensive conformational searching and testing various methods, we were unable to reproduce the non-enzymatic reaction results with calculated **TS-1** and **TS-2**. Therefore, we cautiously analysed the enzymatic reactions only in terms of reactivity and structure (Supplementary Note 6). Based on the planar geometry of the electron density for the ligand around the C17–C18 bond (Extended Data Fig. 2) as well as the structural arrangement of S65 and N364 residues in the crystal structure (Fig. 2b), we deduced that the substrate was protonated at the C17 carbonyl group and adopted the typical enol tautomer of tetramic acids in the active site. This arrangement optimizes the interactions between the carbonyl groups of the tetramic acid ring and the side chain groups of S65 and N364. For simplicity of the model, we also protonated the carboxylate to represent the active site. The results of the computational analysis with this protonation state indicate that CghA distorts the substrate from the lowest energy conformation (**R-1**) to an (*E*)-tetramic acid isomer (**R-1′**) in order to maximize binding of the tetramic acid (Fig. 4). This distortion increases the intrinsic *endo* and *exo* cycloaddition barriers by 4 kcal mol^{-1} based on our calculations (transition states **TS-1** versus **TS-1′** and **TS-2** versus **TS-2′**), but the stabilization from the hydrogen-bonding network in the enzyme pocket must outweigh this substrate destabilization. Calculations suggest that CghA exerts control over the *endo* versus *exo* selectivity (**TS-1′** versus **TS-2′**) by a secondary orbital interaction between the exocyclic alcohol of the tetramic acid and the 2-position of the diene. We propose that this secondary orbital interaction leads to a 4 kcal mol^{-1} stabilization of the *endo* **TS-1′**. Strikingly, the mutant CghA can overcome this intrinsic preference for the *endo* adduct and catalyse the formation of the *exo* adduct via **TS-2′**. The conformations of **TS-1′** and **TS-2′** do not lead directly to natural product **1** and the *exo* diastereomer **4**. These transition states lead to geometric isomers **1′** and **4′** that, upon product release, are deprotonated in solution and equilibrate with **1** and **4**.

One question often asked of a DAase is how it circumvents product inhibition, which is thought to occur due to high structural similarity between the transition state and the product[3]. The substantially lower kinetic parameters of the catalytic antibodies 13G5 (k_{cat} 1.20×10^{-3} min^{-1} for >98% disfavoured *exo* adduct formation)[25] and 1E9 (k_{cat} 13 min^{-1})[26] compared to that of CghA ($k_{cat(app)}$ of 30.4 ± 1.0 min^{-1}) are in part blamed on product inhibition occurring in these primitive artificial enzymes. However, the kinetic assay of CghA with **8** in the presence of increasing concentrations of the reaction product **9** indicated that CghA does not suffer significantly from product inhibition (Extended Data Fig. 7). Optimization of the active site towards the higher-energy intermediate and the tendency of the product to ionize to the well-solvated anion upon release from the enzyme are likely to be the strategies employed by CghA to avoid product inhibition and achieve high reaction turnover. The crystal structure of the CghA–**1** complex indicates that the residue S65, which is observed to hydrogen bond with the amide oxygen of the tetramic acid moiety (Fig.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2b) and was proposed earlier to lower the LUMO energy of the dienophile to promote the cycloaddition reaction, could also play a vital role in substrate preparation and product release. In line with this hypothesis, the S65A mutant exhibited the largest loss of activity among all of the mutants we examined (Fig. 3b, Extended Data Fig. 3 and Supplementary Fig. 44 in Supplementary Note 4 and Supplementary Fig. 50 in Supplementary Note 5), clearly indicating that S65 plays an important role in the catalytic mechanism of CghA presumably by hydrogen bonding. Similarly, the N82A mutant also lost almost all activity (Fig. 3b, Extended Data Fig. 3a and Supplementary Fig. 44 in Supplementary Note 4 and Supplementary Fig. 50 in Supplementary Note 5), suggesting that the catalytic activity of N82 is likely to further enhance the LUMO-lowering electron withdrawal by S65 through the water-mediated hydrogen-bonding interaction (Extended Data Fig. 3b, Supplementary Note 1). The free energy of the (*Z*)-isomeric **TS-1** for non-enzymatic cyclization was calculated to be 4.2 kcal mol^{-1} more stable than **TS-1′**. This indicates that CghA has evolved to stabilize a different TS of the reaction it catalyses compared to the TS favoured for the spontaneous reaction. By this strategy, CghA is likely to avoid product inhibition by compensating for the higher energy cost through pre-organizing its active site to optimally stabilize the higher-energy **TS-1′**. To support this proposal, we compare docking scores (predicted binding affinities) for products and transition states (Supplementary Method, Supplementary Fig. 7). Interestingly, amino acid sequence analysis of lipocalin-type DAase homologues showed that the DAases can be divided into two types based on the types of tetramic acid moiety present in the product (Extended Data Fig. 1). This indicates that the catalytic mechanism employed by the pyrrolidine-2,4-dione-type lipocalin-type DAases (type B, Extended Data Fig. 1b) differs from the mechanism currently proposed for the type A lipocalin-type DAases, including CghA, Eqx3, Fsa2, Phm7 and PvhB. Further studies are ongoing to investigate the mechanistic differences among these octalin-forming lipocalin-type DAases.

In summary, structural characterization and mutagenesis of CghA yields insight into the catalytic mechanism of this enzyme and provides further examples of authentic DAases. This strictly *endo*-selective enzyme was converted into a predominantly *exo*-selective enzyme by rational site-directed mutagenesis to yield the disfavoured *exo* adduct. This finding is important, as only artificially designed antibodies[27] and the natural enzymes PvhB (ref.[10]) and solanapyrone synthase[28] have been observed to perform *exo* adduct formation. The result also indicates that subtle tuning of the shape complementarity of the binding pocket to the reaction transition state is key to establishing desired diastereoselectivity in octalin-forming DAases. The combination of the CghA–**1** complex structure and computational analyses of the CghA-catalysed reaction showed that a water-mediated hydrogen-bonding network among two active-site residues and the enol tautomer of the substrate enhances reactivity by withdrawing electron density from the dienophile and plays a key role in how CghA catalyses the octalin-forming pericyclic reaction. Furthermore, our study also suggested that CghA is optimized to bind higher-energy transition states than those involved in the non-enzymatic reaction pathways that give alternative isomers. This optimal binding of higher-energy transition states and the tendency of the product to become ionized and solvated upon release from the active site appear to confer to CghA its ability to avoid product inhibition and realize high reaction turnover. Taken together, our results can

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

serve as a basis for further efforts towards developing DAases that can make the processes of medicinal chemistry and industrial organic syntheses clean and more efficient.

Methods

Strains and general techniques for DNA manipulation.

PCR was performed using PrimeSTAR GXL DNA polymerase (Takara Bio., Inc.). *E. coli* XL1-Blue (Stratagene) was used for plasmid propagation and BL21(DE3) was used for preparation of recombinant CghA. Sequences of constructed plasmids were confirmed through DNA restriction enzymes and were used as recommended by the manufacturer (Thermo Fisher Scientific). Sequences of PCR products were confirmed through DNA sequencing (Macrogen Japan Corporation).

Spectroscopic analyses.

NMR spectra were obtained with a Bruker BioSpin AVANCE III HD 500 MHz spectrometer (^{1}H 500 MHz, ^{13}C 125 MHz). ^{1}H NMR chemical shifts are reported in ppm using the proton resonance of residual solvent as reference: $CDCl_3$ ($\delta = 7.26$) and CD_3OD ($\delta = 3.31$)[29]. ^{13}C NMR chemical shifts are reported relative to CDCl ($\delta = 77.16$) and CD_3OD ($\delta = 49.0$)[29]. Mass spectra were recorded with a Thermo Fisher Scientific ACCELA Exactive liquid chromatography–mass spectrometer (LC–MS) by using both positive and negative electrospray ionization. Samples were separated for analysis on an ACQUITY UPLC 1.8 μm, 2.1 × 50 mm C18 reversed-phase column (Waters) using a linear gradient of 5–100% v/v MeCN in H_2O supplemented with 0.05% v/v formic acid at a flow rate of 0.5 ml min^{-1}.

Construction of pKW20460 for expression of *cghA*.

The *cghA* gene was synthesized by Eurofins Genomics to optimize its codon usage for expression in *E. coli*. The synthesized *cghA* was amplified by PCR using two primers, pKW20460-f and pKW20460-r (see Supplementary Table 3 for primer sequence information) and integrated with pET21c(+) expression vector (EMD Millipore Corporation) that was digested with *Nde* I (10 units) and *Xho* I (10 units) to generate pKW20460 (Supplementary Fig. 1) using GeneArt Seamless Cloning and Assembly kit (Thermo Fisher Scientific). The identity of the resulting vector pKW20460 was confirmed by DNA sequencing (Macrogen Japan Corporation). See Supplementary Fig. 1 for the plasmid map. This plasmid was used to express *cghA* for preparing purified samples of recombinant CghA to be used in in vitro assays and crystallization.

Construction of plasmids for expression of CghA mutants.

The oligonucleotide primers used for generating the mutants and their sequences are listed in Supplementary Tables 2 and 3. The plasmid pKW20460 containing the wild-type *cghA* gene was used as the template for PCR-based site-directed mutagenesis. Two primers pKW19225_Fw1 and pKW19225_Rv1 were used for W235A mutation and the resulting plasmid carrying the mutant gene was named pKW19225. Similarly, the W183A, N364A, S65A, K352A, N82A, H94A, A242S, V391L and M257V mutants of CghA were constructed using the primer pairs pKW19249_Fw1/pKW19249_Rv1, pKW19252_Fw1/pKW19252_Rv1, pKW19253_Fw1/pKW19253_Rv1,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pKW19254_Fw1/pKW19254_Rv1, pKW19255_Fw1/pKW19255_Rv1, pKW19256_Fw1/pKW19256_Rv1, pKW19274_Fw1/pKW19274_Rv1, pKW19275_Fw1/pKW19275_Rv1 and pKW19273_Fw1/pKW19273_Rv1, respectively and the resulting plasmids carrying the mutant genes were named pKW19249, pKW19252, pKW19253, pKW19254, pKW19255, pKW19256, pKW19236, pKW18569 and pKW19273, respectively. The plasmid pKW19274 containing the A242S/M257V double mutant gene was prepared by using pKW19273 as the template for PCR-based site-directed mutagenesis. Two primers pKW19274_Fw1 and pKW19274_Rv1 were used for creating the A242S/M257V mutations. The A242Q/M257V, A242N/M257V, A242T/M257V, A242M/M257V, M257V/V391K, M257V/V391Q, M257V/V391T and M257V/V391F mutants of CghA were constructed using the primer pairs CghA-A242Q-F/CghA-A242Q-R, CghA-A242N-F/CghA-A242N-R, CghA-A242T-F/CghA-A242T-R, pKW19251_Fw1/pKW19251_Rv1, CghA-V391K-F/CghA-V391K-R, CghA-V391Q-F/CghA-V391Q-R, CghA-V391T-F/CghA-V391T-R and CghA-V391F-F/CghA-V391F-R, respectively and the resulting plasmids carrying the mutant genes were named pKW19273-A242Q, pKW19273-A242N, pKW19273-A242T, pKW19273-A242M, pKW19273-V391K, pKW19273-V391Q, pKW19273-V391T and pKW19273-V391F, respectively. The plasmid pKW4552 containing the W183A/W235A double mutant gene was prepared by using pKW19225 as the template for PCR-based site-directed mutagenesis. Two primers pKW19249_Fw1 and pKW19249_Rv1 were used for creating the W183A/W235A mutations. The plasmid pKW18570 containing the A242S/V391L double mutant gene was prepared by using pKW19236 as the template for PCR-based site-directed mutagenesis. Two primers pKW19275_Fw1 and pKW19275_Rv1 were used for creating the A242S/V391L mutations. The plasmid pKW18571 containing the M257V/V391L double mutant gene was prepared by using pKW19273 as the template for PCR-based site-directed mutagenesis. Two primers pKW19275_Fw1 and pKW19275_Rv1 were used for creating the M257V/V391L mutations. The plasmid pKW19275 containing the A242S/M257V/V391L triple mutant gene was prepared by using pKW19274 as the template for PCR-based site-directed mutagenesis. Two primers pKW19275_Fw1 and pKW19275_Rv1 were used for creating the A242S/M257V/V391L mutations. The A242S/M257V/V391K and A242S/M257V/V391T mutants of CghA were constructed by using pKW19274 as the template for PCR-based site-directed mutagenesis using the primer pairs CghA-V391K-F/CghA-V391K-R and CghA-A242T-F/CghA-A242T-R, respectively and the resulting plasmids carrying the mutant genes were named pKW19273-SK and pKW19273-ST, respectively. The A242M/M257V/V391L and A242Q/M257V/V391L mutants of CghA were constructed by using pKW18571 as the template for PCR-based site-directed mutagenesis using the primer pairs pKW19251_Fw1/pKW19251_Rv1 and CghA-A242Q-F/CghA-A242Q-R, respectively and the resulting plasmids carrying the mutant genes were named pKW19273-ML and pKW19273-QL, respectively. The A242N/M257V/V391Q, A242N/M257V/V391T and A242N/M257V/V391L mutants of CghA were constructed by using pKW19273-A242N as the template for PCR-based site-directed mutagenesis using the primer pairs CghA-V391Q-F/CghA-V391Q-R, CghA-V391T-F/CghA-V391T-R and CghA-A242N-F/CghA-A242N-R, respectively and the resulting plasmids carrying the mutant genes were named pKW19273-NQ, pKW19273-NT and pKW19273-NL, respectively. The accuracy of the DNA sequences of the coding regions of the plasmids prepared here, including the mutagenized positions,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

was confirmed by DNA sequencing (Macrogen Japan Corporation). See Supplementary Fig. 2 for maps of the plasmids constructed.

Production and purification of CghA and its mutants.

Overexpression and subsequent protein purification of CghA was performed as follows: *E. coli* BL21(DE3) harbouring plasmid pKW20460 was grown overnight in 20 ml of LB medium with 100 μg ml^{-1} ampicillin at 37 °C. Five litres of fresh LB medium with 100 μg ml^{-1} ampicillin was inoculated with 20 ml of the overnight culture and incubated at 37 °C until the optical density at 600 nm (OD_{600}) reached 0.8. Then expression of the gene was induced with 100 μM isopropylthio-β-D-galactoside at 15 °C. Incubation was continued for another 24 h, after which cells were harvested by centrifugation at 10,000*g* for 10 min. All subsequent procedures were performed at 4 °C or on ice. Harvested cells were resuspended in disruption buffer (0.1 M Tris-HCl at pH 7.4, 0.1 M NaCl and 20 mM imidazole). Cells were disrupted using a French Press and the lysate was clarified by centrifugation at 10,000*g* for 10 min. The supernatant and precipitate were recovered as soluble and insoluble fractions, respectively. The soluble fraction containing protein was applied onto a Ni-NTA Agarose (Thermo Fisher Scientific) column. After washing the column with 10 mM imidazole and 0.1 M NaCl in 0.1 M Tris-HCl (pH 7.4), the target protein was eluted with 0.1 to 1.0 M imidazole and 0.1 M NaCl in 0.1 M Tris-HCl (pH 7.4). Further purification was carried out on an anion-exchange column (HiTrap Q, 5 ml × 2, GE Healthcare). A gradient of 0.0–1.0 M NaCl in 0.1 M Tris-HCl (pH 7.4) was applied at a flow rate of 2 ml min^{-1} over 95 min. Fractions of 2 ml were collected and fractions with the desired protein, which typically eluted at 0.2 M NaCl, were pooled and further concentrated using an Amicon Ultra centrifugal concentrator (EMD Millipore Corporation). Then the protein was further purified using a Superdex 75 gel filtration column (16 × 600 mm, GE Healthcare) in a buffer containing 10 mM Tris-HCl (pH 7.4) and 0.1 M NaCl at a flow rate of 1 ml min^{-1}. The purified proteins in 10 mM Tris-HCl (pH 7.4) and 0.1 M NaCl were pooled and concentrated to 25 mg ml^{-1} using an Amicon Ultra centrifugal concentrator. Protein concentration was estimated using the Bio-Rad protein assay kit with bovine serum albumin as a standard. This sample was analysed using sodium dodecyl sulfate–polyacrylamide gel electrophoresis using Tris-HCl and 10% polyacrylamide gel stained with Coomassie Brilliant Blue R-250 stain solution (Nacalai Tesque) (Supplementary Fig. 3a). The *cghA* mutants were overexpressed and the mutant enzymes were purified using essentially the same procedure as described above (Supplementary Fig. 3b).

In vitro assay of CghA and its mutants against 8.

The assay mixtures (25 μl) containing various concentrations of the wild-type CghA or its mutants (specified in the chromatogram or in the figure legends) and 100 μM of **8** in the reaction buffer (100 mM potassium phosphate, 100 mM NaCl, pH 8.4) were incubated at 25 °C for 10 or 20 min. Relatively high concentrations of the enzymes were used only for poorly active mutants, for example 100 μM for many of the triple mutants, to be able to reliably observe the formation of the products. Heat-inactivated samples of the wild-type CghA or its mutants were used as negative controls. After the 10-min incubation, the reactions were quenched by addition of 25 μl of methanol containing 20 μM of *N*-Boc L-tryptophan methyl ester as an internal standard. The mixtures were centrifuged to remove

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

debris and the supernatants were subjected to LC–MS analysis performed with a Thermo Scientific Q-Exactive liquid chromatography mass spectrometer using both positive and negative electrospray ionization. LC was performed using an ACQUITY UPLC 1.8 μm, 2.1 × 50 mm C18 reversed-phase column (Waters) and separated on a linear gradient of 10–50% v/v acetonitrile in H_2O supplemented with 0.05% v/v formic acid at a flow rate of 500 μl min^{-1}. Peak heights of different samples were standardized by scaling to the heights of the internal standard peaks in all samples. The resulting HPLC chromatograms are given in Extended Data Figs. 3 and 6 and Supplementary Note 4.

Analysis of the stereoselectivity of CghA and its mutants.

For more detailed kinetic characterizations of the wild-type CghA and the mutants, the assay mixture (25 μl) containing an appropriate concentration of CghA or its mutant (for example, 0.1 μM for the wild-type CghA and 0.5–4.0 μM for most of the relatively active mutants, well within 10% of the substrate concentration) and 150 μM of **8** in the reaction buffer (100 mM potassium phosphate, 100 mM NaCl, pH 8.4) was incubated at 25 °C for 3–6 min. After the incubation, the reaction was quenched by addition of 25 μl of methanol containing 20 μM of *N*-Boc L-tryptophan methyl ester as an internal standard. The mixture was centrifuged to remove the debris and the supernatant was subjected to LC–MS analysis performed with a Thermo Scientific Q-Exactive liquid chromatography mass spectrometer using both positive and negative electrospray ionization modes. LC was performed using an ACQUITY UPLC 1.8 μm, 2.1 × 50 mm C18 reversed-phase column (Waters) and separated on a linear gradient of 10–50% v/v acetonitrile in H_2O supplemented with 0.05% v/v formic acid at a flow rate of 500 μl min^{-1}. Recording of the chromatogram was performed at 288 nm. Peak heights of different samples were standardized by scaling to the heights of the internal standard peaks in all samples. To obtain the apparent turnover number $k_{cat(app)}$ of the enzyme, each of the areas of the peaks for the *endo* and *exo* adducts was measured. The peak area for the *endo* and *exo* adducts upon enzymatic conversion of the substrate can be calculated using the formulae $((A–B)/(2C–1))\times C$ and $((A–B)/(2C–1))\times(1–C)$ for the *endo* and *exo* peak areas, respectively, where A is the *endo* peak area measured at a specific reaction time, B is the *exo* peak area measured at the same specific reaction time and C and $(1–C)$ are the *endo* and exo *peak* areas determined upon complete conversion of the substrates, respectively. These peak areas are converted to the amount of adduct by the pre-determined conversion constant for each peak and divided by the concentration of the enzyme and the reaction time to determine the $k_{cat(app)}$ value. The resulting HPLC chromatograms and kinetic analyses are given in Fig. 3a, Extended Data Figs. 5 and 6 and Supplementary Note 5.

Assay on product inhibition of CghA by the *endo* adduct 9.

Prior to starting the reactions, the *endo* adduct **9** was added at concentrations of 0, 50, 100, 200 and 500 μM to the reaction mixtures containing 0.4 μM of wild-type CghA in the reaction buffer (100 mM potassium phosphate, 100 mM NaCl, pH 8.4). The reactions were initiated with an addition of 50 μM of **8** and were incubated at 25 °C for 2 min. The subsequent quenching and analysis of the reactions were carried out following the same protocol described above. Triplicate measurements were taken for each data point and the calculated standard errors are shown in the plot as error bars. The results are shown in Extended Data Fig. 7.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Production and purification of seleno-L-methionine-substituted CghA.

For preparation of seleno-L-methionine (SeMet)-substituted CghA, plasmid pKW20460 was transformed into electrocompetent BL21(DE3) *E. coli* cells. The BL21(DE3) harbouring plasmid pKW20460 was cultured overnight at 37 °C in 20 ml of M9 minimal medium containing 100 μg ml^{-1} ampicillin. This overnight culture was used to inoculate a total of 5 l of fresh M9 minimal medium in a fermentation vessel. After 9 h of incubation at 37 °C with agitation at 800 rpm, an OD_{600} of 0.3 was achieved. After addition of 100 mg l^{-1} of L-lysine, 100 mg l^{-1} of L-phenylalanine, 100 mg l^{-1} of L-threonine, 50 mg l^{-1} of L-isoleucine, 50 mg l^{-1} of L-valine and 60 mg l^{-1} of SeMet, incubation was continued for another 2.5 h. All amino acid vitamin solutions were filtered through a 0.22-μm filter (Sartorius). Protein expression was induced with isopropylthio-β-D-galactoside at a final concentration of 0.1 mM. After incubation for 24 h at 15 °C with agitation at 800 rpm, cells were harvested by centrifugation. All of the subsequent steps were performed at 4 °C. The cells were resuspended in the previously described disruption buffer and disrupted using a French Press. SeMet-substituted CghA was purified using essentially the same procedure as described above.

Protein crystallization and data collection.

A broad screening of crystallization conditions was performed for CghA using a sitting-drop method. Crystallization conditions that yielded microcrystals were reproduced and optimized in terms of pH, precipitant concentration and drop volume. The best crystal of CghA was obtained by mixing 1.0 μl of the protein (16.5 mg ml^{-1}) with 1.0 μl of a mother liquor of 15% PEG 4000 and 100 mM MES at pH 6.5. Crystals appeared after a few days at 20 °C and grew as octahedral crystals (Supplementary Fig. 4a). On the other hand, the best crystal of SeMet-substituted CghA could be obtained by mixing 1.0 μl of the protein (14.7 mg ml^{-1}) with 1.0 μl of a mother liquor of 15% PEG 6000 and 100 mM MES at pH 6.5. Crystals appeared after a few days at 20 °C and grew as octahedral crystals (Supplementary Fig. 4b). Prior to data collection, crystals were transferred to a cryoprotectant solution (15% ethylene glycol in the mother liquour) and immediately cryo-cooled in liquid nitrogen. Crystals of both CghA and SeMet-substituted CghA belong to the space group $P4_32_12$ and have one molecule per asymmetric unit. X-ray diffraction data were collected at the Photon Factory and processed using the program XDS version Nov. 1, 2016 (ref.[30]).

Co-crystallization and data collection.

Crystals of CghA in complex with **1** were grown at 20 °C by the hanging-drop vapour diffusion method. The protein solution was prepared by mixing 90.1 μM CghA, 0.5 mM **1** in 2.5% DMSO, 100 mM sodium chloride and 10 mM Tris-HCl at pH 7.4. The best crystal of CghA was obtained by mixing 1.0 μl of the protein solution with 1.0 μl of a mother liquor of 14.5% PEG 4000 and 100 mM MES at pH 6.5. CghA and **1** were mixed in a molar ratio of 1:5.5 with the final concentrations of CghA and **1** being 2.0 mg ml^{-1} and 111.3 mg ml^{-1}, respectively. Crystals appeared after a few days at 20 °C and grew as octahedral crystals (Supplementary Fig. 4c). Prior to data collection, crystals were transferred to a cryoprotectant solution (15% ethylene glycol in the mother liquor) and immediately cryo-cooled in liquid nitrogen. Crystals of CghA in complex with **1** belong to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the space group $P4_32_12$ and have one molecule per asymmetric unit. X-ray diffraction data were collected at the Photon Factory and processed with the program XDS version Jun. 1, 2017 (ref.[30]). Details of the computational methods, including protein structure refinement, model completion and analysis, ligand docking, homology model construction and amino acid sequence analyses are given in the Supplementary Methods.

Computational methods.

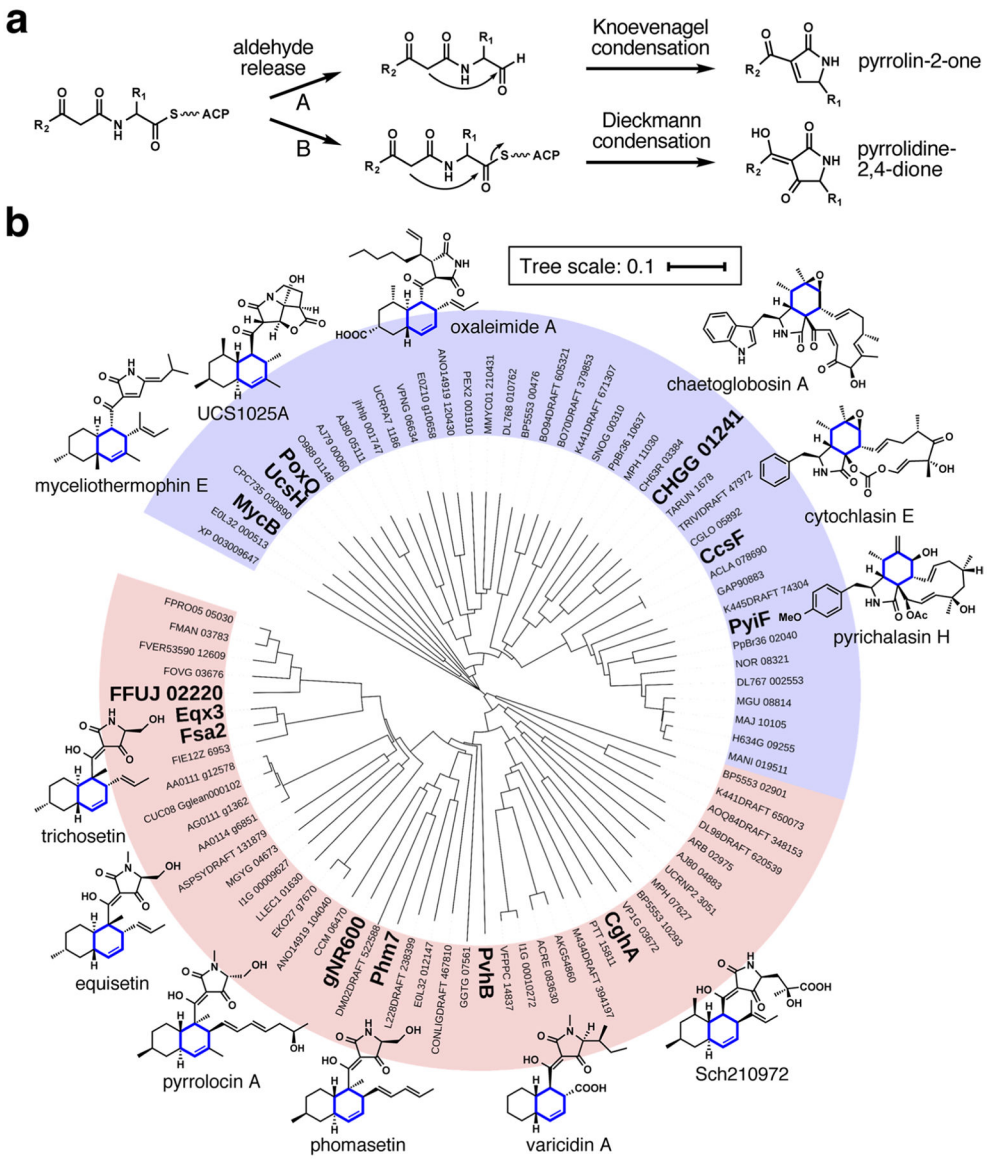
Initial conformational searches were completed using Maestro 2017-2 version 11.2.014 (Schrödinger)[31]. Conformers were recalculated quantum mechanically in Gaussian 16 Rev. A.03 (sse4)[32] using density functional theory with the ωB97X-D/6–31G(d)[33–38] functional. This functional was chosen for its ability to reproduce CCSD geometry calculations of asynchronous Diels–Alder reactions as well as its general applicability for accurately calculating reaction barriers[39,40]. Ground states were optimized using the default `opt` keyword in Gaussian. Transition states were found by first scanning the transition state region (b1 = 1.9–2.2 Å, b2 = 2.5–3.0 Å) with the keyword `opt = modredundant` and then optimizing the highest energy point with the Berny algorithm and without checking eigenvalues by specifying `opt = (TS, NoEigen)`. Using the basis set suggested by Head-Gordon for energetics[40], single-point energies were computed at the ωB97X-D/def2-QZVPP and CPCM(H_2O)-ωB97X-D/def2-QZVPP levels of theory[41–44]. To further check the validity of our calculations, we compared our energetics against the RI-SCS-MP2 wave function method using a cc-pVTZ basis set in the gas phase and the SMD(H_2O) model in solution as implemented in ORCA 4.1.1 (refs.[45–50]). The following keywords, as recommended by the ORCA Input Library (https://sites.google.com/site/orcainputlibrary/home), were specified in ORCA for the wave function method single-point calculations: `RI-SCS-MP2 cc-pVTZ cc-pVTZ/C TIGHTSCF`. Thermochemistry was calculated using GoodVibes 3.0.0 (ref.[51]) with the entropic quasi-harmonic corrections due to Grimme and the enthalpic quasi-harmonic corrections due to Head-Gordon[52,53] using frequency cut-offs of 100 wavenumbers. All reported structures were stationary points on the energy surface and were characterized as transition states or minima using frequency calculations. However, despite many efforts, structure **1** had a small imaginary frequency (approximately −60 wavenumbers) that was unable to be removed. For the method for computational docking experiments using Autodock Vina, please refer to Supplementary Methods.

Reporting Summary.

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

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Extended Data

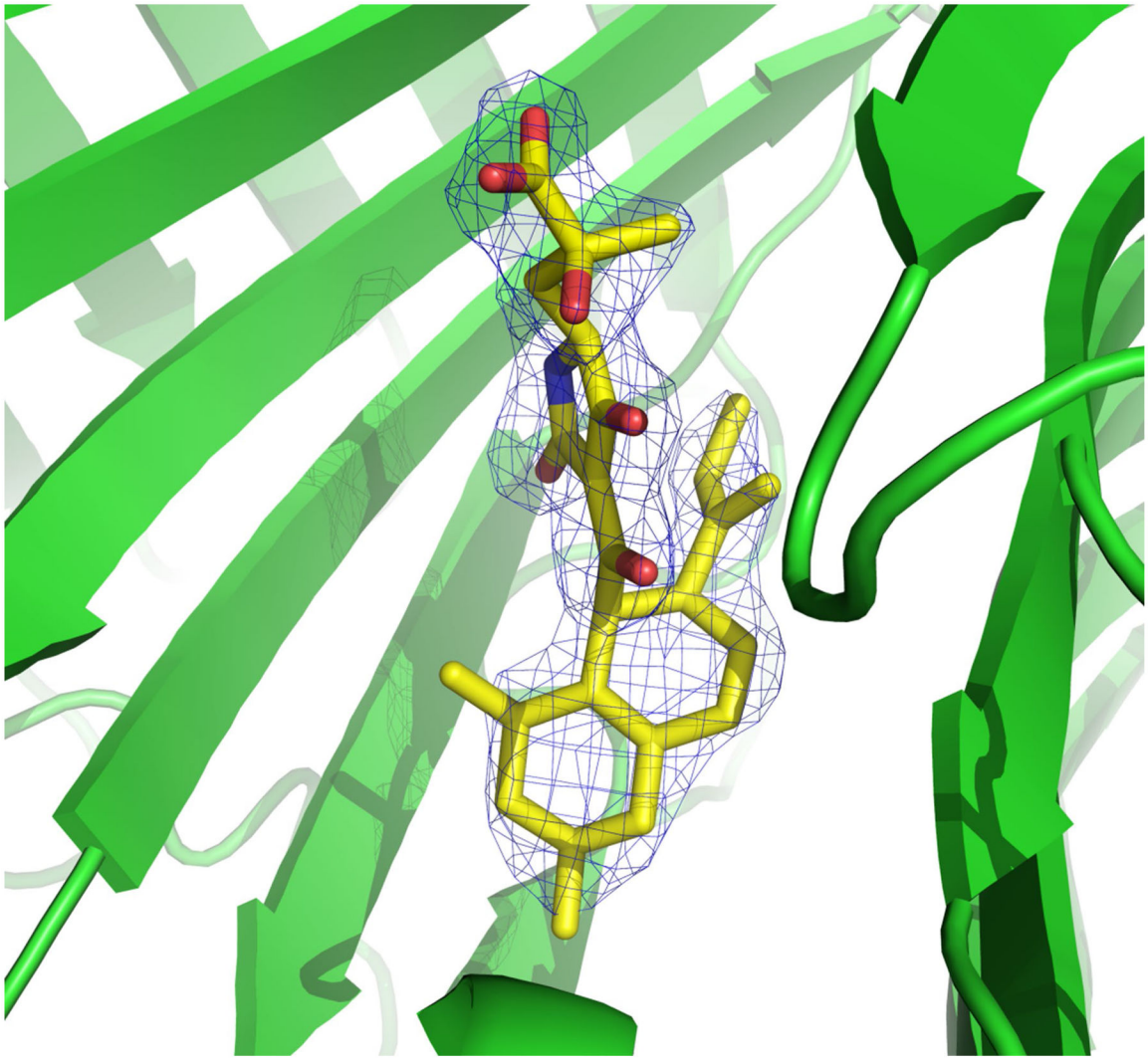


Extended Data Fig. 1 |. Classification of DAases by the mechanism of product release from the NRPS.

a, Proposed two distinct mechanisms of substrate release from the NrPS. One type of the embedded terminal reductase domain of an NrPS catalyses reduction and Knoevenagel condensation of the thioester intermediate to give a pyrrolin-2-one product via a rote A. Another type catalyses a Dieckmann condensation of the intermediate to give a pyrrolidine-2,4-dione product via a rote B. **b**, Phylogenetic tree of the lipocalin-type DAase homologs. Phylogenetic analysis of the amino acid sequences of the representative homologs found in the protein database revealed that they could be divided clearly into two groups based on the types of tetramic acid moieties described above that the corresponding products bear. Pink represents enzymes predicted to catalyse Knoevenagel condensation to release the pyrrolin-2-one-type (type A) product. Purple represents enzymes predicted to

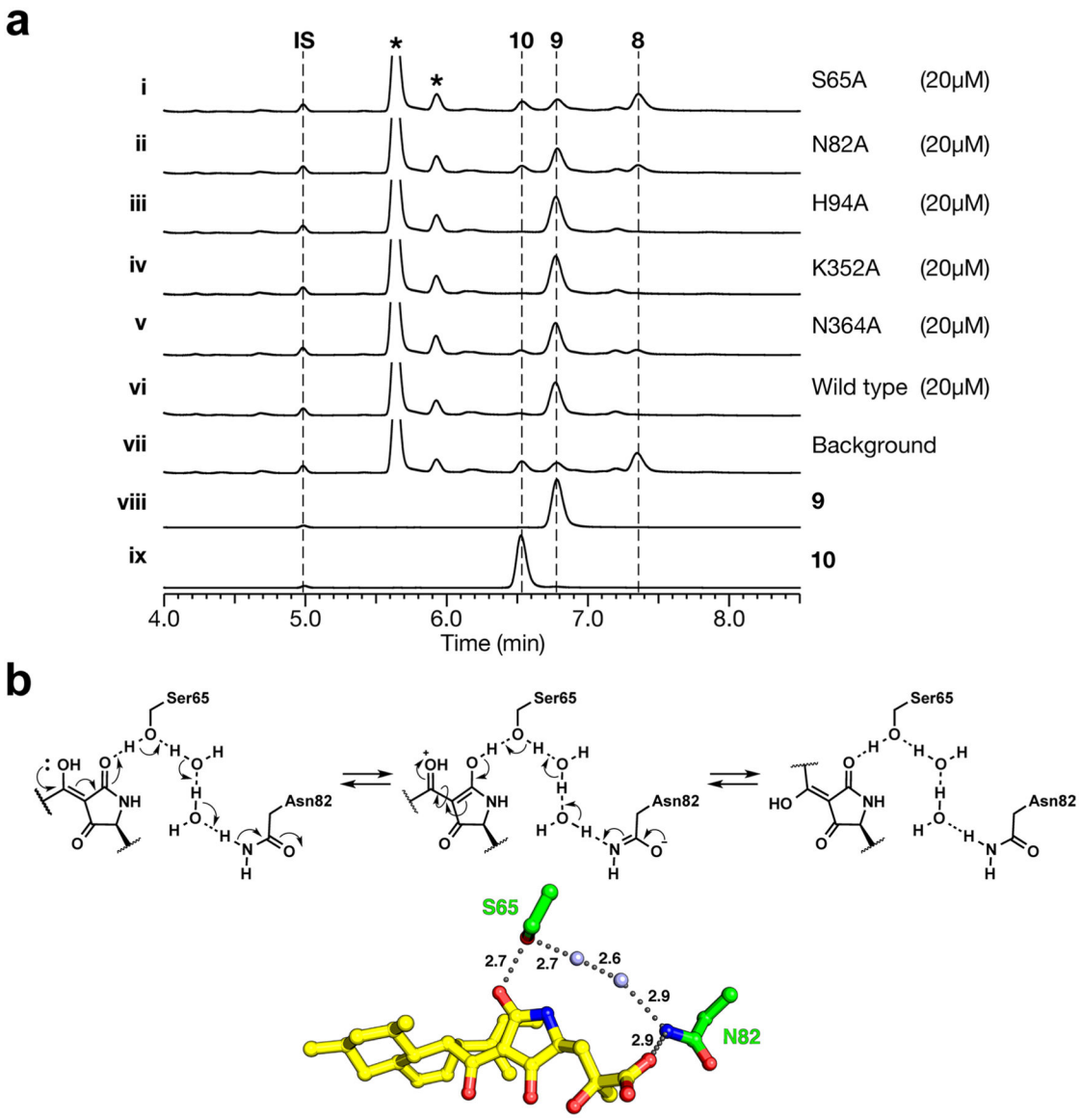
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

catalyse Dieckmann condensation to release the pyrrolidine-2,4-dione-type (type B) product. The length of the black bar is scaled to represent the evolutionary distance of 0.1 amino acid substitution per site. The enzyme names are abbreviated for the ease of representation.



Extended Data Fig. 2 |. Electron density of the bound ligand in the CghA co-crystal structure. The *Fo–Fc* map contoured at 3.0σ represented as thin blue mesh is shown to indicate the electron density of Sch210972 bound within the active site pocket of CghA. Colouring scheme for the ligand is the same as in Fig. 2a.

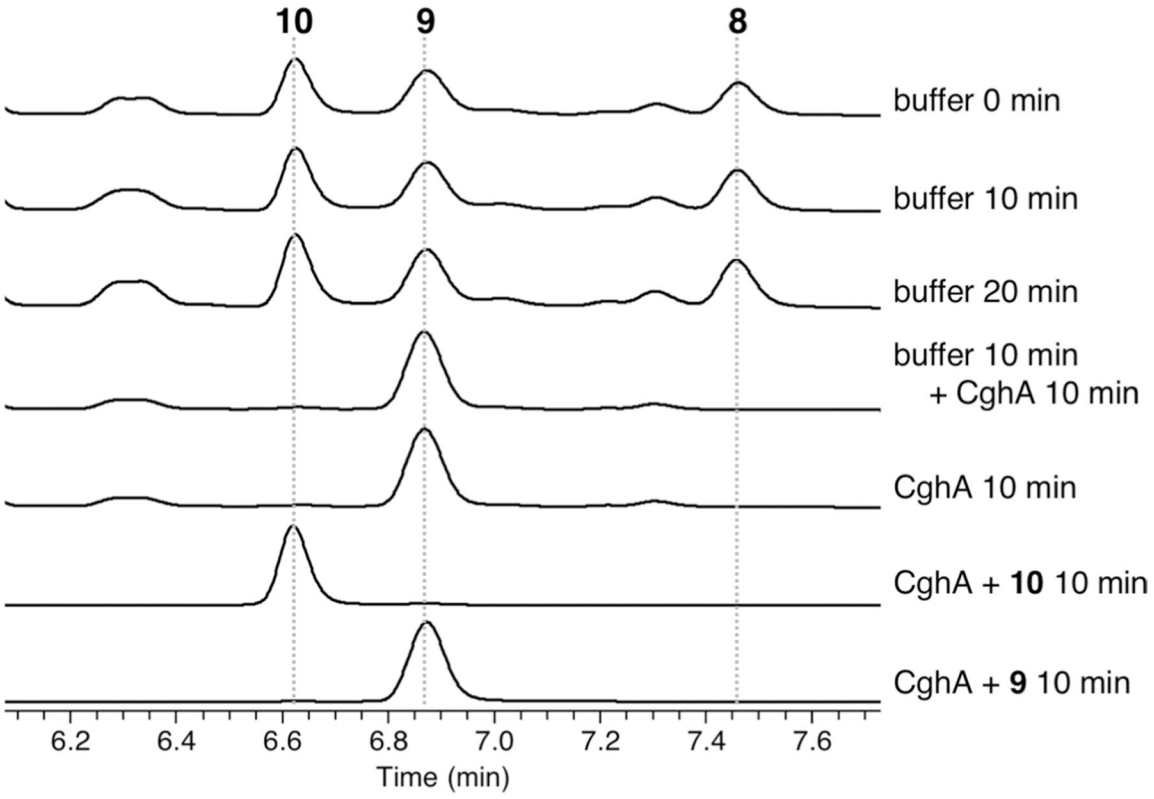
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 3 |. Activity assays on the wild-type CghA and its mutants to examine the roles of the active-site residues.

The in vitro analyses were performed using the substrate analogue **8** for the formation of the *endo* **9** and *exo* **10** adducts. Detailed reaction conditions are described in Methods unless otherwise specified below. **a**, hPLC profiles of the reaction mixtures containing **8** as a substrate and (i) the S65N mutant; (ii) the N82A mutant; (iii) the h94A mutant; (iv) the K352A mutant; (v) the N364A mutant; (vi) the wild-type CghA; (vii) buffer, as a background reaction showing the conversion of **8** to **9** and **10** after 10 minutes of incubation; (viii) the authentic reference of **9**; and (ix) the authentic reference of **10**. The reactions were monitored at 288 nm. *N*-Boc-L-tryptophan methyl ester was used as an internal standard (IS) throughout the study. See Methods for details. Peaks denoted by asterisks are impurities from the chemical synthesis of **8**. **b**, hydrogen-bonding interaction between the bound ligand **P-1** and the S65 and N82 side chain groups. Two water molecules also link the S65 and N82 side chain groups through hydrogen bonds.

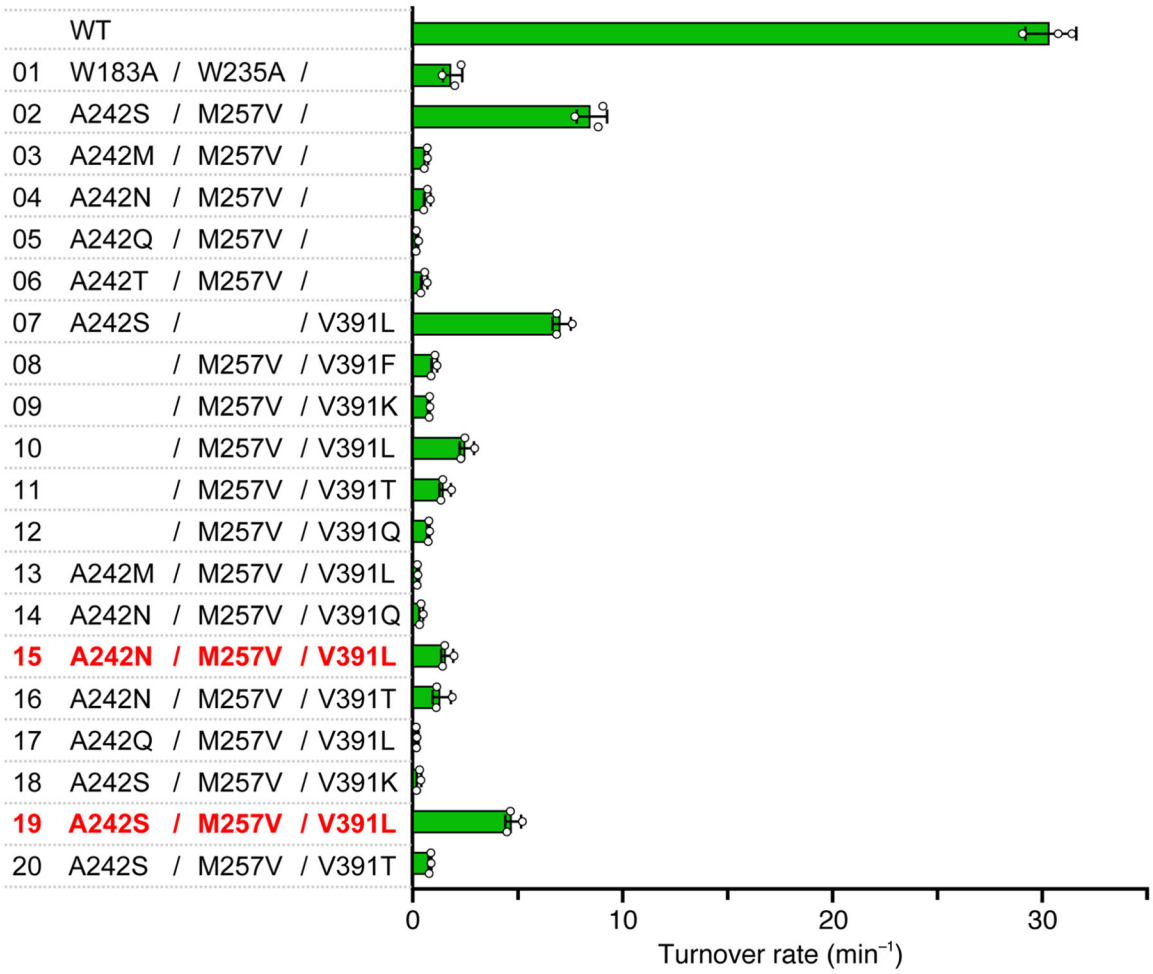
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 4 |. Background conversion of the substrate analogue 8.

Spontaneous transformation of the substrate analogue **8** into the *endo* **9** and *exo* **10** adducts occurs almost entirely in the hPLC solvent supplemented with formic acid. **8** (150 μM) was incubated for 0 (labelled "buffer 0 min" in the plot), 10 (buffer 10 min) and 20 min (buffer 20 min) at 25 °C without CghA in the reaction buffer (100 mM potassium phosphate, 100 mM NaCl, ph 8.4) in a total reaction volume of 25 μl. After incubation, the reaction was quenched with 25 μl of MeOh and then centrifuged for removal of debris. Subsequently, the sample was subjected to LC–MS analysis. For the control experiment "buffer 10 min + CghA 10 min", **8** (150 μM) was incubated for 10 min at 25 °C without CghA in the reaction buffer to allow the spontaneous transformation of **8**. Subsequently, 0.1 μM of the wild-type CghA was added to the reaction mixture and the mixture was incubated for another 10 min at 25 °C in a total reaction volume of 25 μl to allow CghA to react on the remaining **8**. For other control experiments "CghA 10 min", "CghA + **10** 10 min" and "CghA + **9**", 150 μM of **8**, **10** and **9** was added to the reaction mixture, respectively. These mixtures were incubated for 10 min at 25 °C with CghA (0.1 μM wild-type CghA) under the same conditions and analysed as described above.

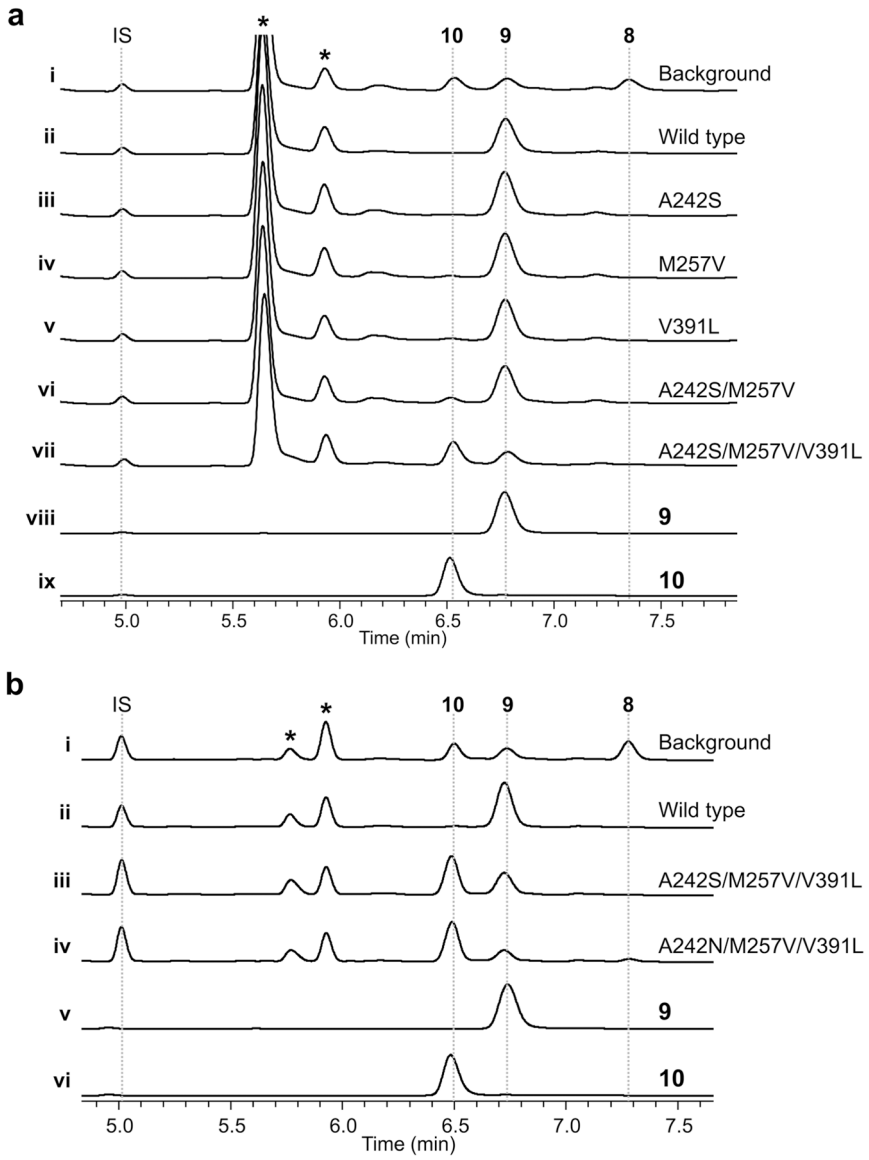
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 5 |. Total turnover rates of the wild-type CghA and its mutants designed for stereospecificity conversion.

The combined apparent turnover rates *k*cat(app) for the formation of **9** and **10** from **8** by the CghA mutants that were designed to examine the role of octalin-packing residues in determining the diastereoselectivity of the adducts formed are presented. The centre value is the mean of triplicate measurements with each data point shown as an open circle and the error bar representing the standard deviation. See Methods and Supplementary Note 5 for details.

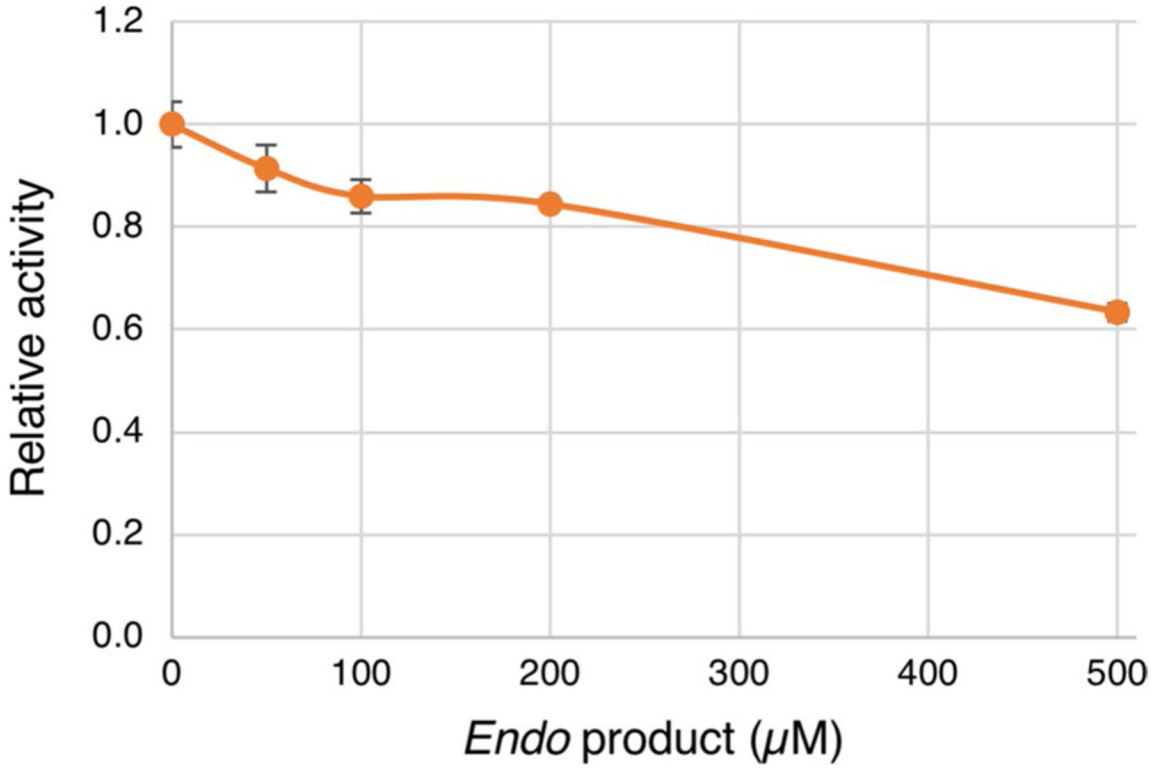
Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 6 |. Activity assays on CghA and its mutants to examine *endo*-to-*exo* stereospecificity conversion.

The *in vitro* analyses were performed using the substrate analogue **8** as the substrate. Detailed reaction conditions are described in Methods unless otherwise specified below. The reactions were monitored at 288 nm. *N*-Boc-L-tryptophan methyl ester was used as an internal standard (IS) throughout the study. See Methods for details. Peaks denoted by asterisks are impurities from the chemical synthesis of **8**. **a**, hPLC profiles of the reaction mixtures containing **8** as a substrate and (i) buffer only as a background reaction; (ii) the wild-type CghA; (iii) the A242S mutant; (iv) the M257V mutant; (v) the V391L mutant; (vi) the A242S/M257V mutant; (vii) the A242S/M257V/V391L mutant; (viii) the authentic reference of **9**; and (ix) the authentic reference of **10**. **b**, hPLC profiles of the reaction mixtures containing **8** as a substrate and (i) buffer, as a background reaction; (ii) the wild-type CghA; (iii) the A242S/M257V/V391L mutant; (iv) the A242N/M257V/V391L mutant; (v) the authentic reference of **9**; and (vi) the authentic reference of **10**.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



Extended Data Fig. 7 |. The kinetic assay to examine product inhibition of CghA.
In vitro analyses of CghA were performed with **8** as the substrate in the presence of an increasing concentration of the reaction product (the *endo* adduct **9**) to examine the possible product inhibition of CghA. The activity at each *endo* product concentration is given relative to the activity in the absence of the product. The centre value is the mean of triplicate measurements with the error bar representing the standard deviation. See Methods for details.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank T. Kan at the University of Shizuoka for providing us with important advice on the chemical synthesis of the substrate analogues. This work was supported by the NIH (GM118056 to Y. Tang and GM124480 to K.N.H.). This work was also supported in part by the Japan Society for the Promotion of Science (JSPS) (K.W., 15KT0068, 26560450 and 19KK0150; M.S., 19K07136) and Innovative Areas from MEXT, Japan (K.W., 16H06449).

Data availability

The crystal structure data are available on the PDB (accession code: 6KAW for the apo CghA structure, 6KBC for the CghA–**1** complex structure). The coordinates of calculated structures are available online at https://doi.org/10.19061/iochem-bd-6-67. The GenBank accession numbers of the amino acid sequences of the enzymes referenced in this study are provided in this paper. The source data underlying Supplementary Fig. 3a are provided as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a Source Data file. Source data are provided with this paper. All other data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Kim HJ, Ruszczycky MW & Liu HW Current developments and challenges in the search for a naturally selected Diels–Alderase. Curr. Opin. Chem. Biol 16, 124–131 (2012). [PubMed: 22260931]
2. Klas K, Tsukamoto S, Sherman DH & Williams RM Natural Diels–Alderases: elusive and irresistable. J. Org. Chem 80, 11672–11685 (2015). [PubMed: 26495876]
3. Minami A & Oikawa H Recent advances of Diels–Alderases involved in natural product biosynthesis. J. Antibiot. (Tokyo) 69, 500–506 (2016). [PubMed: 27301662]
4. Sato M et al. Involvement of lipocalin-like CghA in decalin-forming stereoselective intramolecular [4 + 2] cycloaddition. ChemBioChem 16, 2294–2298 (2015). [PubMed: 26360642]
5. Yang SW et al. Chemokine receptor CCR-5 inhibitors produced by *Chaetomium globosum*. J. Nat. Prod 69, 1025–1028 (2006). [PubMed: 16872138]
6. Kakule TB et al. Native promoter strategy for high-yielding synthesis and engineering of fungal secondary metabolites. ACS Synth. Biol 4, 625–633 (2015). [PubMed: 25226362]
7. Kato N et al. A new enzyme involved in the control of the stereochemistry in the decalin formation during equisetin biosynthesis. Biochem. Biophys. Res. Commun 460, 210–215 (2015). [PubMed: 25770422]
8. Li L et al. Biochemical characterization of a eukaryotic decalin-forming Diels–Alderase. J. Am. Chem. Soc 138, 15837–15840 (2016). [PubMed: 27960349]
9. Kato N et al. Control of the stereochemical course of [4 + 2] cycloaddition during *trans*-decalin formation by Fsa2-family enzymes. Angew. Chem. Int. Ed. Engl 57, 9754–9758 (2018). [PubMed: 29972614]
10. Tan D et al. Genome-mined Diels–Alderase catalyses formation of the *cis*-octahydrodecalins of varicidin A and B. J. Am. Chem. Soc 141, 769–773 (2019). [PubMed: 30609896]
11. Zhang Z et al. Enzyme-catalysed inverse-electron demand Diels–Alder reaction in the biosynthesis of antifungal ilicicolin H. J. Am. Chem. Soc 141, 5659–5663 (2019). [PubMed: 30905148]
12. Hantke V, Skellam EJ & Cox RJ Evidence for enzyme catalysed intramolecular [4 + 2] Diels–Alder cyclization during the biosynthesis of pyrichalasin H. Chem. Commun 56, 2925–2928 (2020).
13. Little R et al. Unexpected enzyme-catalysed [4 + 2] cycloaddition and rearrangement in polyether antibiotic biosynthesis. Nat. Catal 2, 1045–1054 (2019).
14. Takao K, Munakata R & Tadano K Recent advances in natural product synthesis by using intramolecular Diels–Alder reactions. Chem. Rev 105, 4779–4807 (2005). [PubMed: 16351062]
15. Chiu HJ et al. Structure of the first representative of Pfam family PF09410 (DUF2006) reveals a structural signature of the calycin superfamily that suggests a role in lipid metabolism. Acta Crystallogr. F 66, 1153–1159 (2010).
16. Zheng Q et al. Enzyme-dependent [4 + 2] cycloaddition depends on lid-like interaction of the N-terminal sequence with the catalytic core in PyrI4. Cell Chem. Biol 23, 352–360 (2016). [PubMed: 26877021]
17. Byrne MJ et al. The catalytic mechanism of a natural Diels–Alderase revealed in molecular detail. J. Am. Chem. Soc 138, 6095–6098 (2016). [PubMed: 27140661]
18. Fage CD et al. The structure of SpnF, a standalone enzyme that catalyses [4 + 2] cycloaddition. Nat. Chem. Biol 11, 256–258 (2015). [PubMed: 25730549]
19. Drulyte I et al. Crystal structure of the putative cyclase IdmH from the indanomycin nonribosomal peptide synthase/polyketide synthase. IUCrJ 6, 1120–1133 (2019).
20. Lakshmi B, Mishra M, Srinivasan N & Archunan G Structure-based phylogenetic analysis of the lipocalin superfamily. PLoS One 10, e0135507 (2015). [PubMed: 26263546]
21. Zheng Y & Thiel W Computational Insights into an enzyme-catalysed [4 + 2] cycloaddition. J. Org. Chem 82, 13563–13571 (2017). [PubMed: 29131960]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

22. Hugot M et al. A structural basis for the activity of retro-Diels–Alder catalytic antibodies: evidence for a catalytic aromatic residue. Proc. Natl Acad. Sci. USA 99, 9674–9678 (2002). [PubMed: 12093912]
23. Ishiuchi K et al. Combinatorial generation of complexity by redox enzymes in the chaetoglobosin A biosynthesis. J. Am. Chem. Soc 135, 7371–7377 (2013). [PubMed: 23611317]
24. Qiao K, Chooi YH & Tang Y Identification and engineering of the cytochalasin gene cluster from *Aspergillus clavatus* NRRL 1. Metab. Eng 13, 723–732 (2011). [PubMed: 21983160]
25. Heine A et al. An antibody *exo* Diels–Alderase inhibitor complex at 1.95 angstrom resolution. Science 279, 1934–1940 (1998). [PubMed: 9506943]
26. Xu J et al. Evolution of shape complementarity and catalytic efficiency from a primordial antibody template. Science 286, 2345–2348 (1999). [PubMed: 10600746]
27. Cannizzaro CE, Ashley JA, Janda KD & Houk KN Experimental determination of the absolute enantioselectivity of an antibody-catalysed Diels–Alder reaction and theoretical explorations of the origins of stereoselectivity. J. Am. Chem. Soc 125, 2489–2506 (2003). [PubMed: 12603137]
28. Kasahara K et al. Solanapyrone synthase, a possible Diels–Alderase and iterative type I polyketide synthase encoded in a biosynthetic gene cluster from *Alternaria solani*. ChemBioChem 11, 1245–1252 (2010). [PubMed: 20486243]
29. Gottlieb HE, Kotlyar V & Nudelman A NMR chemical shifts of common laboratory solvents as trace impurities. J. Org. Chem 62, 7512–7515 (1997). [PubMed: 11671879]
30. Kabsch W XDS. Acta Crystallogr. D 66, 125–132 (2010). [PubMed: 20124692]
31. Schrödinger. Release 2017–2: MacroModel v.11.2.014 (Schrödinger, LLC, 2017).
32. Frisch MJ et al. Gaussian 16 Rev. A.03 (Gaussian Inc., 2016).
33. Krishnan R, Binkley JS, Seeger R & Pople JA Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. J. Chem. Phys 72, 650–654 (1980).
34. Rassolov VA, Pople JA, Ratner MA & Windus TL 6–31G* basis set for atoms K through Zn. J. Chem. Phys 109, 1223–1229 (1998).
35. Francl MM et al. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. J. Chem. Phys 77, 3654–3665 (1982).
36. Dill JD & Pople JA Self-consistent molecular orbital methods. XV. Extended Gaussian-type basis sets for lithium, beryllium, and boron. J. Chem. Phys 62, 2921–2923 (1975).
37. Hehre WJ, Ditchfield R & Pople JA Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. J. Chem. Phys 56, 2257–2261 (1972).
38. Chai JD & Head-Gordon M Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. Phys. Chem. Chem. Phys 10, 6615–6620 (2008). [PubMed: 18989472]
39. Linder M & Brinck T On the method-dependence of transition state asynchronicity in Diels–Alder reactions. Phys. Chem. Chem. Phys 15, 5108–5114 (2013). [PubMed: 23450171]
40. Mardirossian N & Head-Gordon M Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. Mol. Phys 115, 2315–2372 (2017).
41. Clark T, Chandrasekhar J, Spitznagel GW & Schleyer PVR Efficient diffuse function-augmented basis sets for anion calculations. III. The 3–21 + G basis set for first-row elements, Li–F. J. Comput. Chem 4, 294–301 (1983).
42. Blaudeau J-P, McGrath MP, Curtiss LA & Radom L Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. J. Chem. Phys 107, 5016–5021 (1997).
43. Schäfer A, Huber C & Ahlrichs R Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. J. Chem. Phys 100, 5829–5835 (1994).
44. Weigend F & Ahlrichs R Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. Phys. Chem. Chem. Phys 7, 3297–3305 (2005). [PubMed: 16240044]
45. Grimme S Improved second-order Møller–Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. J. Chem. Phys 118, 9095–9102 (2003).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

46. Dunning THJ Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J. Chem. Phys 90, 1007–1023 (1989).
47. Weigend F, Köhn A & Hättig C Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. J. Chem. Phys 116, 3175–3183 (2002).
48. Marenich AV, Cramer CJ & Truhlar DG Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J. Phys. Chem. B 113, 6378–6396 (2009). [PubMed: 19366259]
49. Neese F The ORCA program system. WIREs Comput. Mol. Sci 2, 73–78 (2012).
50. Neese F Software update: the ORCA program system, version 4.0. WIREs Comput. Mol. Sci 8, e1327 (2018).
51. Paton L, Rodríguez-Guerra J, Chen J & Funes-Ardoiz I bobbypaton/GoodVibes: GoodVibes v3.0.0 (Zenodo, 2019).
52. Grimme S Supramolecular binding thermodynamics by dispersion-corrected density functional theory. Chem. Eur. J 18, 9955–9964 (2012). [PubMed: 22782805]
53. Li Y-P, Gomes J, Mallikarjun Sharada S, Bell AT & Head-Gordon M Improved force-field parameters for QM/MM simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies. J. Phys. Chem. C 119, 1840–1850 (2015).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

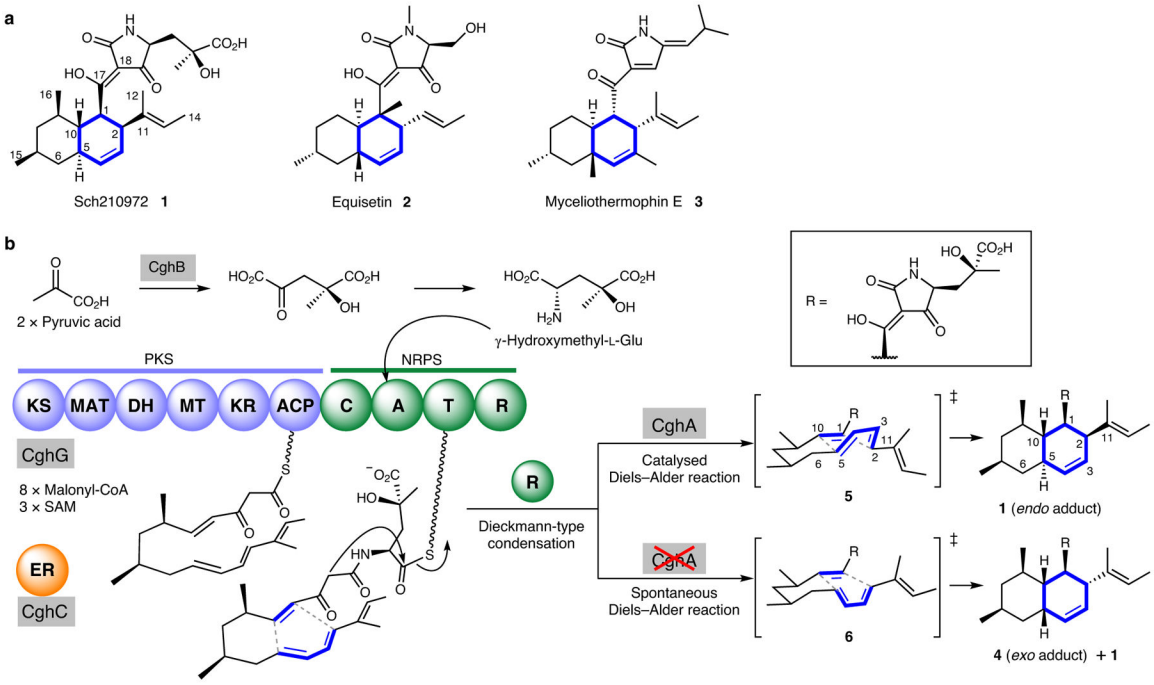


Fig. 1 |. Octalin-containing fungal secondary metabolites and proposed involvement of Diels–Alder reactions in their biosynthesis.

a, Sch210972 (**1**), equisetin (**2**), myceliothermophin E (**3**). **b**, The proposed biosynthesis of **1**. The hybrid polyketide synthase (PKS)–nonribosomal peptide synthetase (NrPS) CghG and the stand-alone enoyl reductase (Er) CghC together form the linear intermediate. The terminal reductase (r) domain of CghG is proposed to catalyse the tetramate moiety-forming Dieckmann-type condensation. CghA is thought to catalyse the cycloaddition reaction. KS, ketosynthase; MAT, malonyl-CoA acyltransferase; Dh, dehydratase; MT, methyltransferase; Kr, ketoreductase; ACP, acyl carrier protein; C, condensation; A, adenylation; T, thiolation; SAM, *S*-adenosyl-L-methionine.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

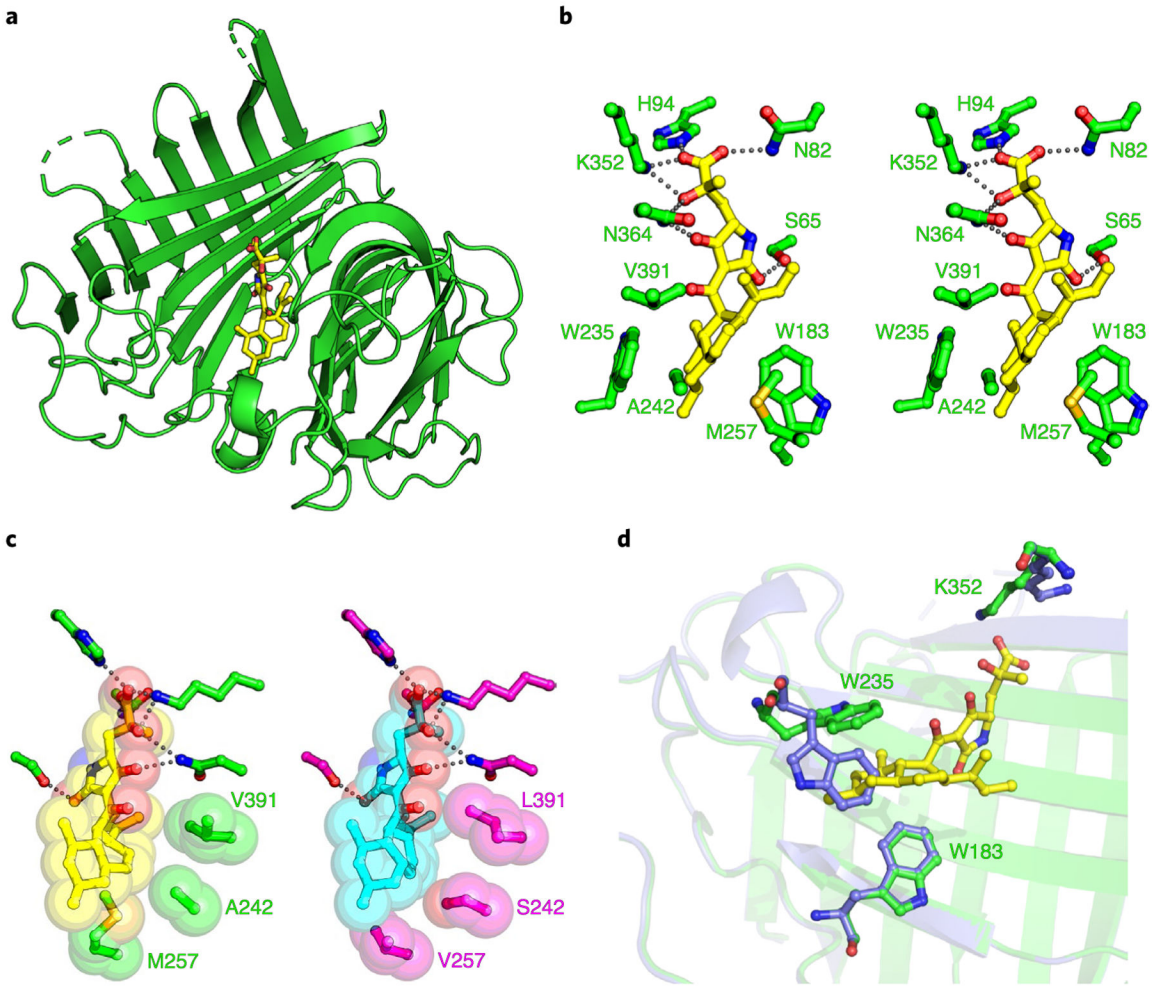


Fig. 2 |. Crystal structure of CghA in complex with 1.

a, Overall structure of CghA and the bound product **1**. **b**, Stereo view of the active site. Carbon atoms of the protein side chains and **1** are in green and yellow, respectively. Oxygen, nitrogen and sulfur atoms are in red, blue and gold, respectively. Dashed lines represent hydrogen bonds. **c**, Comparison of the CghA (green)–*endo* adduct (yellow) complex and a modelled A242S/M257V/V391L mutant (magenta)–*exo* adduct (cyan) complex. **d**, Conformational differences between the apo (purple) and complex (green) structures at W183, W235 and K352.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

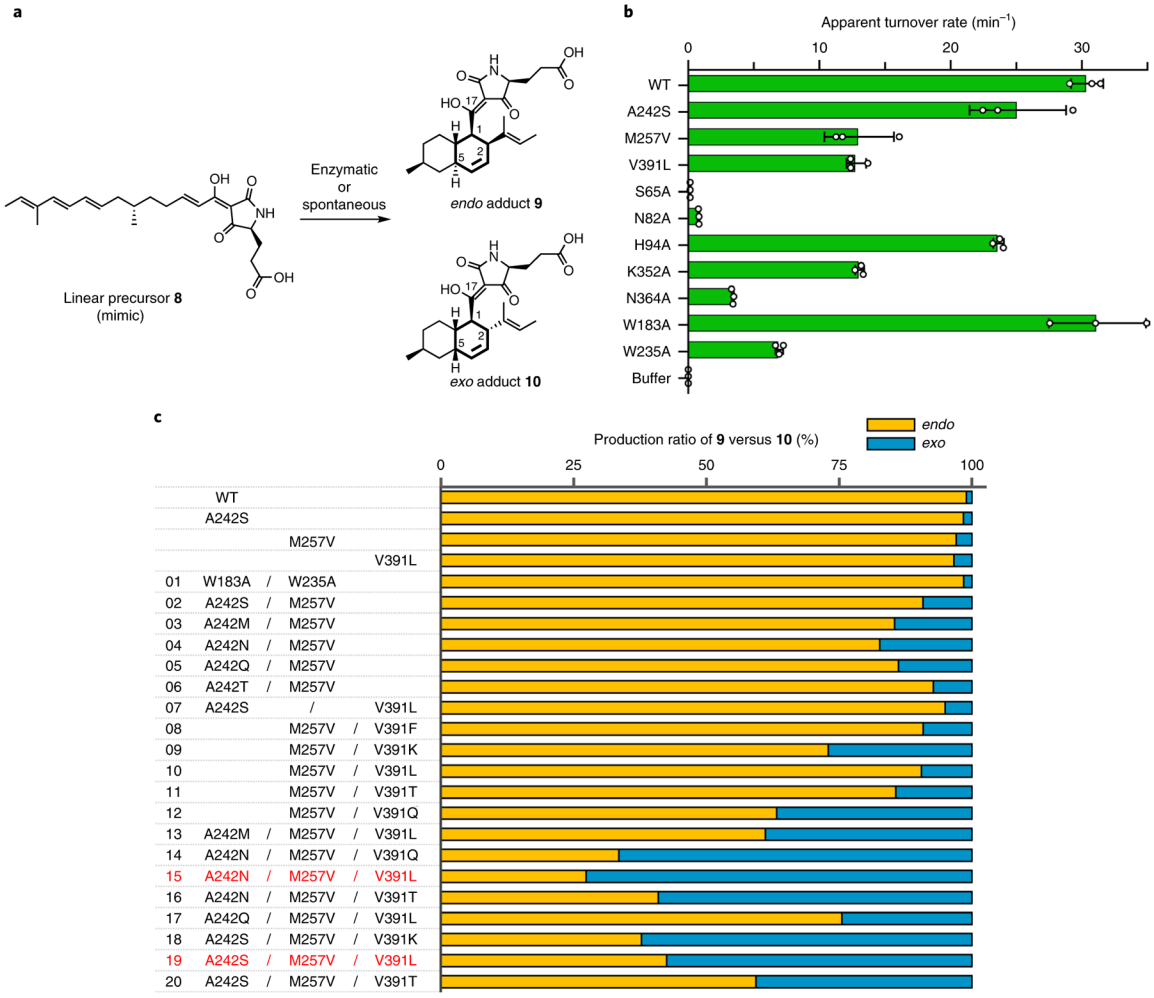


Fig. 3 |. In vitro assays of the CghA mutants.

a, The expected *endo* **9** and *exo* **10** adducts formed from the substrate analogue **8** by the spontaneous or enzyme-catalysed Diels–Alder reaction. **b**, Apparent turnover rates $k_{cat(app)}$ for the conversion of **8** to the adducts by the mutants for examining the role of active-site residues. The centre value is the mean of triplicate measurements (open circles) with the error bar representing the standard deviation. **c**, relative production levels of **9** versus **10** by different mutants for exploring the roles of octalin-packing residues. See Methods and Supplementary Note 5 for details.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

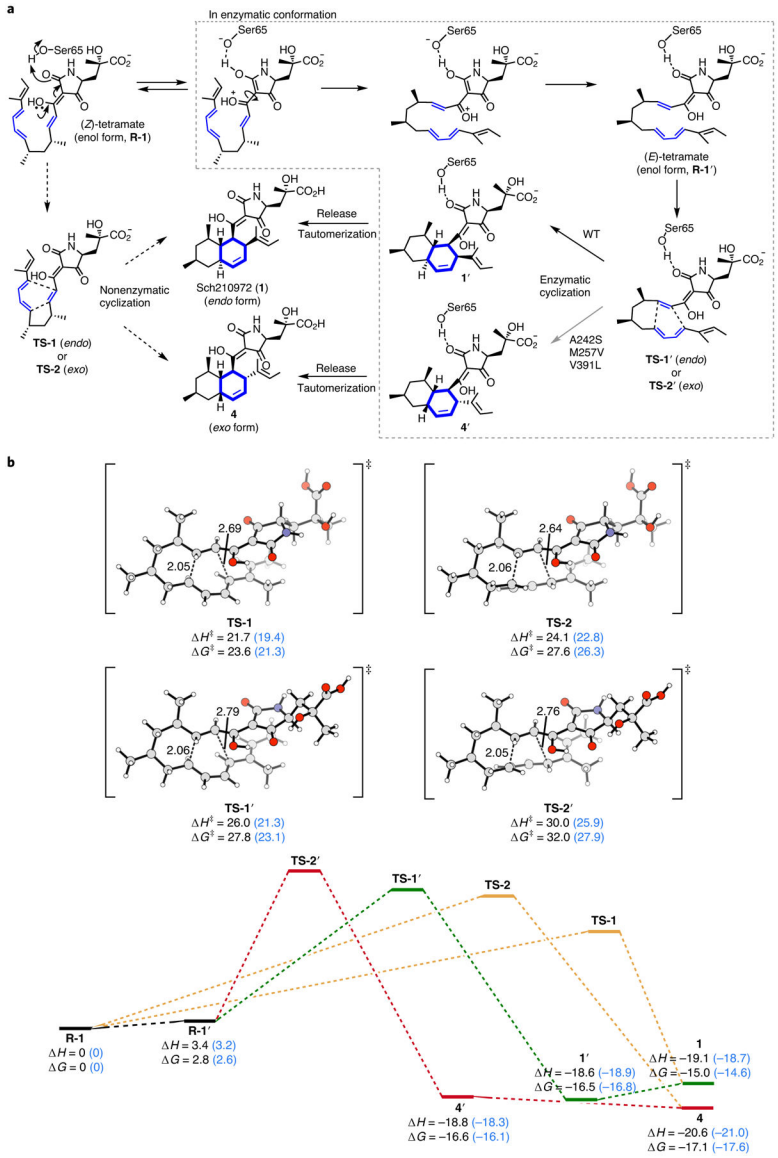


Fig. 4 |. Computational analyses of the transition states and energetics of the octalin-forming intramolecular Diels–Alder reactions.

a, Steps in the CghA-catalysed reaction leading to *endo* **1** and *exo* **4** adducts from enol-form (*Z*)-tetramate **R-1** via enol-form (*E*)-tetramate **R-1′** formed upon rotation of the tetramate moiety. **TS-1** and **TS-2** are the transition states for the *endo* and *exo* pathways, respectively. The non-enzymatic reactions occur directly from **R-1** via **TS-1** and **TS-2**. **b**, The computed transition state structures and the energy diagram of the reaction pathways. Energies were calculated in the gas phase (black) and using the CPCM implicit water model (blue).