

Structural basis for olivetolic acid formation by a polyketide cyclase from *Cannabis sativa*

Xinmei Yang¹, Takashi Matsui¹, Takeshi Kodama¹, Takahiro Mori², Xiaoxi Zhou¹, Futoshi Taura³, Hiroshi Noguchi⁴, Ikuro Abe^{2*} & Hiroyuki Morita^{1*}

¹Institute of Natural Medicine, University of Toyama, 2630-Sugitani, Toyama 930-0194, Japan.

²Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

³Graduate School of Medicine and Pharmaceutical Sciences, Faculty of Pharmacy and Pharmaceutical Sciences, University of Toyama, 2630-Sugitani, Toyama 930-0194, Japan.

⁴School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga, Shizuoka 422-8526, Japan.

Address correspondence to Hiroyuki Morita (Institute of Natural Medicine, University of Toyama, 2630-Sugitani, Toyama 930-0194, Japan. TEL: +81-76-434-7625; FAX: +81-76-434-5059, E-mail: hmorita@inm.u-toyama.ac.jp), or Ikuro Abe (Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. TEL: +81-3-5841-4740; FAX: +81-3-5841-4744, E-mail: abei@mol.f.u-tokyo.ac.jp).

Running title: Crystal structure of olivetolic acid cyclase

Abbreviations: ACP, acyl carrier protein; AtHS1, *Arabidopsis thaliana* heat stable protein; Betv1, *Betula verrucosa* birch pollen allergen; DABB, dimeric $\alpha+\beta$ barrel; GST, Glutathione *S*-transferase; IPTG, isopropyl- β -D-thiogalactopyranoside; MLMI, 4-methylmuconolactone methylisomerase; NAC, N-Acetylcysteamine; NTF-2, nuclear transport factor 2; OA, olivetolic acid; OAC, olivetolic acid cyclase; PF, Photon Factory; PKS, polyketide synthase; SAD, single-wavelength anomalous diffraction; Se-Met, selenomethionine; SP1, *Populus tremula* boiling stable protein; Tcm ARO/CYC, tetracenomycin aromatase/cyclase; TcmI, tetracenomycin F2 cyclase; TKS, tetraketide synthase.

Enzyme EC numbers: Olivetolic acid cyclase, EC 4.4.1.26; tetraketide synthase, EC 2.3.1.206.

Data deposition: Structural data reported in this paper are available in the Protein Data Bank under the accession numbers **5B08** for the OAC apo, **5B09** for the OAC-OA binary complex, and **5B0A**, **5B0B**, **5B0C**, **5B0D**, **5B0E**, **5B0F**, and **5B0G** for the OAC His5Q, Ile7F, Tyr27F, Tyr27W, Val59M,

Tyr72F, and His78S mutant enzymes, respectively.

Keywords: crystal structure, enzyme mechanism, olivetolic acid cyclase, *Cannabis sativa*, DABB protein

Conflict of interest: The authors declare that they have no conflicts of interest.

Abstract

In polyketide biosynthesis, ring formation is one of the key diversification steps. Olivetolic acid cyclase (OAC) from *Cannabis sativa*, involved in cannabinoid biosynthesis, is the only known plant polyketide cyclase. In addition, it is the only functionally characterized plant $\alpha+\beta$ barrel (DABB) protein that catalyzes the C2-C7 aldol cyclization of the linear pentyl tetra- β -ketide CoA as the substrate, to generate olivetolic acid (OA). Herein, we solved the OAC apo and OAC-OA complex binary crystal structures at 1.32 and 1.70 Å resolutions, respectively. The crystal structures revealed that the enzyme indeed belongs to the DABB superfamily, as previously proposed, and possesses a unique active-site cavity containing the pentyl-binding hydrophobic pocket and the polyketide binding site, which have never been observed among the functionally and structurally characterized bacterial polyketide cyclases. Furthermore, site-directed mutagenesis studies indicated that Tyr72 and His78 function as acid/base catalysts at the catalytic center. Structural and/or functional studies of OAC suggested that the enzyme lacks thioesterase and aromatase activities. These observations demonstrated that OAC employs unique catalytic machinery utilizing the acid/base catalytic chemistry for the formation of the precursor of OA. The structural and functional insights obtained this work thus provide the foundation for analyses of the plant polyketide cyclases that will be discovered in the future.

Introduction

Polyketide cyclases, such as *Streptomyces nogalater* SnoaL [1, 2] (Fig. 1A), *S. glaucescens* tetracenomycin aromatase/cyclase (Tcm ARO/CYC) [3, 4] (Fig. 1B), and *S. glaucescens* tetracenomycin (Tcm) F2 cyclase (TcmI) [5, 6] (Fig. 1C), are key enzymes that diversify the structures of pharmaceutically and biologically important natural polyketides. The enzymes catalyze aldol cyclization(s) to generate new cyclic or aromatic ring system(s) of the polyketides, such as nogalamycin [1] and Tcm C [3, 5]. The polyketide cyclases are structurally divided into three types, nuclear transport factor (NTF2)-like polyketide cyclase, *Betula verrucosa* birch pollen allergen (Betv1)-like polyketide cyclase, and dimeric $\alpha+\beta$ barrel (DABB)-type polyketide cyclase. The crystal structures of the NTF2-like polyketide cyclase SnoaL, involved in nogalamycin biosynthesis, and the Betv1-like polyketide cyclase Tcm ARO/CYC, involved in Tcm C biosynthesis, revealed the detailed enzyme reaction mechanisms, as well as the active-site architectures [2, 4]. Further structure-based mutagenesis studies suggested that SnoaL utilizes the single catalytic residue Asp121 as the acid/base catalyst for the C10-C9 aldol cyclization of nogalonic acid methyl ester as the substrate, to generate nogalaviketone. In contrast, TcmARO/CYC utilizes the catalytic triad Tyr35-Arg69-Arg82 and a water molecule as the acid/base catalysts for the C14-C9 and C16-C7 aldol cyclizations and aromatization of the acyl carrier protein (ACP)-tethered acetyl- β -deca ketide as the substrate, to generate the precursor of ACP-tethered Tcm F2. The crystal structure of the DABB-type polyketide cyclase TcmI, involved in the biosynthesis of Tcm C, has also been characterized. However, the detailed mechanism of the C9-C10 aldol cyclization of Tcm F2, for the formation of Tcm F1, has remained elusive [6].

The recently identified olivetolic acid cyclase (OAC) from *Cannabis sativa* is the only known plant polyketide cyclase that has been proposed to be involved in the biosynthesis of cannabinoid [7]. The *in vitro* functional analysis, using the recombinant OAC and the recombinant *C. sativa* type-III polyketide synthase (PKS), tetraketide synthase (TKS), suggested that OAC accepts a

linear pentyl tetra- β -ketide CoA as the substrate, without interacting with TKS in a similar manner to chalcone reductase [8], and performs the C2–C7 aldol cyclization, thioester bond cleavage, and aromatization reactions to generate olivetolic acid (OA), without requiring any co-factors (Fig. 1D) [7]. The enzyme is a homodimeric protein, with each subunit consisting of 101 amino acid residues [7, 9], and shares 30–50% identity with the other plant DABB superfamily proteins, such as the structurally characterized and functionally uncharacterized heat stable protein (AtHS1) from *Arabidopsis thaliana* [10, 11], boiling stable protein (SP1) from *Populus tremula* [12], and At5g22580 from *A. thaliana* [13]. In contrast, the enzyme exhibits low identity (less than 20%) with the bacterial DABB proteins (Fig. 2), such as TcmI [6], the functionally distinct, structurally characterized ActVA-Orf6 monooxygenase from *S. coelicolor* [14], and 4-methylmuconolactone methylisomerase (MLMI) from *Pseudomonas reinekei* [15]. The structure modeling study of OAC suggested that OAC is a member of the DABB-type polyketide cyclases and possesses a hydrophobic tunnel as the active-site cavity in each monomer [7], as in the cases of the other structurally characterized DABB proteins. Furthermore, site-directed mutagenesis studies suggested that three His residues (His5, His57, and His78) play crucial roles in the OA-forming activity [7]. Thus, OAC is likely to employ acid/base catalytic chemistry for the formation of OA, as proposed for the structurally and functionally distinct SnoaL and Tcm ARO/CYC from OAC [2, 4]. However, the catalytic mechanism underlying the substrate and product specificities, the aldol cyclization and aromatization reactions, and the thioester bond cleavage, as well as the catalytic role of the three His residues in the OA-forming activity, have remained unclear.

Herein, we present the crystal structure of OAC, the only known plant polyketide cyclase. The OA complex binary crystal structure of OAC clearly revealed the active-site architecture, containing a 9 Å long pentyl moiety binding pocket and a putative polyketide binding site. Furthermore, site-directed mutagenesis studies allowed us to identify Tyr72/His78 as the catalytic residues for the formation of OA. Our crystallographic analysis of OAC provides insight into the

structural basis for the catalytic mechanism of this plant polyketide cyclase.

Results

Overall structure of OAC

The apo crystal structure of OAC was also solved by the Se-SAD method and refined at 1.32 Å resolution. The asymmetric unit contains two monomers, which further form a dimer (Fig. 3A). Significant conformational differences between monomers A and B were observed in the residues 65-88, with root mean square (r. m. s.) deviation values of 1.9 Å for all C α atoms relative to each other (Fig. 3B). The monomer A consists of a four-stranded antiparallel β -sheet and three α -helices ($\alpha 1$ - $\alpha 3$), while the monomer B consists of a four-stranded antiparallel β -sheet and two α -helices ($\alpha 1'$ and $\eta 2'$). The outer surface of the antiparallel β -sheets faces to each other and form a central α + β barrel core. Upon dimerization, each monomer buries approximately 1290 Å² of the surface, corresponding to 19.5% of the total surface area.

In contrast, the OAC and OA binary complex structure was solved by the molecular replacement method, and refined at 1.70 Å resolution. The asymmetric unit of the binary complex crystal contains a monomer, which further forms a biologically active symmetric dimer with a crystallographic two-fold axis (Fig. 3C, D). As in the case of monomer A in the OAC apo structure, the monomer of the binary complex consists of a four-stranded antiparallel β -sheet and three α -helices ($\alpha 1$ - $\alpha 3$) (Fig. 3A, C). The outer surfaces of the antiparallel β -sheets face each other and form a central α + β barrel core. Upon dimerization, each monomer buries approximately 1,420 Å² of the surface, corresponding to 22.3% of the total surface area. Several hydrogen bonds and hydrophobic interactions are involved in the dimerization (Fig. 3D). The electron density in the binary complex structure of OAC clearly indicated the presence of the OA molecule in the cavity (Fig. 3C, D, E). Thus, the binary complex structure indicated that OAC possesses an active-site cavity in the interior of the

α + β barrel in each monomer, and the active-site entrance is open at the center of α 2, α 3, and β 4, in a location and orientation similar to those of the hydrophobic cavities observed in AtHS1, SP1, and At5g22580 [10–13] (Fig. 4A–D).

A comparison of the overall structure of OAC apo and the binary complex revealed that the structure of OAC-OA binary complex is nearly identical to the structure of monomer A of the apo structure, with a root-mean-square r. m. s. deviation value of 0.7 Å (Fig. 3F). There are no significant backbone changes between the monomer A and OAC-OA binary complex structures, except for residues 47-54 in the β 2- β 3 loop, in which residues 51-53 are disordered in the binary complex structure (Fig. 3F). However, the comparison of the overall structure of the binary complex and monomers A and B of the OAC apo indicated that the monomer B shows an artificial structure, due to great influence of the crystallographic packing. In particular, the structural conversion of the α 2 and α 3 (residues 65-73 and 76-88, respectively) and an α 2- α 3 loop (residues 74-75) in monomer A into β 3- η 2' and η 2'- β 4 loops (residues 65-69 and 74-88, respectively) and a one-turn helix (η 2', residues 70-73) in monomer B by the crystallographic packing resulted in the disappearance of the active site cavity from monomer B (Fig. 3B). Due to the high structural conservation between monomer A and the binary complex structures and the artificial structure of monomer B, further crystal structure analysis of OAC was carried out by using the structure of the OAC-OA binary complex.

The three-dimensional structure of OAC adopts almost the same overall structure observed in the plant and bacterial DABB proteins. A structure-based similarity search revealed that the overall structure of the OAC-OA binary complex exhibits r. m. s. deviation values of 1.1 Å, 1.4 Å, and 2.5 Å with the plant DABB proteins, AtHS1 (PDB code **1Q4R**, 48% identity with OAC) [10, 11], SP1 (PDB code **1TR0**, 38% identity with OAC) [12], and At5g22580 (PDB code **1RJJ**, 32% identity with OAC) [13], and of 1.8 Å, 2.2 Å, 2.3Å, 2.4 Å, 2.4, and 2.9 Å with the bacterial DABB proteins, *S. nogalater* SnoaB (PDB code **3KG1**, 10% identity with OAC) [16], *Mycobacterium tuberculosis* Rv0793 (PDB

code **1Y0H**, 7% identity with OAC) [17], MLMI (PDB code **3HDS**, 13% identity with OAC) (15), ActVA-Orf6 (PDB code **1N5S**, 15% identity with OAC) [14], *P. aeruginosa* PA3566 (PDB code **1X7V**, 7% identity with OAC) [18], and TcmI (PDB code **1TUW**, 17% identity with OAC) [6], respectively (Figs. 2 and 4).

Active-site architecture of OAC

The active-site cavity of OAC is created with His5, Ile7, Leu9, Phe23, Phe24, Tyr27, Val28, Leu30, Val40, Val59, Tyr72, Ile73, His78, Phe81, Gly82, Trp89, Leu92, and Ile94 (Fig. 5A). A 9-Å long, narrow hydrophobic tunnel, named the pentyl-binding pocket, is located deep inside the active-site cavity. The side chains of the previously proposed crucial residues, His5 and His78, create a hydrophilic region on the tunnel wall near the entrance, together with the hydroxy groups of Tyr27 and Tyr72. The side chains of His5 and His78 also participate in the formation of the active-site entrance, together with Ile73, Gly82, Trp89, Leu92, and Ile94. In contrast, Asp96 is located just outside the active-site cavity, although the Ala substitution of this residue was previously reported to decrease the OA-forming activity [7]. The estimated total volume of the active-site cavity is 270 Å³, which is large enough to accommodate the pentyl moiety and part of the tetra-β-ketide moiety of the substrate. The OA binding binary structure lacked other large cavities or tunnels, suggesting that the CoA portion of the substrate is attached on the protein surface or exposed to the solvent, and does not bind to the enzyme. Furthermore, His57, reported as the crucial residue for the OA-forming activity, is located far from the pocket and protrudes toward the protein surface (Fig. 3D). Lys4, Lys12, Lys38, Asp45, and His75, which exhibited decreased OA-forming activities upon substitutions with Ala [7], are also located far from the pocket and protrude toward either the protein surface or the dimerization interface (Fig. 3D). These residues are thus proposed to be important for the protein folding or the CoA-binding, rather than the catalytic functions.

The crystal structure of the binary complex of OAC and OA clearly indicated that the pentyl moiety of the OA molecule is accommodated in the pentyl-binding pocket (Fig. 5A). The pentyl-binding pocket is thus filled by the pentyl moiety. In contrast, the dihydroxybenzoate moiety of OA is located near the entrance of the active-site cavity. Two hydroxy groups, one of the carboxyl oxygens, and part of the aromatic ring of the dihydroxybenzoate moiety are thus exposed to the solvent (Fig. 5A). The other carboxyl oxygen forms hydrogen bonds with the side chains of His5 and Tyr72, and His5 makes another hydrogen bond with Asp96, in a location outside the active-site cavity. Tyr72 makes two more hydrogen bonds with the side chains of Tyr27 and His78, located deep inside the active-site cavity and near the active-site entrance, respectively. Notably, the side chain of His78 protrudes toward the C1 and C6 positions of the aromatic ring of OA, corresponding to the C2 and C7 positions of the aldol cyclization point on the substrate. The N ϵ 2 atom of the His78 side chain is thus located at 3.9 Å and 4.3 Å distances from the C1 and C6 positions, respectively, and is closer than the 4.8 Å and 5.2 Å distances between the hydroxy group of Tyr72 and the C1 and C6 positions, respectively. Furthermore, no metal ions and water molecules were observed in the active-site cavity. These observations indicated that the His5/Tyr27/Tyr72/His78/Asp96 arrangement and the pentyl-binding pocket play crucial roles in controlling the substrate and product specificities of OAC.

Site-directed mutagenesis of OAC and crystal structure analyses of selected OAC mutant enzymes

To further clarify the catalytic mechanism of OAC, we constructed site-directed mutants of the hydrophilic and selected key hydrophobic residues lining the active-site cavity, and investigated the mechanistic consequences of the point mutations. We also constructed the previously reported His5A and His78A mutants [7], to reassess the effects of the mutations. Thus, all mutants were overexpressed in *E. coli*, as GST-fused proteins, and after the GST-tag was removed, the mutant enzymes were used

for the enzyme reactions. Among them, the mutant enzymes His5A/D/N, Leu9A/W, Phe23A/I/L/W, Phe24A/W, Val28F, Val40F, Val59A/F, and His78A/D, including the His5A and His78A mutants, formed inclusion bodies in *E. coli*. Thus, these mutants were not analyzed in our study.

We performed the OAC mutant enzyme and TKS co-incubation assay to assess the effects of the mutations, since the linear pentyl tetra- β -ketide-CoA is not available. Thus, the effects of the mutations were evaluated on the basis of related activities of the obtained mutant enzymes for the formation of OA, as compared with that of wild-type. The HPLC profiles and the related activities are summarized in Table 1 and Fig. 6. A co-incubation of the OAC His78S mutant and TKS with hexanoyl-CoA and malonyl-CoA as the substrates revealed that OA is no longer produced by the mutant enzyme. The LC-MS analysis thus detected only the hexanoyl triacetic acid lactone and pentyl triketide pyrone that are produced by TKS from condensation of three and two molecules of malonyl-CoA with the hexanoyl-CoA, respectively in the reaction mixture. This suggested that, as in the case of the previously reported His78A mutant enzyme [7], the His78S mutant completely lost the enzyme activity. As in the case of the His78S mutants, the set of the His78N/Q mutants lost the enzyme activity. Remarkably, the Tyr72F substitution also abolished the enzyme activity. The crystal structure analysis of the His78S mutant enzyme revealed that the His78S substitution only disrupts the hydrogen bond between Tyr72 and His78 at the active-site (Fig. 7A, B). The crystal structure of the Tyr72F mutant enzyme also indicated that the side chain of Phe72 rotates approximately 17° toward the α 1 helix, in comparison with that of Tyr72 in the wild-type, thereby disrupting the hydrogen bond network between Tyr72 and His78 at the active-site (Fig. 7A, C). However, significant conformational changes between the OAC wild-type and Tyr72F mutant enzyme structures were not observed in the other active site residues. These results suggested that the hydroxy group of Tyr72 plays an important role in elevating the basicity of His78.

However, in contrast to the previous report for the His5A mutant enzyme [7], the His5Q,

His5S, and His5L substitutions retained the enzyme ability and exhibited 78%, 53%, and 64% decreases of the OA-forming activity, respectively. The crystal structure of the His5Q mutant enzyme demonstrated that the His5Q substitution did not affect the conformations of any of the active-site residues and maintains the hydrogen bonds between His78, Tyr72, and Tyr27 (Fig. 7A, D). Remarkably, Gln5 in the mutant enzyme and His5 in the wild-type are superimposable, with nearly identical positions in those structures, and retained the hydrogen bond between the side chains of Gln5 and Asp96. The only significant difference in the mutant enzyme is thus regarded as the acidity of Gln5, suggesting that the substitution solely disrupts the approximate contact between His5 and the carboxyl group of OA. A comprehensive assessment of the results of the site-directed mutagenesis and structural studies suggested that the set of His5S/L substitutions hinders the substrate binding, but does not affect the conformations of the other residues. His5 is thus proposed to be more crucial for the substrate binding affinity, rather than playing the role of an acid/base catalyst. Presumably, the Asp96A substitution in the previous report shifted the location and/or orientation of His5, thereby disrupting the approximate contact of His5 with the β -keto moiety of the substrate, and leading to the previously reported decrease in the OA-forming activity [7].

Interestingly, despite the lack of a hydroxy group, the Tyr27F substitution displayed a 62% increase in the OA-forming activity. In contrast, the Tyr27M, Tyr27W, and Tyr27L substitutions showed 70%, 63%, and 52% decreases in the OA-forming activity, respectively. The aromatic ring of Tyr27 not only forms the hydrogen bond network with His78 and Tyr72, but also is involved in the binding of the pentyl moiety of OA in the OAC-OA binary complex structure. Therefore, the substitution may change the size, shape, and hydrophobicity of the active-site cavity, as reflected in the observed activities of the mutant enzymes. As expected, the crystal structure of the Tyr27F mutant enzyme indicated that the Tyr27F substitution disrupts the hydrogen bond between Tyr27 and Tyr72 (Fig. 7A, E) observed in the wild-type, but maintains all of the conformations of the active-site

residues in the mutant enzyme, in locations and orientations very similar to those in the wild-type. The effect of the loss of the hydroxy group of Tyr27 on the hydrogen bond network at the active-site is thus regarded to increase both the hydrophobicity of the pentyl-binding pocket and the electrostatic interaction between Tyr72 and His78, which enhance the OA-forming activity. Conversely, the crystal structure of the Tyr27W mutant enzyme demonstrated that the Tyr27W substitution rearranges the hydrogen bond networks with Tyr27, Tyr72, and His78 in the wild-type to those of Trp27, Tyr72, and His78, but also results in the protrusion of the Trp27 side chain into the bottom of the pentyl-binding pocket, thereby preventing the access of the pentyl moiety of the substrate (Fig. 7A, F).

Similar cases of steric constraint were also observed in our site-directed mutagenesis studies of the pentyl-binding pocket and the crystal structures of the Ile7F and Val59M mutant enzymes. In line with these mutants, the Phe24L, Val59M, Ile7F, and Ile7L substitutions showed 51%, 35%, 36%, and 15% decreases in the OA-forming activity. The structures of the Ile7F and Val59M mutant enzymes indicated that the Ile7F and Val59M substitutions narrow the pentyl-binding pocket by constricting the bottom of the cavity, but maintain the hydrogen bonds between His78, Tyr72, and Tyr27, and between His5 and Asp96 (Fig. 7A, G, H). The Ile7F, Tyr27W, and Val59M mutant enzymes exhibited estimated total cavity volumes of 247 Å³, 204 Å³, and 222 Å³, which are slightly less than the 270 Å³ volume of the wild-type OAC. These observations strongly favor the possible catalytic roles of Tyr72 and His78 over their roles in the substrate binding affinity, and the crucial roles of His5 and the pentyl-binding pocket in the binding affinities of the polyketide and the pentyl moiety of the substrate, respectively. OAC is thus suggested to employ only acid/base chemistry for the C2-C7 aldol cyclization of the substrate, in a similar manner to the structurally and functionally distinct enzymes SnoaL and Tcm ARO/CYC [2, 4].

Thioesterase and aromatase activities of OAC

As mentioned above, OAC has also been suggested to catalyze the thioester bond cleavage and aromatization reactions to generate OA. However, no residues, metal ions, or water molecules that may be involved in the thioester bond cleavage and aromatization were observed in the OAC-OA binary complex structure. This suggested that OAC lacks both the thioesterase and aromatase activities. To test the thioesterase activity of OAC, we performed an enzyme reaction of OAC using OA thioester linked *N*-acetyl cysteamine (OA-NAC) as the possible assay system, since the linear pentyl tetra- β -ketide- and cyclized intermediate-CoAs, -NACs, corresponding acids, and CoA-linked OA thioester are not available. NAC and NAC-linked acyl-thioester analogues have been successfully utilized as the substrate analogues of the functionally and mechanistically distinct, CoA-SH or CoA thioester-requiring other type of enzymes such as the type III polyketide synthases and the CoA ligases to investigate their functions, since NAC is the terminal portion of CoA-SH [19-22]. However, the LC-MS analysis did not reveal any differences in the production of OA in the reaction mixture of OAC with OA-NAC, as compared with that in the control experiments (Fig. 8). The result may be account for lacking the hydrolysis activity of OAC, although a possibility of the result derived from the structural differences between CoA-linked and NAC-linked OAs is not excluded. Furthermore, a docking studies of the OAC substrate, pentyl tetra- β -ketide-CoA into the OAC structure suggested that His78 may be located near the C2 carbon and C7 carbonyl oxygen of the aldol cyclization point on the substrate, and Tyr72 recognizes the side-chain of His78 and thioester carbonyl oxygen of the substrate (Fig. 9A). In addition, a docking simulation of the cyclized OAC-intermediate into the OAC structure suggested that the side chain of His78 may be located near the carbonyl oxygen of the cyclized intermediate, corresponding to that of C7 position of pentyl tetra- β -ketide-CoA (Fig. 9B). However, the docking models did not show any residues that can activate water molecules or metal ions to cleave the thioester bond of the substrate and abstract protons of the cyclic ring on the cyclized intermediate. These observations may suggest that OAC lacks the thioesterase and aromatase activities.

Discussion

Our crystal structure and site-directed mutagenesis studies of OAC revealed that OAC possesses the unique active-site architecture, as well as catalytic roles of Tyr72 and His78 as acid/base catalysts at the catalytic center. Furthermore, structural and/or functional studies of OAC suggested that the enzyme lacks thioesterase and aromatase activities. On the basis of these observations, we propose two elements for the OAC catalytic mechanism. One is that the catalytic function of OAC is to form the CoA-linked cyclized product, corresponding to the precursor of OA. The other is that OAC employs the acid/base catalysis chemistry, facilitated by the presence of ~~the proton-exchangeable~~ Tyr72 and His78, for the formation of the precursor of OA.

Thus, after pentyl tetra- β -ketide CoA is loaded into the active-site cavity, (i) OAC facilitates a nucleophilic attack by ~~deprotonated~~ His78 ~~activated by~~ Tyr72 to abstract the proton at the C2 of the substrate, and produces an enol (or enolate) intermediate (Fig. 10A). Subsequently, (ii) the nucleophilic attack of the enol (enolate) intermediate on the C7 carbonyl carbon by the keto/enol tautomerization and (iii) the sequential proton abstraction from the protonated His78 by the C7 carbonyl oxygen proceed to facilitate the C2-C7 aldol cyclization, ~~with the restoration of His78~~ (Fig. 10B, C). Finally, (iv) the enzyme releases the CoA-linked cyclized product, which is immediately subjected to aromatization and CoA bond cleavage to form OA, in a non-enzymatic and spontaneous manner (Fig. 10D, E).

Interestingly, despite the lack of OA-forming activity as previously reported, a comparison of the crystal structure of OAC with other DABB proteins revealed that the structurally closest, functionally uncharacterized AtHS1 conserves the His5/Tyr72/His78/Asp96 arrangement in the structure, in a location and orientation very similar to those in OAC (Fig. 5A, B). Most of the other residues, including Tyr27 of OAC, are also superimposable in nearly identical positions in AtHS1. The

difference between both proteins is only the shape and size of the active-site cavity. This is basically due to the distinctive sequence of Leu15, Ala17, Leu31, Phe48, Phe93, Leu94, Leu97, Val100, and Val102 in AtHS1 (Fig. 2). Thus, the side-chains of Phe93 and Leu94 in AtHS1 uniquely participate in the formation of the hydrophobic tunnel of AtHS1, instead of Phe23 and Trp89 in OAC. The side-chains of Leu15 and Tyr35 of AtHS1, corresponding to Ile7 and Tyr27 in OAC, also protrude toward the tunnel, as compared to those of OAC. Thereby, the hydrophobic pocket of AtHS1 has a different shape and size from the active-site cavity of OAC. According to these observations, we propose that AtHS1 is another polyketide cyclase that accepts an acyl poly- β -ketide CoA as the substrate to generate another plant polyketide, by employing the same reaction mechanism as that of OAC.

However, a comparison of the crystal structures of OAC and TcmI indicated that the active-site architecture of OAC is conformationally different from that of TcmI, although both enzymes share the same overall folding and possess the active-site cavity in the interior of the α + β barrel. Notably, the Tyr72/His78 arrangement is not conserved in the structure of TcmI (Fig. 5A, C). Furthermore, site-directed mutagenesis studies of TcmI also suggested that the substitutions of His26, Arg40, and His51, located deep inside the active-site cavity, only decrease the Tcm F1-forming activity. These findings suggested that TcmI employs a distinct catalytic mechanism from that of OAC for the C9-C10 aldol cyclization of Tcm F2, to produce Tcm F1. Gagne and co-workers proposed that OAC and the bacterial DABB-type cyclases are an example of convergent evolution, with the polyketide cyclizing activity arising independently in plants and bacteria [7]. The crystal structure of OAC also supports this hypothesis.

In conclusion, our structural and functional analyses revealed the intimate structural details of the catalytic mechanism of OAC, for the formation of OA. The mechanism is facilitated by the acid/base catalytic chemistry, as suggested in the similar catalytic mechanisms of the structurally and functionally distinct bacterial polyketide cyclases, SnoaL and Tcm ARO/CYC [2, 4]. Interestingly, the

catalytic Tyr72 and His78 observed in OAC are conserved in homologous proteins widely distributed in plants, such as POP3 in *Medicago sativa* and the POP3-like proteins deposited with unknown functions, as in the case of AtHS1. It is thus assumed that the proteins also possess hydrophobic tunnels equivalent to the pentyl-binding pocket of OAC, and may function as polyketide cyclases. Furthermore, recent studies have suggested that the biosyntheses of plant polyketides, such as the anthranoid produced by *Aloe arborescens* [23, 24], require additional enzymes for the proper folding and cyclization of the linear poly- β -keto intermediate to generate the final products, as in the cases of cannabinoid and bacterial polyketide biosyntheses. The present results provide insights not only into the functional diversity of the DABB family enzymes, but also the structural basis for analyses of plant polyketide cyclases that will be discovered in the future.

Materials and methods

Materials

OA was synthesized from olivetol, as previously reported [25]. OA-NAC was synthesized from OA, according to the synthesis of benzoyl-NAC with some modifications [26]. Standard chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA), Wako Pure Chemical Industry (Osaka, Osaka, Japan), Hampton Research (Aliso Viejo, CA, USA), and QIAGEN (Germantown MD, USA). The pET22b expression plasmid encoding TKS was purchased from Eurofins Genomics (Tokyo, Tokyo, Japan).

Expression and purification of Non- and SeMet-labeled wild-type OACs

The non-labeled recombinant OAC, with a GST-tag at the N-terminus, was overexpressed in *Escherichia coli* M15 and purified, as previously reported [9]. Se-Met labeled OAC was overexpressed in *E. coli* M15, grown in M9 medium containing 1 mM MgSO₄, 0.1 mM CaCl₂, 1 × BME vitamins solution (Sigma-Aldrich), 0.4%(w/v) glucose, and 100 µg/mL ampicillin, at 37 °C until the OD₆₀₀ reached 0.6. For the expression of Se-Met labeled OAC, 25 mg/L selenomethionine, 100 mg/L each of L-Lys, L-Thr, L-Ile, L-Leu, L-Val, and L-Phe, and 1.0 mM isopropyl-β-D-thiogalactopyranoside (IPTG) were then added, and the cells were cultured further for 24 hours at 17 °C. The SeMet-labeled OAC was purified using the same procedure as that for the non-labeled recombinant OAC [9].

Site-directed mutagenesis and purification and crystallization of mutant enzymes

All of the mutant enzymes were constructed with a QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA), according to the manufacturer's protocol. The mutant enzymes were expressed, extracted, and purified by the same procedure used for the non-labeled wild-type OAC [9].

Crystallization and structure determination

All crystallization experiments were performed using the sitting-drop vapor-diffusion technique, with incubation at 5 °C. Diffraction quality Se-Met labeled and native wild-type OAC apo crystals were obtained as previously reported [9]. Crystals of the OAC-OA complex were obtained under conditions with 100 mM Tris-HCl (pH 8.8) and 25%(w/v) PEG6000, by co-incubating 25 mg/mL OAC with 50 mM OA. Diffraction quality crystals of the OAC H5Q, I7F, Y27F, Y27W, V59M, Y72F, and H78S mutants were obtained in the optimized reservoir conditions listed in Table 2, after initial screening in the same manner as that for the wild-type apo crystals. All crystals were transferred into a cryoprotectant solution, consisting of each crystallization solution with 10%(v/v) glycerol. After a few seconds, the crystals were picked up in a nylon loop and then flash-cooled at -173 °C in a nitrogen-gas stream. Single-wavelength anomalous diffraction (SAD) data of the Se-Met labeled OAC apo crystals and diffraction data of the OAC apo, OAC I7F, Y27F, Y27W, V59M, and H78S mutant crystals were collected on beamline NW-12A, at the Photon Factory (PF). Diffraction data of the OAC-OA binary complex crystals were collected on beamline NE-3A at PF. Diffraction data of the OAC Y72F mutant crystals were collected on beamline BL-17A at PF. The 0.97908 Å wavelength at NW12A was used for the SAD data collection of the OAC crystals, on the basis of the fluorescence spectrum of the Se *K* absorption edge [27]. Diffraction data of H5Q were collected using in-house X-ray diffraction equipment, consisting of a Rigaku MicroMax-7HF high-intensity microfocus rotating anode X-ray generator and an R-Axis VII detector (Rigaku, Akishima, Tokyo, Japan). These diffraction data were processed and scaled with the HKL-2000 program package [28] for the OAC I7F mutant enzyme and XDS [29] for the other enzymes.

The Se sites were determined and refined, and then the initial phase of the Se-Met labeled OAC apo structure was calculated with AutoSol [30] in Phenix [31], and the native structure of apo

OAC with the highest resolution was determined by the molecular replacement method, using the Se-Met labeled OAC structure as a search model, with Molrep [32] in the CCP4 suite [33]. The structure refinement of apo OAC was performed using Phenix.refine [34] with the twin operator $(-h, -k, l)$ at 0.08 of twin fractions. The merohedral twin operator of the apo OAC crystals was found, using Xtrriage [31]. Molecular replacements were performed with the apo OAC structure as the search model, using Phaser [35] in the CCP4 suite [33], to solve the structures of the OAC-OA binary complex and the OAC H5Q mutant enzyme, and Molrep [32] in the CCP4 suite [33] for the other enzymes. The structures were modified manually with Coot [36] and refined with Phenix.refine [34]. The structure refinement of the OAC V59M mutant enzyme was performed using Phenix.refine [34] with the twin operator $(-h, -k, l)$ at 0.03 of twin fractions, after the merohedral twinning of the mutant enzyme crystals was solved using Xtrriage [31]. The detailed data collection and refinement statistics are summarized in Table 2. The quality of the final models was assessed with Molprobity [37]. The Dali program was used for the structure-based similarity search [38]. The active site cavity volume was calculated by the CASTp program (<http://cast.engr.uic.edu/cast/>). All crystallographic figures were created with either the CueMol (<http://www.cuemol.org>) or PyMOL (<http://www.pymol.org>) program.

Expression and purification of TKS

The pET22b vector, encoding full-length TKS [20], was transformed into *E. coli* BL21. The cells harboring the plasmid were cultured to an OD₆₀₀ of 0.6 in LB medium containing 100 µg/mL ampicillin, at 37 °C. IPTG was added to a final concentration of 0.5 mM to induce protein expression, and the culture was incubated for a further 20 h at 20 °C.

All of the following procedures were performed at 4 °C. The *E. coli* cells were harvested by centrifugation at 6,000 g for 20 min, and then resuspended in 50 mM Tris-HCl buffer (pH 8.0), containing 200 mM NaCl and 5% (v/v) glycerol (buffer A). The cells were disrupted by sonication,

and the lysate was centrifuged at 8,000 g for 10 min. The supernatant was loaded on a Ni Sepharose 6 Fast Flow open column (GE Healthcare Life Sciences, Pittsburgh, PA, USA) equilibrated with buffer A. After washing with buffer A containing 10 mM imidazole, the recombinant TKS was eluted with buffer A containing 600 mM imidazole. The resultant protein solution was concentrated to 5 mL, and further purified by gel-filtration chromatography on HiLoad 16/60 Superdex 200 pg (GE Healthcare Life Sciences). Finally, the purified enzyme was concentrated to 26 mg/mL in 20 mM HEPES-NaOH (pH 7.5) buffer, containing 25 mM NaCl and 5% (v/v) glycerol.

Standard enzyme reaction

The reaction mixture contained 60 μ M of hexanoyl-CoA, 100 μ M of malonyl-CoA, 10 μ g of the purified TKS, and 30 μ g of the purified wild-type or mutant OAC enzyme, in a final volume of 500 μ L of 100 mM potassium phosphate buffer (KPB, pH 7.0). The boiled enzymes were used for control experiments. The each experiment was performed in triplicate. Incubations were performed at 20 °C for 16 h, and were stopped by the addition of 50 μ L of 20% HCl. The products were then extracted with 3 mL of ethyl acetate, and were analyzed by an online LC-EIS-MS system with an Agilent Technologies series 1100 HPLC (Santa Clara, CA, USA) coupled to a Bruker Daltonics esquire 4000 ion-trap mass spectrometer fitted with an ESI source (Bruker Daltonik GmbH, Bremen, Germany). HPLC separations were performed on a TSK-gel ODS-80Ts column (4.6 i.d. \times 150 mm, TOSOH, Tokyo, Tokyo, Japan), at a flow rate of 0.9 mL/min. Gradient elution was performed with H₂O and MeOH, both containing 0.1% trifluoroacetic acid: 0–5 min, 30% MeOH; 5–17 min, linear gradient from 30 to 60% MeOH; 17–25 min, 60% MeOH; 25–27 min, linear gradient from 60 to 70% MeOH; and 27–40 min, linear gradient from 70 to 100% MeOH. Elutions were monitored by a multichannel UV detector (MULTI 340, JASCO, Hachioji, Tokyo, Japan), and UV spectra (198–400 nm) were recorded every 0.4 s. The ESI capillary temperature and the capillary voltage were 350 °C and 4.0 V,

respectively. The tube lens offset was set at 20.0 V. All spectra were obtained in the positive mode over a mass range of m/z 50–800, and at a range of one scan every 0.2 s. The collision gas was helium, and the relative collision energy scale was set at 30.0% (1.5 eV). The peak area data of OA at the 260 nm on the HPLC elution profile obtained from each enzyme reaction was employed to derive the percentages of the OA production relative to that of wild-type OAC (average of triplicates \pm SD).

OAC and OA-NAC co-incubation assay

The assay mixture contained 30 μ g OAC and 0.3 nmol OA-NAC, in a final volume of 500 μ L of 100 mM KPB (pH 7.0). The boiled OAC and TKS were used for the control experiments. Incubations were performed at 20 °C for 16 h, and were stopped by the addition of 50 μ L of 20% (v/v) HCl. The products were then extracted with 1 mL of ethyl acetate and were analyzed by an online LC-ESI-MS system, using the same ESI-MS method as described above. HPLC separations were performed on a TSK-gel ODS-80Ts column (4.6 i.d. \times 150 mm, TOSOH), at a flow rate of 0.6 mL/ min. Gradient elution was performed with 25 mM KPB (pH 6.5) and CH₃CN: 0–5 min, 20% CH₃CN; 5–17 min, linear gradient from 20% to 35% CH₃CN; 17–25 min, 35% CH₃CN; 25–27 min, linear gradient from 35% to 40% CH₃CN; and 27–40 min, linear gradient from 40% to 70% CH₃CN.

Docking model of OAC complexed with the substrate and the cyclized intermediate

Three-dimensional structures of pentyl- β -ketide- and its cyclized intermediate-CoAs were generated by the Chem3D Ultra 10 program (CambridgeSoft, Waltham, MA, USA), and their geometries were optimized using elbow in Phenix [31]. The OAC models bound the substrate and cyclized intermediate-CoA were manually built by superimposing pentyl and polyketide thioester moieties of the substrate and intermediate molecules on the electron-density of the OA molecule in the OAC-OA binary complex structure with Coot, respectively. The OAC-substrate and -intermediate binary

complex models were then performed energy minimizations by CNS 1.2 [36], and are used for further analyses.

Acknowledgements

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to HN, IA, and HM), and by grants from the Japan Foundation for Applied Enzymology and the NAGASE Science Technology Foundation (to H.M.). The X-ray diffraction experiments were performed at the Photon Factory, under proposals 13G515, 14G504, 14G521, 14G530, and 15G007.

Author contribution

HN, IA, and HM designed the research; XY, FT, and TK conducted the substrate preparations and the enzyme reactions and analyzed the enzyme reaction products; XY and XZ performed the protein purifications and crystallizations. TM performed the structure refinements of the enzymes. XY, TM, MT, and HM analyzed the crystal structure data; XY, TM, IA, and HM wrote the paper. All authors reviewed the results and approved the final version of the manuscript.

References

1. Torkkell S, Kunnari T, Palmu K, Hakala J, Mäntsälä P & Ylihonko K (2000) Identification of a cyclase gene dictating the C-9 stereochemistry of anthracyclines from *Streptomyces nogalater*. *Antimicrob Agents Chemother* **44**, 396–399.
2. Sultana A, Kallio P, Jansson A, Wang JS, Niemi J, Mäntsälä P & Schneider G (2004) Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. *EMBO J* **23**, 1911–1921.
3. Shen B & Hutchinson CR (1996) Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. *Proc Natl Acad Sci USA* **93**, 6600–6604.
4. Ames BD, Korman TP, Zhang W, Smith P, Vu T, Tang Y & Tsai SC (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides. *Proc Natl Acad Sci USA* **105**, 5349–5354.
5. Shen B & Hutchinson CR (1993) Tetracenomycin F2 cyclase: Intramolecular aldol condensation in the biosynthesis of tetracenomycin C in *Streptomyces glaucescens*. *Biochemistry* **32**, 11149–11154.
6. Thompson TB, Katayama K, Watanabe K, Hutchinson CR & Rayment I (2004) Structural and functional analysis of tetracenomycin F2 cyclase from *Streptomyces glaucescens*: A type II polyketide cyclase. *J Biol Chem* **279**, 37956–37963.
7. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM & Page JE (2012) Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *Proc Natl Acad Sci USA* **109**, 12811–12816.
8. Bomati EK, Austin MB, Bowman ME, Dixon RA & Noel JP (2005) Structural elucidation of chalcone reductase and implications for deoxychalcone biosynthesis. *J Biol Chem* **280**, 30496–30503.

9. Yang X, Matsui T, Mori T, Taura F, Noguchi H, Abe I & Morita H (2015) Expression, purification, and crystallization of a plant polyketide cyclase from *Cannabis sativa*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **71**, 1470–1474.
10. Lytle BL, Peterson FC, Kjer KL, Frederick RO, Zhao Q, Thao S, Bingman C, Johnson KA, Phillips GN Jr & Volkman BF (2004) Letter to the Editor: Structure of the hypothetical protein At3g17210 from *Arabidopsis thaliana*. *J Biomol NMR* **28**, 397–400.
11. Bingman CA, Johnson KA, Peterson FC, Frederick RO, Zhao Q, Thao S, Fox BG, Volkman BF, Jeon WB, Smith DW, Newman CS, Ulrich EL, Hegeman A, Sussman MR, Markley JL, & Phillips GN Jr (2004) Crystal structure of the protein from gene At3g17210 of *Arabidopsis thaliana*. *Proteins* **57**, 218–220.
12. Dgany O, Gonzalez A, Sofer O, Wang W, Zolotnitsky G, Wolf A, Shoham Y, Altman A, Wolf SG, Shoseyov O & Almog O (2004) The structural basis of the thermostability of SP1, a novel plant (*Populus tremula*) boiling stable protein. *J Biol Chem* **279**, 51516–51523.
13. Cornilescu G, Cornilescu CC, Zhao Q, Frederick RO, Peterson FC, Thao S & Markley JL (2004) Letter to the Editor: Solution structure of a homodimeric hypothetical protein, At5g22580, a structural genomics target from *Arabidopsis thaliana*. *J Biomol NMR* **29**, 387–390.
14. Sciara G, Kendrew SG, Miele AE, Marsh NG, Federici L, Malatesta F, Schimperna G, Savino C & Vallone B (2003) The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *EMBO J* **22**, 205–215.
15. Marín M, Heinz DW, Pieper DH & Klink BU (2009) Crystal structure and catalytic mechanism of 4-methylmuconolactone methylisomerase. *J Biol Chem* **284**, 32709–32716.
16. Grocholski T, Koskiniemi H, Lindqvist Y, Mäntsälä P, Niemi J & Schneider G (2010) Crystal structure of the cofactor-independent monooxygenase SnoaB from *Streptomyces nogalater*: Implications for the reaction mechanism. *Biochemistry* **49**, 934–944.

17. Lemieux MJ, Ference C, Cherney MM, Wang M, Garen C & James MN (2005) The crystal structure of Rv0793, a hypothetical monooxygenase from *M. tuberculosis*. *J Struct Funct Genomics* **6**, 245–257.
18. Sanders DA, Walker JR, Skarina T & Savchenko A (2005) The X-ray crystal structure of PA3566 from *Pseudomonas aureginosa* at 1.8 Å resolution. *Proteins* **61**, 209–212.
19. Oguro S, Akashi T, Ayabe S, Noguchi H, & Abe I (2004) Probing biosynthesis of plant polyketides with synthetic N-acetylcysteamine thioesters. *Biochem Biophys Res Commun* **325**, 561–567.
20. Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y & Morimoto S (2009) Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Lett* **583**, 2061–2066.
21. Arora P, Vats A, Saxena P, Mohanty D & Gokhale RS (2005) Promiscuous fatty acyl CoA ligases produce acyl-CoA and acyl-SNAC precursors for polyketide biosynthesis. *J Am Chem Soc* **127**, 9388–9389.
22. Hughes AJ, Keatinge-Clay A. (2011) Enzymatic extender unit generation for in vitro polyketide synthase reactions: structural and functional showcasing of *Streptomyces coelicolor* MatB. *Chem Biol* **18**, 165–176.
23. Abe I, Oguro S, Utsumi Y, Sano Y & Noguchi H (2005) Engineered biosynthesis of plant polyketides: Chain length control in an octaketide-producing plant type III polyketide synthase. *J Am Chem Soc* **127**, 12709–12716.
24. Abdel-Rahman IA, Beuerle T, Ernst L, Abdel-Baky AM, Desoky Eel-D, Ahmed AS & Beerhues L (2013) *In vitro* formation of the anthranoid scaffold by cell-free extracts from yeast-extract-treated *Cassia bicapsularis* cell cultures. *Phytochemistry* **88**, 15–24.
25. Lin H, Annamalai T, Bansod P, Tse-Dinh YC & Sun D (2013) Synthesis and antibacterial

- evaluation of anziaic acid and its analogues as topoisomerase I inhibitor. *Med Chem Commun* **4**, 1613–1618.
26. Kusebauch B, Brendel N, Kirchner H, Dahse HM & Hertweck C (2011) Assessing oxazole bioisosteres as mutasynthones on the rhizoxin assembly line. *ChemBioChem* **12**, 2284–2288.
 27. Rice LM, Earnest TN & Brunger AT (2000) Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr D Biol Crystallogr* **56**, 1413–1420.
 28. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology; Macromolecular Crystallography Part A*, eds. Charles WC, Jr. and Sweet, RM (Academic Press), **276**, 307–326.
 29. Kabsch W (2010) Xds. *Acta Crystallogr D Bio. Crystallog.* **66**, 125–132.
 30. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the *PHENIX AutoSol* wizard. *Acta Crystallogr D Biol Crystallogr* **65**, 582–601.
 31. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Echols N, Headd JJ, Hung LW, Jain S, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner RD, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2011) The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106.
 32. Vagin A & Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66**, 22–25.
 33. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
 34. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M,

- Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352–367.
35. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
36. Emsley P, Lohkanp B, Scott WG, Cowtan K (2010) Features and Development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.
37. Chen VB, Arendall WB3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21.
38. Holm L& Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549.
39. Brunger AT (2007) Version 1.2 of the crystallography and NMR system. *Nat Protoc* **2**, 2728–2733.

Figure Legends

Figure 1. Enzymatic formation of polyketides by polyketide cyclases. (A) Nogalaviketone, (B) Tcm F2-ACP, (C) Tcm F1, and (D) OA.

Figure 2. Primary sequence comparison of OAC with other DABB proteins. The secondary structures of OAC are delineated as follows: α -helices (green rectangles), β -strands (orange arrows), and loops (red, bold lines). The proposed catalytic residues of OAC and the equivalent residues of AtHS1 are colored red (numbering according to OAC-OA). Abbreviations (Uniprot codes): *Arabidopsis thaliana* AtHS1 (Q9LUV2), *Populus tremula* SP1 (P0A881), *Arabidopsis thaliana* At5g22580 (Q9FK81), *Streptomyces glaucescens* TcmI cyclase (P39890), *S. coelicolor* ActVA-Orf6 (Q53908), *Pseudomonas reinekei* MLMI (C5MR76), *S. nogalater* SnoaB (O54259), *Mycobacterium tuberculosis* Rv0793 (O86332), *P. aeruginosa* PA3566 (Q9HY51).

Figure 3. Overall structure of OAC. (A) The apo OAC structure in an asymmetric unit. (B) Structure comparisons of monomer A with monomer B. (C) Overall structures of the OAC-OA binary complex in an asymmetric unit. (D) The homodimeric structure of OAC, formed by a crystallographic two-fold axis. (E) Stereo view of active site cavity of OAC-OA binary complex. (F) Structure comparisons of monomer A in OAC apo with the OAC-OA binary complex. Monomers A and B in the OAC apo and monomer in the asymmetric unit in the OAC-OA binary complex and its symmetric monomer are shown in salmon, sky blue, white, and wheat cartoon models, respectively. His5, Tyr27, Tyr72, His78, and Asp96, previously reported mutated residues, and the residues forming the dimer interface are depicted by stick models. The previously reported mutated residues are highlighted in red residue numbers. The residues forming the dimer interface are labeled with blue residue numbers, and the hydrogen bonds are depicted by light blue dashed lines, respectively. The OA molecule bound in the

OAC structure is shown by a green stick model. Electron density maps of OA are displayed with blue meshes ($F_o - F_c > 2.0 \sigma$).

Figure 4. Structure comparison of OAC with other plant and bacterial DABB proteins. (A) OAC-OA, (B) AtHS1, (C) SP1, (D) At5g22580, (E) SnoaB, (F) Rv0793, (G) MLMI, (H) ActVA-Orf6, (I) PA3566, and (J) TcmI. Each main structure is represented by a cartoon model. The proposed catalytic residues are depicted by stick models. Olivetolic acid, 3-methylmuconolactone, and acetyl dithranol, bound in the OAC, MLMI, and ActVA-Orf6 structures, respectively, are depicted by green stick models.

Figure 5. Close-up views of the active-site cavity of OAC, AtHS1, and TcmI. (A) OAC, (B) AtHS1, and (C) TcmI. The active-site cavity of OAC is shown in stereo. The OA molecule bound in the OAC structure and the hydrogen bonds are depicted by a green stick model and light blue dashed lines, respectively.

Figure 6. HPLC elution profiles of enzyme reaction products from hexanoyl-CoA and malonyl-CoA with TKS and OAC wild-type or mutant enzymes.

Figure 7. The active-site architectures of wild-type OAC and the OAC mutant enzymes. The active-site architectures of (A) OAC wild-type, and the OAC mutant enzymes (B) H78S, (C) Y72F, (D) H5Q, (E) Y27F, (F) Y27W, (G) I7F, and (H) V59M. The OA molecule bound in the wild-type OAC structure is superimposed on the active-site cavity of the mutant enzymes, and depicted by a green stick model. The van der Waals force of OA and the hydrogen bonds are depicted by dotted surfaces and light blue dashed lines, respectively.

Figure 8. HPLC elution profiles of enzyme reaction products obtained by the OAC and OA-NAC co-incubation assay. Enzyme reaction products of (A) OAC with OA-NAC, (B) TKS with OA-NAC, and (C) boiled OAC and TKS with OA-NAC. (D) Authentic OA and OA-NAC.

Figure 9. Stereo views of docking models of OAC. (A) Model structure of OAC bound the pentyl- β -ketide-CoA molecule and (B) model structure of OAC bound the cyclized intermediate-CoA. The pentyl tetra- β -ketide- and its cyclized intermediate-CoA were shown as green stick models. His5, Tyr27, Tyr72, His78, Asp96 are depicted by stick models.

Figure 10. Proposed mechanism for the formation of OAC by OA. (A) ~~The proton exchange between His78 and Tyr72 and Proton~~ abstraction from His78 at C2 of the substrate. (B and C) The C2-C7 aldol cyclization of the intermediate to generate the OAC final product, ~~and the restoration of His78 to its initial state in the enzyme reaction.~~ (D and E) Spontaneous aromatization and CoA bond cleavage of the CoA-linked cyclized product released by OAC.

Table 1. Activities of mutant OAC enzymes relative to the wild-type.

Enzymes	Relative activity* (%)	Enzymes	Relative activity* (%)
Wild-type	100 ± 0.03	Y27L	48 ± 0.05
H5L	36 ± 0.11	Y27M	30 ± 0.07
H5Q	22 ± 0.81	Y27W	37 ± 0.15
H5S	47 ± 0.35	V59M	65 ± 0.11
I7L	85 ± 0.23	Y72F	0
I7F	64 ± 0.04	H78N	0
F24L	49 ± 0.22	H78Q	0
Y27F	162 ± 0.63	H78S	0

* Activities are percentages of OA production relative to that of wild-type OAC (average of triplicates ± SD).

Table 2. Data collection, phasing and refinement statistics

Sample name (PDB code)	OAC (Se-Met)	OAC (Native, 5B08)	OAC-OA (5B09)
Sample conditions			
Reservoir condition	0.1 M Tris-HCl, pH 8.8 25% (w/v) PEG6000 0.1 M Sodium malonate	0.1 M Tris-HCl, pH 8.8 25% (w/v) PEG6000 0.1 M Sodium malonate	0.1 M Tris-HCl, pH 8.8 25% (w/v) PEG6000 50 mM olivetolic acid 5% (v/v) methanol
Cryoprotectant ^a	10% (v/v) glycerol	10% (v/v) glycerol	10% (v/v) glycerol
Data collection			
Beamline	PF NW12A	PF NW12A	PF NE3A
Wavelength (Å)	0.97908	1.00000	1.00000
Space group	<i>P</i> 3 ₂ 21	<i>P</i> 3 ₂ 21	<i>C</i> 2
a, b, c (Å)	47.7, 47.7, 176.2	47.4, 47.4, 175.9	86.5, 30.5, 36.9
α, β, γ (°)			$\beta = 108.5$
Resolution range (Å)	50.0 - 1.40 (1.49 - 1.40)	50.0 - 1.32 (1.40 - 1.32)	50.0 - 1.70 (1.80 - 1.70)
Completeness (%)	99.7 (98.0)	98.6 (93.2)	98.5 (92.3)
$\langle I/\sigma(I) \rangle$	19.0 (3.8)	28.0 (11.4)	27.9 (5.1)
R_{merge} (%) ^b	6.0 (38.9)	4.9 (16.9)	4.1 (31.0)
Multiplicity	5.5 (5.4)	9.9 (9.7)	6.8 (4.7)
No. of observed reflections	484,317	479,942	70,068
No. of unique reflections	87,398	53,774	10,098
Refinement			
Resolution (Å)		39.94 - 1.32	32.11 - 1.70
Overall R_{work} (%) ^c		19.7	19.7
Overall R_{free} (%) ^d		21.5	22.3
Chain order in the unit cell		A 2 - 101 B -2 - 101	3 - 50, 54 - 101
Total Atoms			
No. of protein atoms		1,995	883
No. of ligand atoms		1,732	811
No. of solvent atoms		-	16
		223	56
Average B factors (Å)			
Protein atoms		8.73	23.5
Ligand atoms		-	31.6
Solvent atoms		12.48	30.5
R. m. s. d. from ideal			
Bond length (Å)		0.007	0.010
Bond angles (°)		1.027	1.156
Ramachandran plot			
Favored region		98.0%	100%
Allowed region		2.0%	0%
Outlier region		0%	0%

Sample name (PDB code)	OAC H5Q (5B0A)	OAC I7F (5B0B)	OAC Y27F (5B0C)
Sample conditions			
Reservoir condition	0.1 M Sodium acetate 25% (w/v) PEG4000 8% (v/v) isopropanol	0.1 M Sodium acetate pH 4.7 25% (w/v) PEG2000MME	0.1 M Sodium HEPES pH 7.5 25% (w/v) PEG3000
Cryoprotectant ^a	10% (v/v) glycerol	10% (v/v) glycerol	10% (v/v) glycerol
Data collection			
Beamline	R - AXIS VII	PF NW12A	PF NW12A
Wavelength (Å)	1.54180	1.00000	1.00000
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁
a, b, c (Å)	36.8, 30.5, 81.5	43.4, 45.2, 94.4	85.9, 30.3, 36.7
α , β , γ (°)	β = 95.5	β = 96.3	β = 108.5
Resolution range (Å)	50.0 - 2.10 (2.22 - 2.10)	50.0 - 2.18 (2.26 - 2.18)	50.0 - 1.60 (1.70 - 1.60)
Completeness (%)	95.8 (91.2)	99.3 (96.0)	99.3 (98.6)
$\langle I/\sigma(I) \rangle$	31.9 (13.0)	21.5 (5.0)	42.3 (13.0)
R_{merge} (%) ^b	2.4 (9.9)	7.4 (29.6)	2.0 (8.9)
Multiplicity	3.5 (3.3)	3.7 (3.6)	3.6 (3.5)
No. of observed reflections	38,289	70,778	86,682
No. of unique reflections	10,898	19,023	23,862
Refinement			
Resolution (Å)	34.65 - 2.10	43.14 - 2.19	34.59 - 1.80
Overall R_{work} (%) ^c	22.2	20.1	19.3
Overall R_{free} (%) ^d	27.4	25.4	23.0
Chain order in the unit cell	A 2 - 50, 54 - 63, 66 - 101 B 2 - 63, 69 - 101	A - D 2 - 101	A, B 2 - 101
Total Atoms			
No. of protein atoms	1,655	3,331	1,913
No. of ligand atoms	1,604	3,144	1,684
No. of solvent atoms	-	16	-
	51	171	229
Average B factors (Å)			
Protein atoms	27.7	25.2	14.8
Ligand atoms	-	23.4	-
Solvent atoms	28.0	31.7	21.7
R. m. s. d. from ideal			
Bond length (Å)	0.011	0.010	0.007
Bond angles (°)	1.118	1.289	1.149
Ramachandran plot			
Favored region	98.9%	99.7%	99.0%
Allowed region	1.1%	0.3%	1.0%
Outlier region	0%	0%	0%

Sample Name (PDB code)	OAC Y27W (5B0D)	OAC V59M (5B0E)	OAC Y72F (5B0F)
Sample conditions			
Reservoir condition	0.1 M MES, pH 6.5 25% (w/v) PEG3000	0.1 M Tris-HCl pH 8.5 15% (w/v) PEG20000	0.3 M tris-Sodium citrate 33% (w/v) PEG6000
Cryoprotectant ^a	10% (v/v) glycerol	10% (v/v) glycerol	10% (v/v) glycerol
Data collection			
Beamline	PF NW12A	PF NW12A	PF BL17A
Wavelength (Å)	1.00000	1.00000	0.98000
Space group	<i>P</i> 2 ₁	<i>P</i> 3 ₂ 2 ₁	<i>P</i> 2 ₁
a, b, c (Å)	82.1, 29.8, 36.4	47.7, 47.7, 176.9	35.8, 63.6, 46.6
α, β, γ (°)	$\beta = 96.6$		$\beta = 108.3$
Resolution range (Å)	50.0 - 1.80 (1.91 - 1.80)	50.0 - 1.60 (1.70 - 1.60)	50.0 - 1.60 (1.70 - 1.60)
Completeness (%)	99.5 (98.8)	99.9 (99.7)	98.0 (96.7)
$\langle I/\sigma(I) \rangle$	13.1 (2.4)	26.2 (7.0)	19.1 (3.0)
R_{merge} (%) ^b	7.5 (58.7)	7.2 (34.7)	5.6 (59.6)
Multiplicity	3.6 (3.5)	10.7 (10.6)	5.7 (5.6)
No. of observed reflections	60,041	339,339	146,242
No. of unique reflections	16,563	31,774	25,752
Refinement			
Resolution (Å)	34.58 - 1.80	41.30 - 1.60	44.23 - 1.60
Overall R_{work} (%) ^c	19.6	21.2	17.5
Overall R_{free} (%) ^d	24.5	23.5	22.3
Chain order in the unit cell	A, B 2 - 101	A 2 - 101 B -2 - 101	A, B 2 - 101
Total Atoms			
No. of protein atoms	1,791	1,873	1,826
No. of ligand atoms	1,672	1,711	1,684
No. of solvent atoms	-	6	6
No. of solvent atoms	119	156	136
Average B factors (Å)			
Protein atoms	24.2	24.1	26.7
Ligand atoms	-	21.2	23.5
Solvent atoms	31.1	26.7	33.5
R. m. s. d. from ideal			
Bond length (Å)	0.009	0.009	0.006
Bond angles (°)	1.199	1.084	1.005
Ramachandran plot			
Favored region	99.0%	98.5%	99.5%
Allowed region	1.0 %	1.5%	0.5%
Outlier region	0%	0%	0%

Sample Name (PDB code)	OAC H78S (5B0G)
Sample conditions	
Reservoir condition	0.1 M MES pH 6.5 30% (w/v) PEG4000
Cryoprotectant ^a	10% (v/v) glycerol
Data collection	
Beamline	PF NW12A
Wavelength (Å)	1.00000
Space group	C2
a, b, c (Å)	85.0, 32.7, 36.6
α, β, γ (°)	$\beta = 108.3$
Resolution range (Å)	50.0 - 1.40 (1.48 - 1.40)
Completeness (%)	99.2 (98.7)
$\langle I/\sigma(I) \rangle$	25.3 (9.1)
R_{merge} (%) ^b	3.3 (13.0)
Multiplicity	3.6 (3.5)
No. of observed reflections	68,894
No. of unique reflections	18,918
Refinement	
Resolution (Å)	40.38 - 1.40
Overall R_{work} (%) ^c	17.4
Overall R_{free} (%) ^d	20.0
Chain order in the unit cell	2 - 101
Total Atoms	994
No. of protein atoms	839
No. of ligand atoms	-
No. of solvent atoms	155
Average B factors (Å)	
Protein atoms	13.4
Ligand atoms	-
Solvent atoms	24.7
R. m. s. d. from ideal	
Bond length (Å)	0.009
Bond angles (°)	1.215
Ramachandran plot	
Favored region	99.0%
Allowed region	1.0%
Outlier region	0%

^a The cryoprotectant was mixed into each crystallization reservoir.

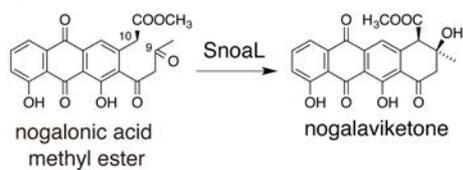
^b $R_{\text{merge}} = \frac{\sum_h \sum_j | \langle I \rangle_h - I_{h,j} |}{\sum_h \sum_j I_{h,j}}$, where $\langle I \rangle_h$ is the mean intensity of symmetry-equivalent reflections.

^c $R_{\text{work}} = \frac{\sum |F_{\text{obs}} - F_{\text{cal}}|}{\sum F_{\text{obs}}}$, where F_{obs} and F_{cal} are observed and calculated structure factor amplitudes.

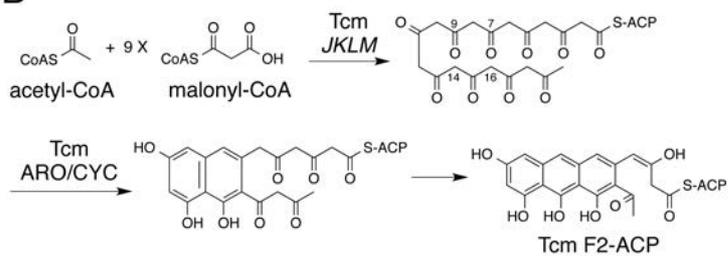
^d R_{free} value was calculated for the R factor, using only an unrefined subset of reflection data.

Fig. 1

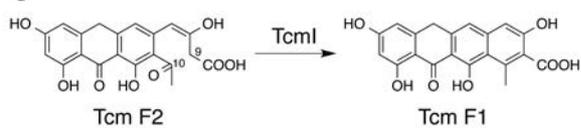
A



B



C



D

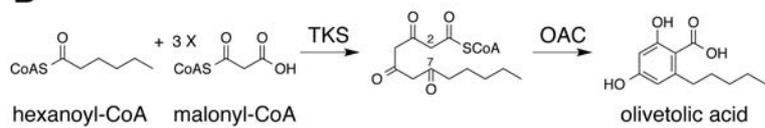


Fig. 3

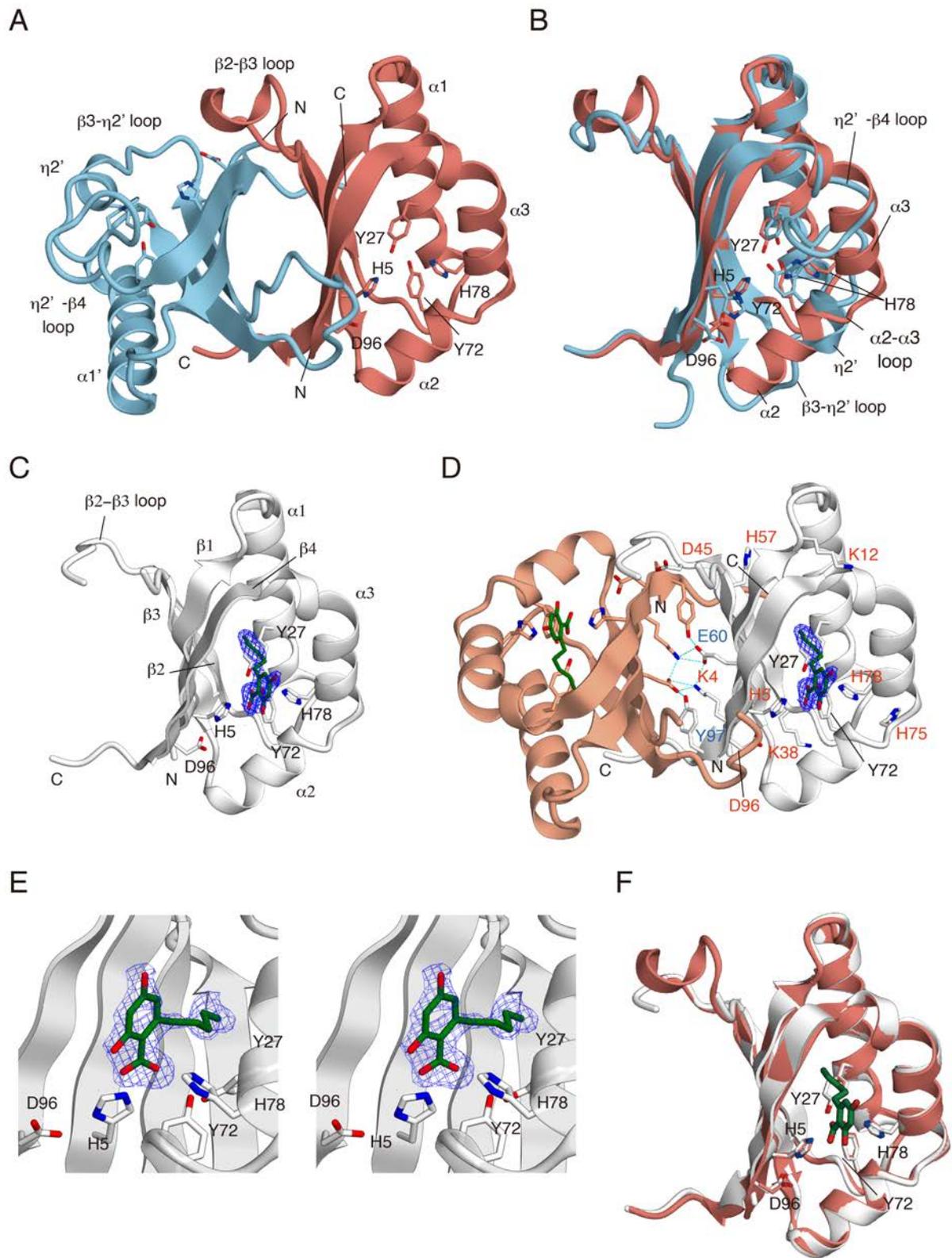


Fig. 4

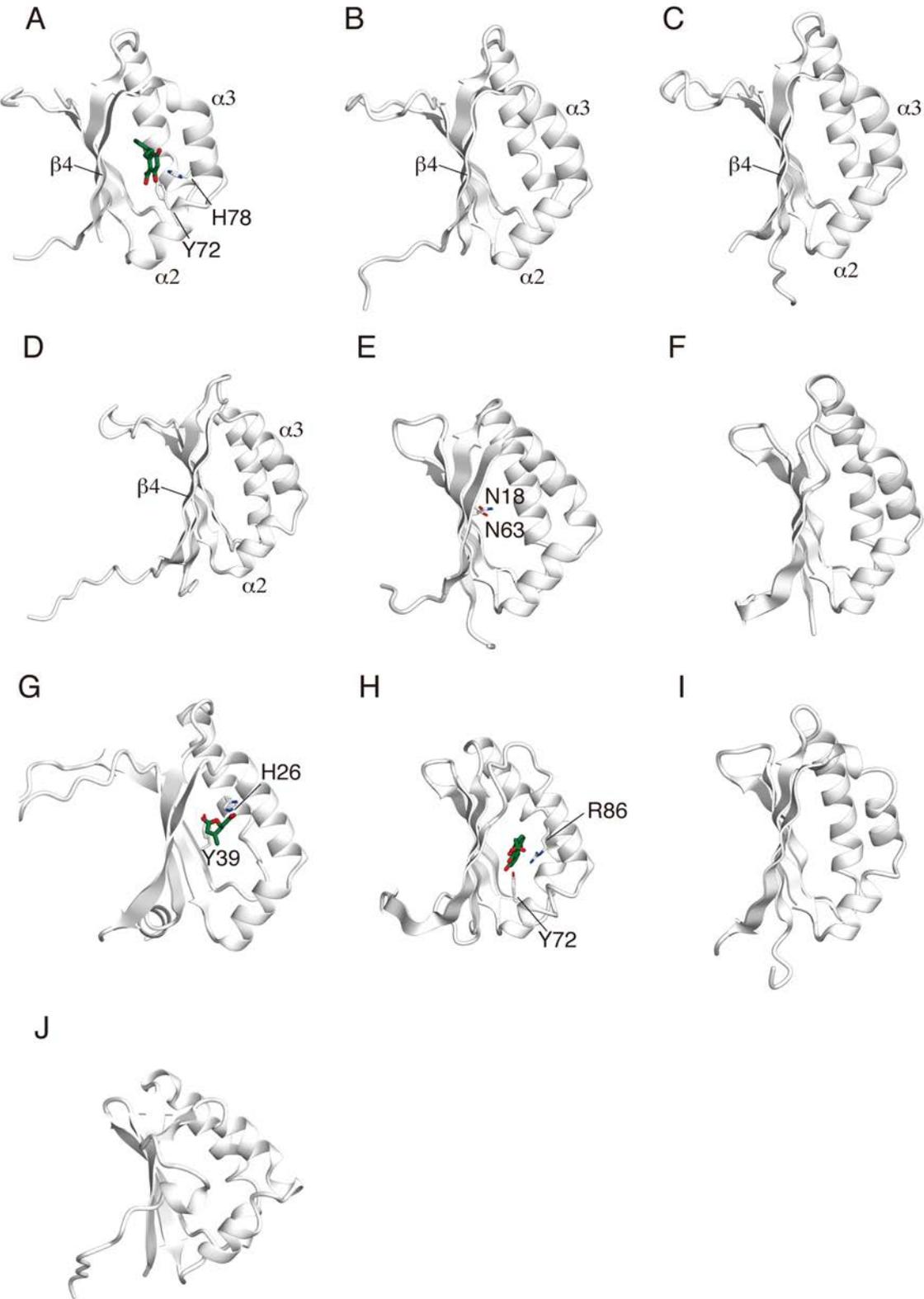


Fig. 5

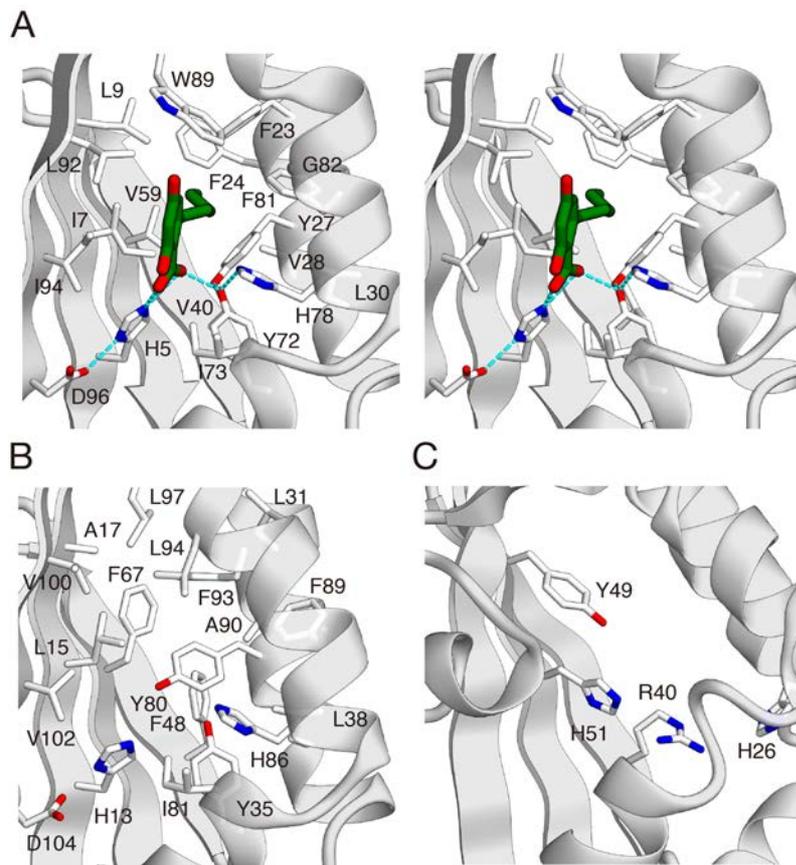


Fig. 6

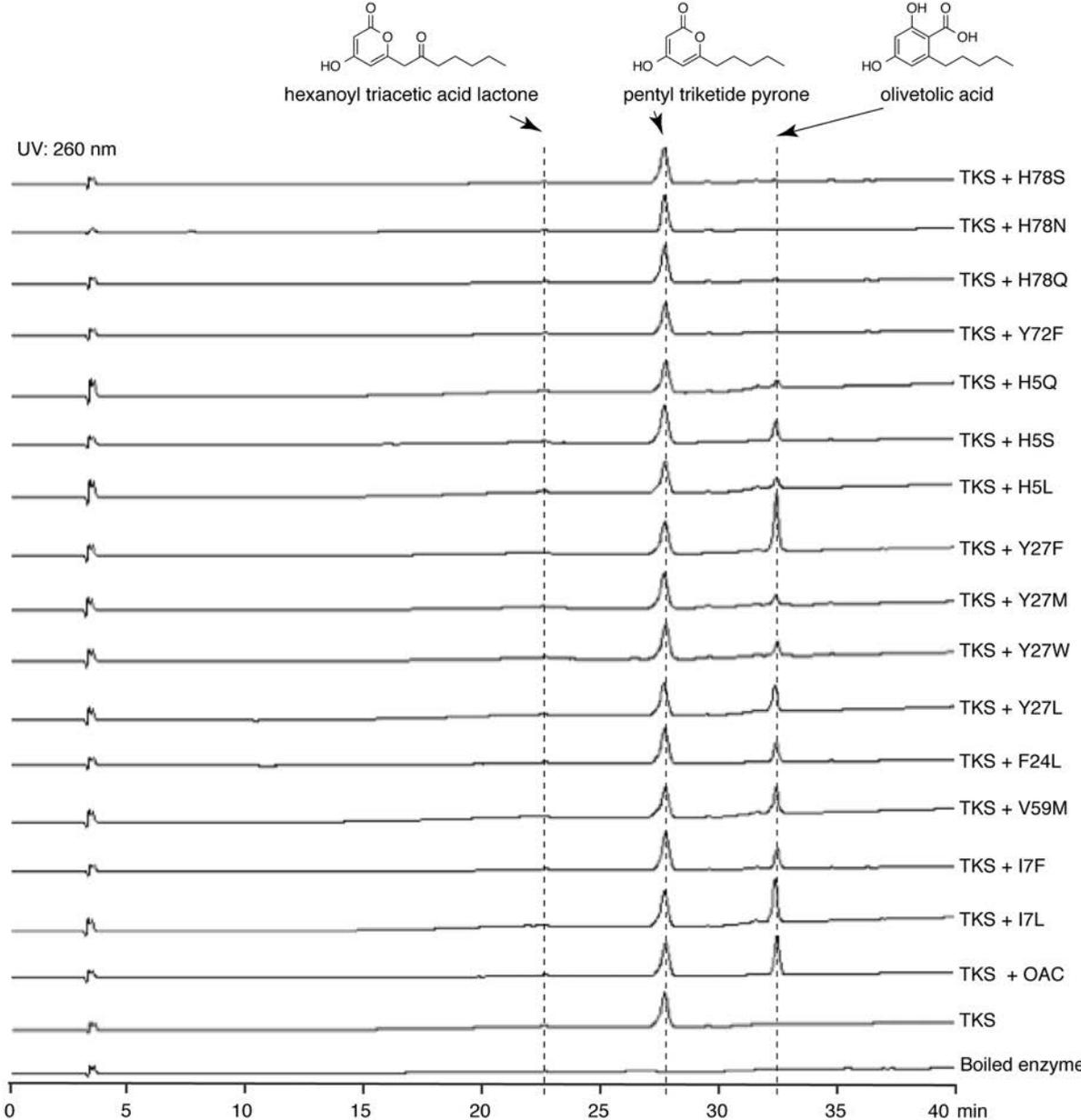


Fig. 7

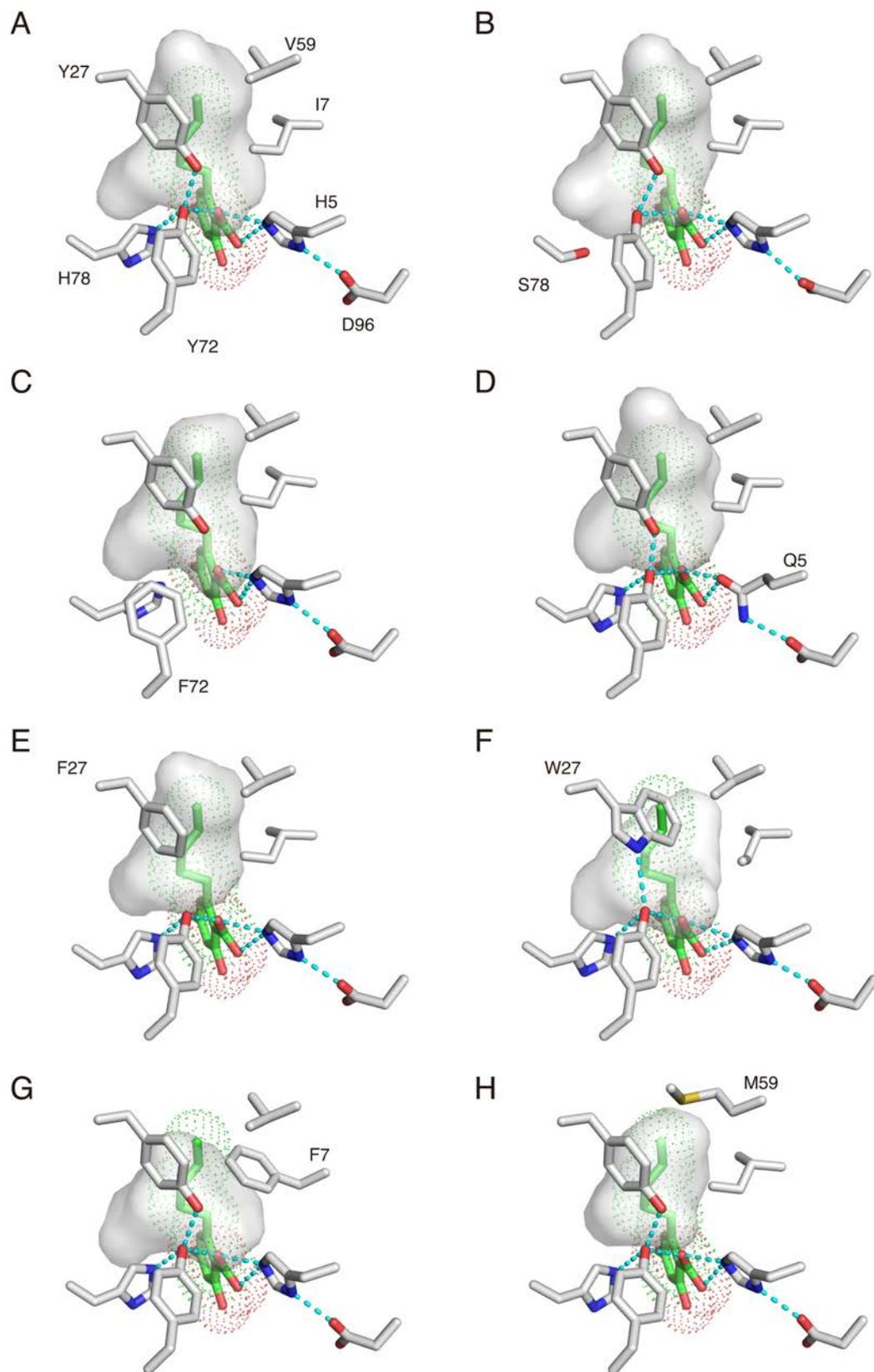


Fig. 8

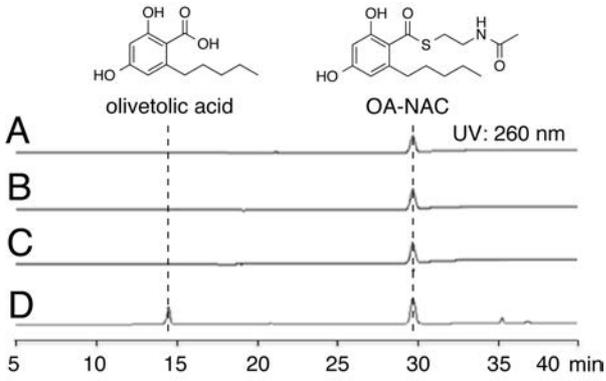
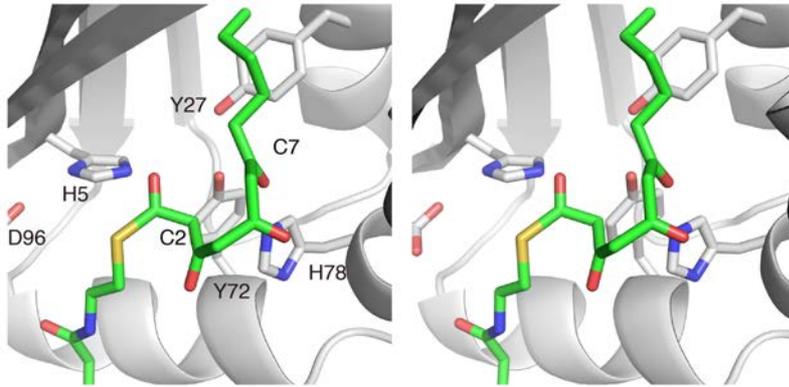


Fig. 9

A



B

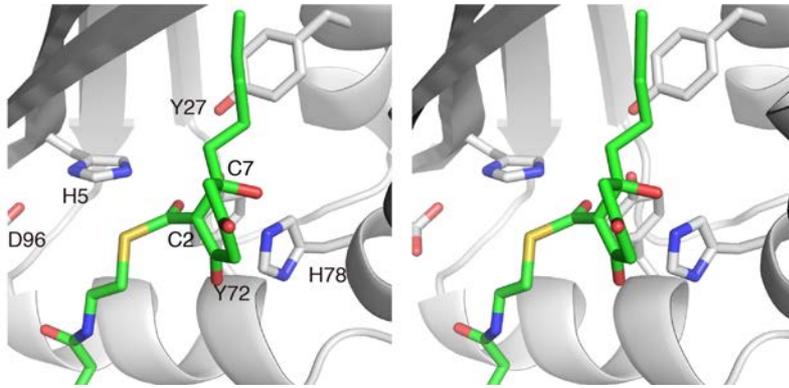


Fig. 10

