

# Excellence in Chemistry Research

## Announcing our new flagship journal

- Gold Open Access
- Publishing charges waived
- Preprints welcome
- Edited by active scientists



## Meet the Editors of *ChemistryEurope*



**Luisa De Cola**

Università degli Studi  
di Milano Statale, Italy



**Ive Hermans**

University of  
Wisconsin-Madison, USA



**Ken Tanaka**

Tokyo Institute of  
Technology, Japan

# Molecular Insights into Bifunctional Ambruticin DH3 for Substrate Specificity and Catalytic Mechanism

Zeqian Du,<sup>[a]</sup> Yongzhen Li,<sup>[a]</sup> Yihan Liu,<sup>[a]</sup> and Ting Shi<sup>\*[b]</sup>

**Abstract:** Dehydratase (DH), a domain located at polyketide synthase (PKS) modules, commonly catalyzes the dehydration of  $\beta$ -hydroxy to an  $\alpha,\beta$ -unsaturated acyl intermediate. As a unique bifunctional dehydratase, AmbDH3 (the DH domain of module 3 of the ambruticin PKS) is verified to be responsible for both dehydration and the following pyran-forming cyclization. Besides, *in vitro* studies showed that its catalytic efficiency varies with different chiral substrates. However, the detailed molecular mechanism of AmbDH3 remains unclear. In this work, the structural rationale for the substrate specificity (2R/2S- and 6R/6S-substrates) in AmbDH3 was elucidated and the complete reaction pathways including dehydration and cyclization were presented. Both MD simulations and binding free energy calculations indicated AmbDH3 had a stronger preference for 2R-substrates (2R6R-2,

2R6S-3) than 2S-substrates (2S6R-1), and residue H51 and G61 around the catalytic pocket were emphasized by forming stable hydrogen bonds with 2R-substrates. In addition, AmbDH3's mild tolerance at C6 was explained by comparison of substrate conformation and hydrogen bond network in 6S- and 6R-substrate systems. The QM/MM results supported a consecutive one-base dehydration and cyclization mechanism for 2R6S-3 substrate with the energy barrier of 25.2 kcal mol<sup>-1</sup> and 24.5 kcal mol<sup>-1</sup>, respectively. Our computational results uncover the substrate recognition and catalytic process of the first bifunctional dehydratase-cyclase AmbDH3, which will shed light on the application of multifunctional DH domains in PKSs for diverse natural product analogs and benefit the chemoenzymatic synthesis of stereoselective pyran-containing products.

## Introduction

Modular polyketide synthases (PKSs) can produce diverse biomolecules with antifungal, anti-cancer, or immunosuppressive activity.<sup>[1]</sup> For synthesizing precursor of these biomolecules, PKS modules minimally have to include an acyltransferase (AT) responsible for loading an extension unit, an acyl carrier protein (ACP) carrying the growing unit between modules, and a ketosynthase (KS) catalyzing the elongation reaction of polyketide. Besides, the combinations with other optional tailoring domains, like keto-reductase (KR) reducing ketone to hydroxyl group, dehydratase (DH) catalyzing the dehydration, and enoyl reductase (ER) adding hydrogens on  $\alpha,\beta$ -unsaturated double

bonds, will further enrich the variety of products by decorating the polyketide chain.<sup>[2]</sup>

DH domains of PKSs own the characteristic  $\alpha + \beta$  double hot dog folds with a large hydrophobic substrate-binding pocket where there is a conserved dyad of His-Asp/Glu residues (Figure S1†).<sup>[3]</sup> Besides, it is generally accepted that DH domains are in charge of the dehydration of  $\beta$ -hydroxy to an  $\alpha,\beta$ -unsaturated acyl intermediate.<sup>[4]</sup> However, recent research has shown that some DH domains also display other additional functions, such as cyclization,<sup>[5]</sup> double-bond isomerization,<sup>[6]</sup> and methyl group epimerization.<sup>[7]</sup> These functions endow PKSs with great potential to produce various products with new biomedical activities.

Ambruticins (varied at C5 position) exert their effects by inducing the high-osmolarity glycerol (HOG) signaling pathway to inhibit the growth of fungi, so they are important drug leads for the development of antifungal agents.<sup>[8]</sup> What's more, AmbDH3 has great potential to significantly simplify the process of synthesis of chiral saturated oxygen heterocycles (CSOH) by intramolecular oxy-Michael addition (IMOMA).<sup>[9]</sup> In 2014, Frank Hahn et al. first confirmed that AmbDH3 had a second catalytic activity as cyclase through *in vitro* assay of the recombinant domain with synthetically-derived substrate surrogates. Different from AmbDH3, AmbDH4 was proposed to be in charge of both dehydration and double bond migration, although they had similar structures. Their research demonstrated that AmbDH3 could exactly catalyze the dehydration process of 2R-substrates rather than 2S-substrates, showing its strict chiral selectivity at the substrate C2 position. Meanwhile, they discovered that AmbDH3 could fully cyclize its 6S-substrate

[a] Z. Du, Y. Li, Y. Liu  
State Key Laboratory of Microbial Metabolism  
Joint International Research Laboratory of Metabolic and Developmental  
Sciences,  
School of Life Sciences and Biotechnology  
Shanghai Jiao Tong University  
Shanghai 200240 (P. R. China)

[b] Prof. T. Shi  
School of Life Sciences and Biotechnology  
Shanghai Jiao Tong University  
No. 800 Dongchuan Rd., Shanghai 200240 (P. R. China)  
E-mail: tshi@sjtu.edu.cn

Supporting information for this article is available on the WWW under  
<https://doi.org/10.1002/chem.202203420>

This manuscript is part of a joint special collection on Mechanisms and  
Selectivities of Organic Reactions – In Celebration of Prof. Kendall N. Houk's  
80th birthday.

to the formation of pyran, while only partly catalyze the 6*R*-substrate to cyclization, indicating its mild tolerant chiral selectivity at C6 (Figure 1). Furthermore, as an isolated PKS domain, AmbDH3 is expected to play a tremendous potential role in synthetic utility, since it presents a superior performance in providing access to preparative amounts of chiral tetrahydropyrans THPs, which will further expand the scope of biocatalysts and accessible CSOH types. The crystal structure of AmbDH3 was not reported until 2017, which encourages us to investigate AmbDH3 with atomistic simulations.<sup>[10]</sup> After all, the AmbDH3-catalyzed molecular mechanism of the consecutive dehydration and cyclization is still unsolved and the structural basis for substrate chiral selectivity at C2 and C6 remains uncovered.

In this work, five complex models with AmbDH3 and different chiral substrates (2*S*6*R*-1, 2*R*6*R*-2, 2*R*6*S*-3, 6*R*-4, and 6*S*-5) are constructed to explore substrate chiral selectivity and catalytic mechanism of AmbDH3 by combining MD simulations and QM/MM calculations. The essential conformational characteristics that contribute to dehydration and cyclization are detected, and they dramatically uncover the chiral selectivity at C2 and C6 of substrates by analyzing the interactions between AmbDH3 and its substrates. QM/MM calculations clarify how the dehydration and cyclization reactions proceed in detail. Our studies unpick the reasons for substrate chiral selectivity in the recognition process, and elucidate the catalytic pathways of AmbDH3. These results will pave the way for the application in the chemoenzymatic synthesis of chiral pyran-containing products by rationally designing AmbDH3, and provide a promising paradigm to rationalize catalytic behavior of enzyme and guide the design of novel mutants with enhancing substrate chirality selectivity and improving biological and therapeutic properties.

## Results and Discussions

### Strict chiral selectivity of AmbDH3 at C2 in recognition

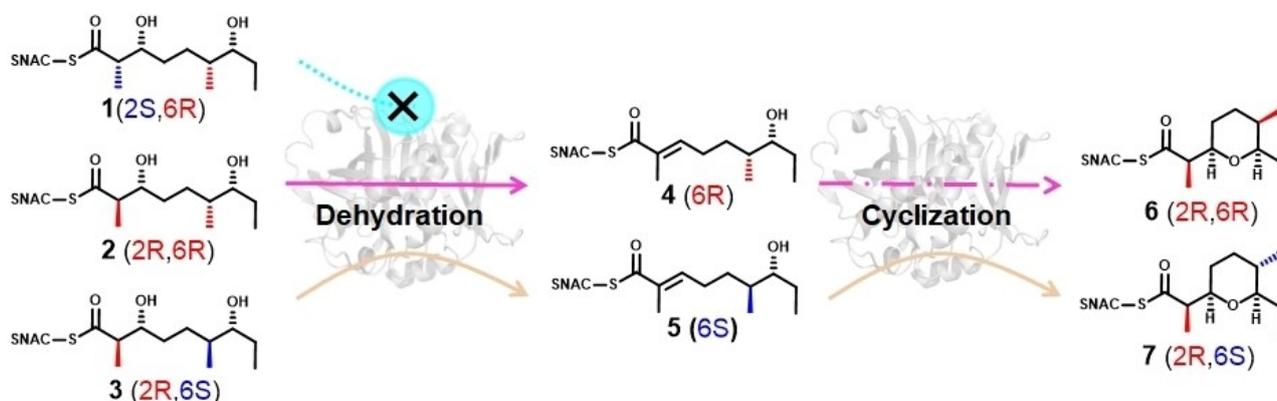
Distance distribution usually was utilized to assess the likelihood of reaction steps and further the formation of the

enzyme-substrate complex with the “reactive state” pose (or Michaelis complex), where the critical interatomic distance was close enough to facilitate the occurrence of reactions. To uncover the substrate selectivity at C2 in AmbDH3, three systems with different chiral substrates 2*S*6*R*-1, 2*R*6*R*-2, 2*R*6*S*-3 were constructed. The distance  $d(H_{\alpha}-N_{\epsilon})$  between the  $H_{\alpha}$  of substrates and the  $N_{\epsilon}$  of H51, representing the beginning of dehydration in the  $H_{\alpha}$ -departure step, was frozen at 2.5 Å by constrained MD simulations in all three systems. After the constrained MD simulations, three replicas of 100 ns MD simulations without the distance restraint were performed in each system.

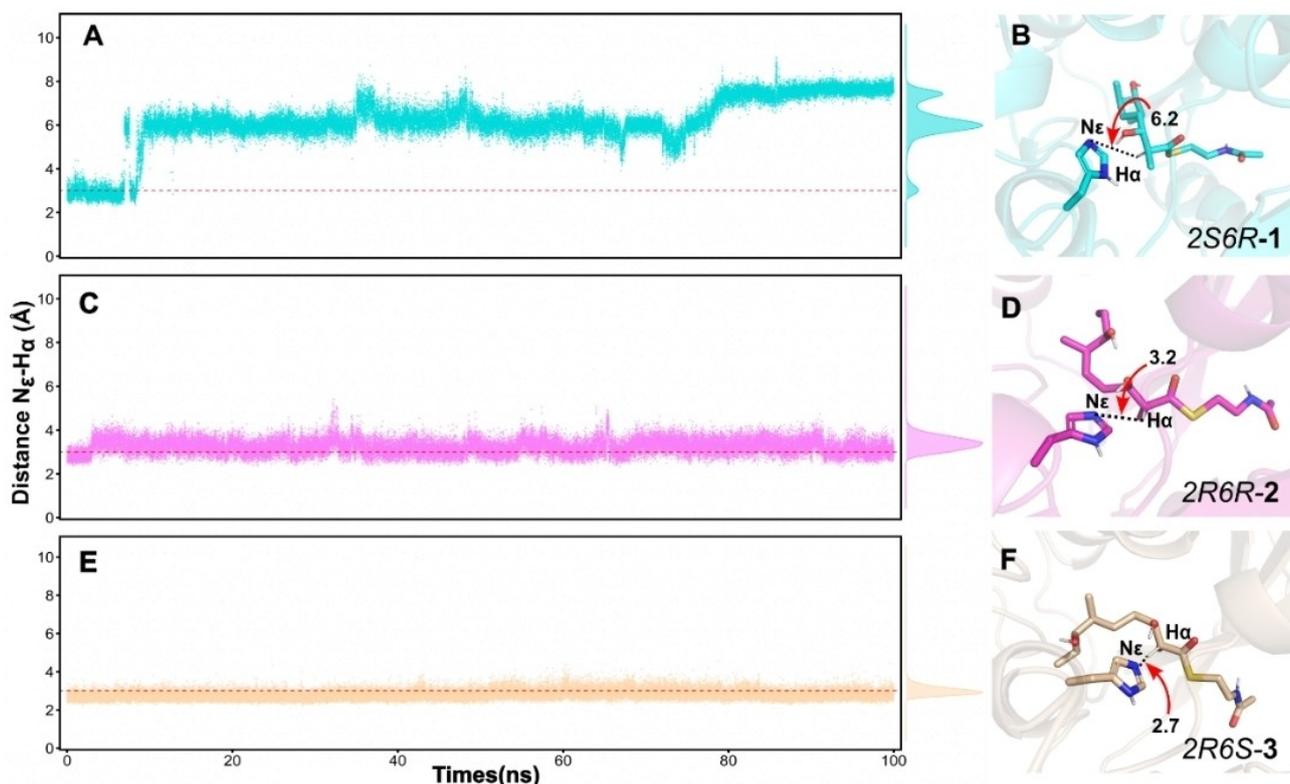
The distance distribution of  $d(H_{\alpha}-N_{\epsilon})$  in all three trajectories was used to evaluate the possibility of the first step  $H_{\alpha}$ -departure, which might be affected by different chiral substrates<sup>[11]</sup> As shown in Figure 2, the distance distribution of  $d(H_{\alpha}-N_{\epsilon})$  in 2*R*6*R*-2 and 2*R*6*S*-3 systems consistently remained close to 3.0 Å, with an average of 3.1 Å and 2.8 Å in their three replicas, respectively, which was appropriate for their following dehydration reactions. Notably, both of them adopted *R*-configuration at C2, while 2*S*6*R*-1 with *S*-configuration at C2 was unfavorable for the following dehydration. The distribution in system 2*S*6*R*-1 was fluctuated and scattered, far from 3.0 Å, with an average of 5.0 Å in all three replicas (Figure S2–3†, Table S1†). From the distance  $d(H_{\alpha}-N_{\epsilon})$  distributions of key reaction coordinates, we concluded that AmbDH3 has strict chiral selectivity at the C2 position of its substrate, preferring 2*R*-substrates and rejecting 2*S*-substrates in recognition, which was well consistent with the experiments reported by Frank Hahn et al.<sup>[10]</sup>

### Critical residues are responsible for substrate selectivity of AmbDH3

Firstly, to find out the critical residues contributing to interactions between AmbDH3 and substrates, MM-GBSA calculations were carried out to get the binding free energies. According to the results, the binding energy for 2*S*6*R*-1 was  $-31.9 \text{ kcal mol}^{-1}$ , the highest in all systems (2*S*6*R*-1, 2*R*6*R*-2, and



**Figure 1.** The consecutive dehydration and pyran-forming cyclization catalyzed by AmbDH3 with specific substrates were reported by Frank Hahn<sup>11</sup>. Unlike substrate 2*S*6*R*-1, 2*R*6*R*-2 and 2*R*6*S*-3 could be dehydrated. After dehydration, 6*R*-4 was partly catalyzed to 2*R*6*R*-6, while 6*S*-5 was fully cyclized to 2*R*6*S*-7.



**Figure 2.** The fluctuations of distance  $d(N_{\epsilon}-H_{\alpha})$  in MD simulations and the dominant structures from cluster analysis of MD simulations (A and B for substrate 2S6R-1 in cyan, C and D for substrate 2R6R-2 in pink, E and F for substrate 2R6S-3 in wheat).

2R6S-3), while that for 2R6R-2 and 2R6S-3 was  $-46.3 \text{ kcal mol}^{-1}$  and  $-41.9 \text{ kcal mol}^{-1}$ , respectively, indicating that 2R-substrates in its reactive pose with AmbDH3 were more favorable than 2S-substrates (Table S2†).

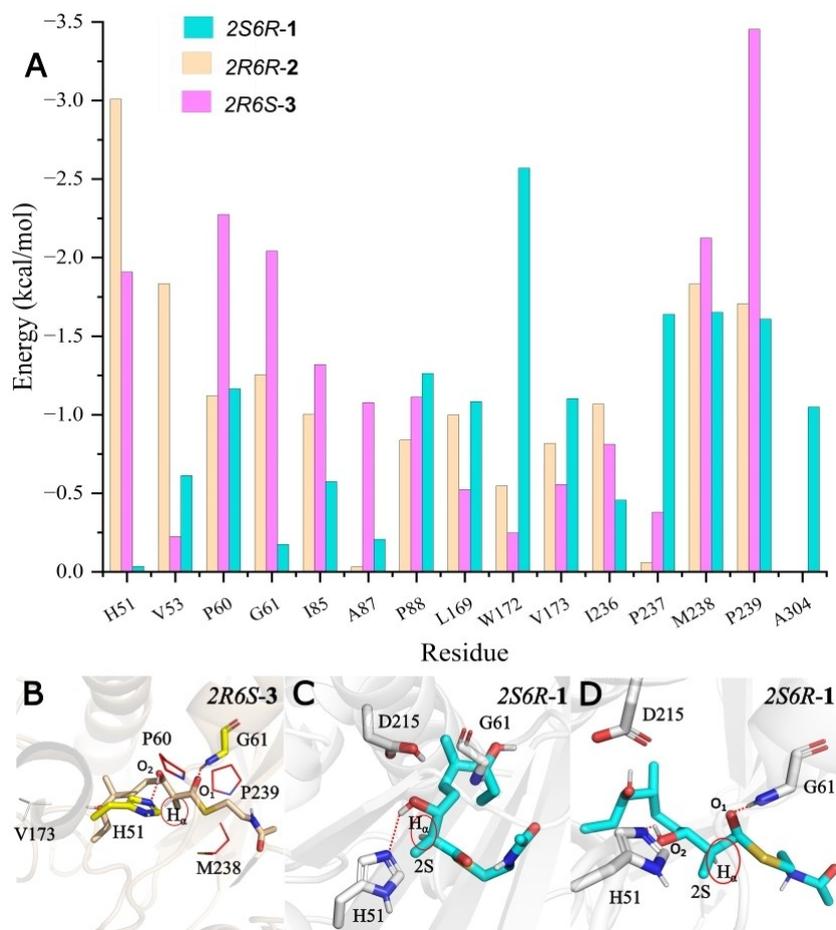
Key residues in the reactive pose were identified by energy decomposition. Fifteen amino acids near the active pocket were highlighted, and each of them contributed more than  $1.0 \text{ kcal mol}^{-1}$  in all systems. They were thought to play important roles in the recognition and catalytic process (Figure 3A).

Secondly, among these fifteen residues, H51 and G61 were noticed because of the great disparity in energy decomposition in system 2R6R-2, 2R6S-3, and 2S6R-1. We carefully distinguished the difference caused by 2R-substrates and 2S-substrate. It was proposed that these two residues might be a decisive factor to determine the strict *R*-configuration selectivity of AmbDH3 at C2. In detail, structure analysis showed that G61 and H51 could form hydrogen bonds with  $O_1$  and  $O_2$  of substrates in system 2R6R-2 and 2R6S-3, making  $H_{\alpha}$  face to H51 and thus easy to be extracted (Figure 3B). On the other hand, in system 2S6R-1, when the constraint of  $d(H_{\alpha}-N_{\epsilon})$  was considered, these two hydrogen bonds were seriously disrupted (Figure 3C), while when the constraint of  $d(H_{\alpha}-N_{\epsilon})$  was abolished, the hydrogen bonds would adjust methyl group to facing H51 rather than  $H_{\alpha}$ , which made it hard for H51 to extract  $H_{\alpha}$  from 2S6R-1 (Figure 3D). Therefore, we summarized that the hydrogen bonds formed by H51 and G61 mediated the recognition

of chiral substrates, which largely rationalized the strict chiral selectivity of AmbDH3 for 2R-substrates from the structural perspective. What's more, H51 was the base amino acid of the catalytic dyad. H51 and G61 were relatively conserved amino acids in the DH domains. The hydrogen bond formed by G51 between the substrate's  $O_1$  atom also existed in GphF DH1 and mammalian DH (corresponding to their Gly1745 and Gly888, respectively) (Figure S4†).<sup>[12]</sup> Experiments verified that the lack of this hydrogen bond between substrate and Gly888 disfavored the  $H_{\alpha}$ -deprotonation.<sup>[13]</sup>

Thirdly, P60, M238, and P239 were discovered to contribute more than  $1.0 \text{ kcal mol}^{-1}$  in all three systems. They participated in the formation of the hydrophobic pocket and helped to mediate the substrates in a proper conformation. What's more, P60 and P239 were highly conserved and reported to play an important role in substrate binding in GphF DH1 (corresponding to P1744 and P1920 respectively). To sum up, these three residues are important for stabilizing the substrate in the active site.

Fourthly, earlier mutational studies demonstrated that the mutant of V173Y and V173H destroyed the cyclized activity of AmbDH3, while the mutant of V173A and V173L didn't.<sup>18</sup> Herein, the energy decomposition also demonstrated that V173 played a crucial part in substrate binding. The energy barrier of pyran-forming cyclization in WT, V173Y, and V173A systems was calculated by umbrella samplings, and the results displayed that mutation of V173 to bulky residue Y would obstruct the



**Figure 3.** (A) Residues contributing more than  $1.0 \text{ kcal mol}^{-1}$  were obtained by energy decomposition; (B) G61 and H51 (labeled by yellow sticks) form hydrogen bonds (labeled by red dotted line) with substrates (labeled by wheat sticks). P60, P239, and M238 are labeled by red lines; (C) With constrained MD simulations, hydrogen bonds with H51 and G61 were disrupted, making the conformation unstable in system 2S6R-1; (D) Without constraint, hydrogen bonds with H51 and G61 make the substrate methyl face to H51 instead of  $H_{\alpha}$ , hindering the  $H_{\alpha}$ -departure step in system 2S6R-1.

cyclization of AmbDH3. The energy barrier was up to  $30.0 \text{ kcal mol}^{-1}$ , which thoroughly destroyed the cyclase activity of AmbDH3. However, the mutation V173A didn't raise the energy barrier. Similar to WT, the energy barrier of cyclization was calculated to be about  $20.0 \text{ kcal mol}^{-1}$ , which made the cyclization occur reasonably in both systems. (Figure S5†)

Finally, in addition to these residues mentioned above, the rest were all hydrophobic residues. They provided hydrophobic microenvironment (Figure S4†) for the hydrophobic main chain of the substrate, which was suitable for the hydrophobic interaction between the substrate and the active site of AmbDH3.

### The dehydration mechanism

Based on our previous computations, the deprotonation of  $\alpha$ -carbon before the  $\beta$ -elimination of the  $\beta$ -hydroxyl have a lower energy barrier, compared with the first elimination of the  $\beta$ -hydroxyl and then the deprotonation of  $\alpha$ -carbon.<sup>[12]</sup> In Frank Hahn's work about AmbDH3, he put forward a hypothesis of a

base-acid dehydration process, which proceeded with the extraction of  $H_{\alpha}$  by H51 and then the elimination of  $\beta$ -hydroxyl with a proton from D215 to release a water molecule. Also, it was a widely accepted dehydration mechanism.<sup>[14]</sup> Unfortunately, we failed to find the transition state of  $\beta$ -elimination with D215 donating proton (Figure S6†). Besides, according to our previous study, the  $\beta$ -elimination with the assistance of D215 should be ignored in consideration of the extremely high energy barrier. Therefore, one-base dehydration and cyclization mechanism was proposed in AmbDH3, which was entirely supported by the gephyronic acid dehydratase DH1.<sup>[12]</sup> For the dehydration step, the proton acceptor H51 extracted the  $H_{\alpha}$  from 2R6S-3, and then this proton transferred from H51 to the substrate's  $\beta$ -hydroxyl to produce one molecule of water. For the cyclization step, the refreshed H51 extracted the hydrogen of  $C_7$ 's hydroxyl, at the same time, the oxygen of  $C_7$ 's hydroxyl attacked the newly created double bond at  $C_3$  to form pyran. Finally, the proton transferred from H51 to  $C_2$  to complete the catalytic cycle.

A direct dehydration pathway was investigated to obtain the energy barrier with a two-layered ONIOM method. For the

first step, the formation of enol intermediate (IM1) required  $24.1 \text{ kcal mol}^{-1}$  in energy. When the distances of  $\text{H}_\alpha\text{-C}_2$  and  $\text{H}_\alpha\text{-N}_\epsilon$  were located at  $1.66$  and  $1.17 \text{ \AA}$ , the transition state (TS1) was observed, whose imaginary frequency was  $533.01i \text{ cm}^{-1}$  with an antisymmetric stretching of  $\text{C}_2\text{-H}_\alpha$  and  $\text{H}_\alpha\text{-N}_\epsilon$  bonds. Comparing TS1 with IM1, we could find that the distances of  $\text{C}_2\text{-N}_\epsilon$  and  $\text{O}_2\text{-N}_\epsilon$  changed obviously ( $2.82$  and  $3.04 \text{ \AA}$  in TS1 and  $3.49$  and  $2.70 \text{ \AA}$  in IM1, Figure S7†), indicating H51 mediated its position to transfer  $\text{H}_\alpha$  from substrate's  $\text{C}_2$  to  $\beta$ -hydroxyl. The second step in dehydration was  $\beta$ -hydroxyl elimination along with the generation of a water molecule. The energy barrier was calculated to be  $25.2 \text{ kcal mol}^{-1}$ , which later was found to be the rate-determining step for the whole dehydration and cyclization reaction. The key distances of  $\text{C}_3\text{-O}_2$  and  $\text{O}_2\text{-H}_\alpha$  were  $1.80 \text{ \AA}$  and  $1.52 \text{ \AA}$ , respectively in TS2, which showed an only imaginary frequency of  $543.44i \text{ cm}^{-1}$  via vibrational analysis, corresponding to the stretching of  $\text{C}_3\text{-O}_2\text{H}_2$  and  $\text{H}_\alpha\text{-O}_2\text{H}_2$ . After TS2, another intermediate (IM2') was generated containing a newly formed double bond accompanied by the generation of one water molecule (Figure 4 and Figure S8†).

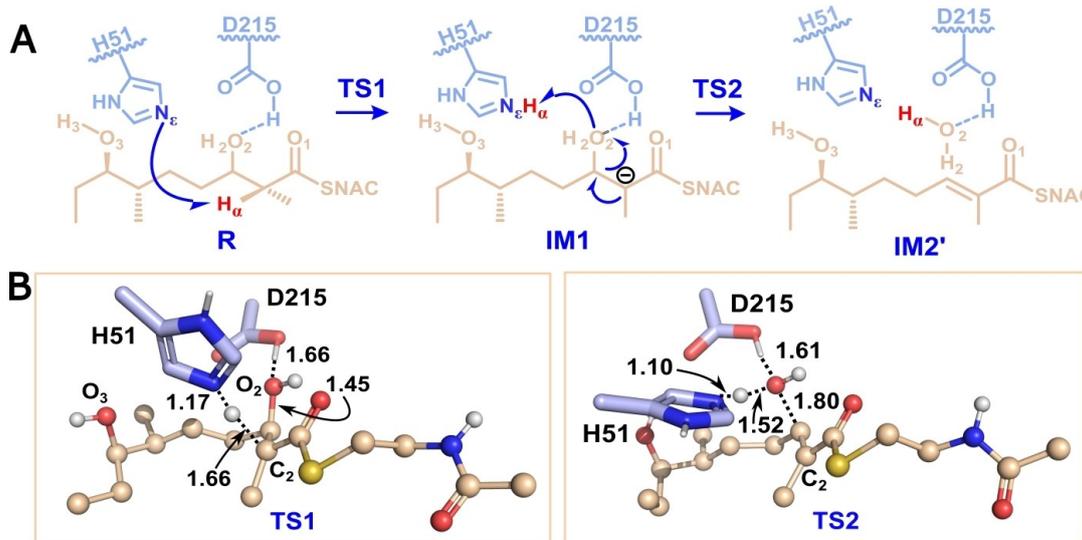
What's more, there was another indirect dehydration pathway, after the formation of enol intermediate (IM1), a bi-hydroxyl-substrate (IM2\_1) was generated, where the proton  $\text{H}_\alpha$  transferred from H51 to  $\text{O}_1$  via  $\text{O}_2$ , rather than directly transferred to  $\text{O}_2$  (Figure S9†). TS2\_1 was observed when the distance of  $\text{N}_\epsilon\text{-H}_\alpha$ ,  $\text{O}_2\text{-H}_\alpha$ ,  $\text{O}_2\text{-H}_2$ , and  $\text{O}_1\text{-H}_2$  was at  $1.08 \text{ \AA}$ ,  $1.54 \text{ \AA}$ ,  $1.17 \text{ \AA}$ ,  $1.27 \text{ \AA}$ , respectively, with an only imaginary frequency  $910.17i \text{ cm}^{-1}$  corresponding to the stretching of  $\text{H}_\alpha\text{-O}_2$  and  $\text{H}_2\text{-O}_1$ . Although this energy barrier was just  $19.7 \text{ kcal mol}^{-1}$ , the following dehydration encountered a relatively high barrier of  $26.9 \text{ kcal mol}^{-1}$ . TS2\_2 was obtained with the distances of  $\text{O}_2\text{-C}_3$ ,  $\text{O}_2\text{-H}_2$ , and  $\text{O}_1\text{-H}_2$  at  $1.74$ ,  $1.07$ , and  $1.45 \text{ \AA}$ , respectively. Also, TS2\_2 had an imaginary frequency of  $540.57i \text{ cm}^{-1}$  corresponding to the stretching of  $\text{O}_2\text{-C}_3$  and  $\text{O}_2\text{-H}_2$  (Fig-

ure S10†). In a word, our calculations preferred the direct dehydration pathway, since its energy barrier ( $25.2 \text{ kcal mol}^{-1}$ ) was a little lower than that in the indirect pathway ( $26.9 \text{ kcal mol}^{-1}$ ).

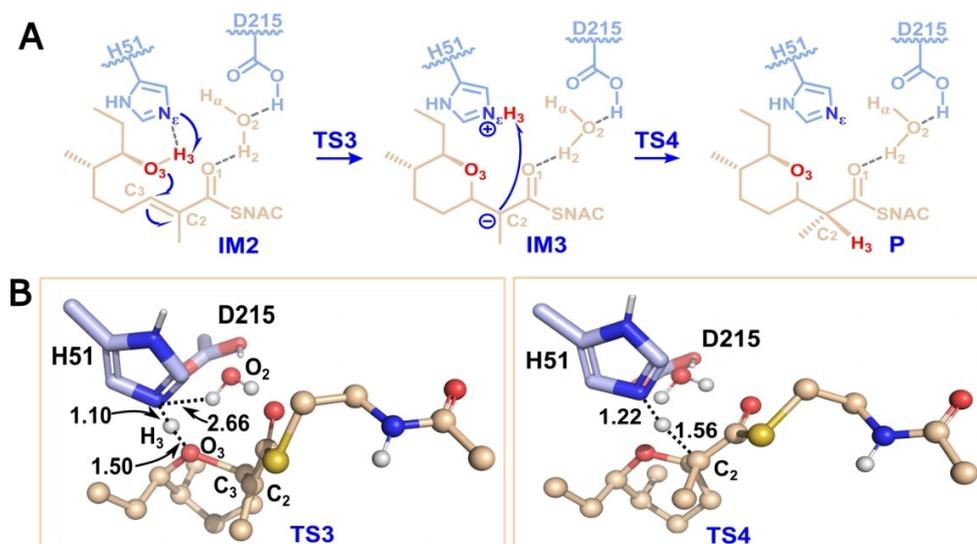
### The cyclization mechanism

After dehydration, IM2' was optimized by flipping its tail close to a ring (IM2), overcoming a tiny energy barrier of  $1.3 \text{ kcal mol}^{-1}$  (Figure S8†). Similar to dehydration, it was a consecutive two-step reaction. Firstly, H51 extracted  $\text{H}_3$  of  $\text{-O}_3\text{H}_3$  accompanied by  $\text{O}_3$  attacking the newly created double bond at  $\text{C}_3$ , and the pyran-ring intermediate (IM3) formed. The energy barrier for the formation of IM3 was  $18.6 \text{ kcal mol}^{-1}$ . As shown in Figure 5, when the distance of  $\text{N}_\epsilon\text{-H}_3$ ,  $\text{O}_3\text{-H}_3$ , and  $\text{C}_3\text{-O}_3$  was located at  $1.10 \text{ \AA}$ ,  $1.50 \text{ \AA}$ , and  $1.85 \text{ \AA}$ , respectively, TS3 was found with the imaginary frequency  $331.94i \text{ cm}^{-1}$ . Besides, TS4 was observed, when the distance of  $\text{N}_\epsilon\text{-H}_3$  and  $\text{C}_2\text{-H}_3$  equaled  $1.22 \text{ \AA}$  and  $1.56 \text{ \AA}$ , with the imaginary frequency of  $1009.67i \text{ cm}^{-1}$ . Finally,  $\text{H}_3$  was donated from  $\text{N}_\epsilon$  of H51 to  $\text{C}_2$  refreshing H51 and forming the final product (P) (Figure S11†). The energy barrier was calculated to be  $24.5 \text{ kcal mol}^{-1}$ .

What's more, on account of a water molecule generated in the dehydration process, we proposed a water-mediated cyclization pathway called the indirect cyclization pathway, where both the proton extracting and its transferring should be mediated by a water molecule (Figure S12†). Firstly, the IM2' was obtained, and then  $\text{O}_2$  of the water extracts  $\text{H}_3$  of  $\text{-O}_3\text{H}_3$  accompanied by  $\text{H}_2$  of the water transferring to  $\text{O}_1$  (IM3'). TS3' was got, when the distance of  $\text{O}_3\text{-H}_3$ ,  $\text{H}_3\text{-O}_2$ ,  $\text{O}_2\text{-H}_2$ , and  $\text{H}_2\text{-O}_1$  reached  $1.13 \text{ \AA}$ ,  $1.29 \text{ \AA}$ ,  $1.04 \text{ \AA}$ , and  $1.48 \text{ \AA}$ , respectively, with an imaginary frequency of  $423.13i \text{ cm}^{-1}$  and the energy barrier of  $20.5 \text{ kcal mol}^{-1}$ . Secondly, H51 extracted the  $\text{H}_\alpha$  of the water accompanied by  $\text{H}_2$  transferring from  $\text{O}_1$  to  $\text{O}_2$  of the water



**Figure 4.** (A) The proposed direct dehydration pathway; (B) Optimized structures of transition states (TS1 and TS2) in the direct dehydration pathway with key distance labeled. The substrate is colored by wheat sticks and spheres. Catalytic dyads are colored by light blue sticks (D215 and H51).



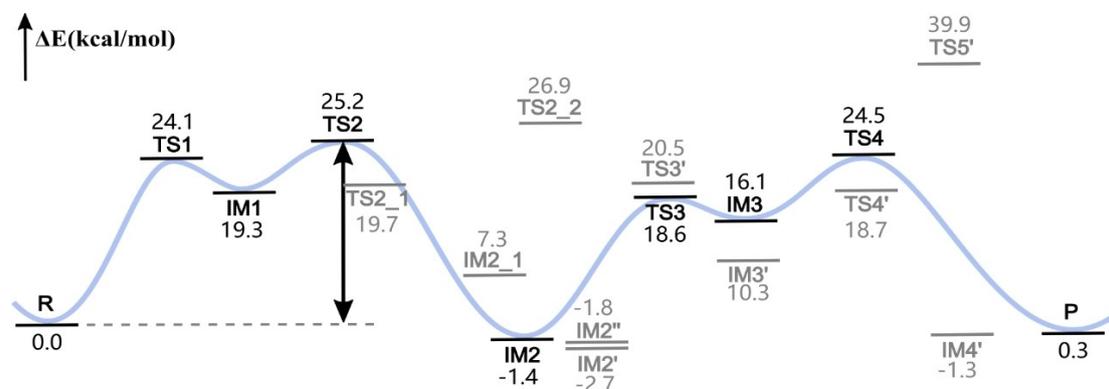
**Figure 5.** (A) The proposed direct cyclization mechanism; (B) Optimized structures of transition states (TS3 and TS4) in the direct cyclization pathway with key distances labeled.

(IM4'). When the distance of H<sub>2</sub>-O<sub>1</sub>, O<sub>2</sub>-H<sub>2</sub>, H<sub>α</sub>-O<sub>2</sub>, and H<sub>α</sub>-N<sub>ε</sub> equaled 1.31 Å, 1.12 Å, 1.24 Å, 1.23 Å, TS4' was obtained with an imaginary frequency of 881.19i cm<sup>-1</sup>. The energy barrier for TS4' was 18.7 kcal mol<sup>-1</sup>. Finally, H51 returned the H<sub>α</sub> to O<sub>2</sub> of the water along with H<sub>3</sub> of the water re-donating to C<sub>2</sub> to release pyran product (P). TS5' was obtained with an imaginary frequency of 522.86i cm<sup>-1</sup> when the distance of H<sub>3</sub>-C<sub>2</sub>, H<sub>3</sub>-O<sub>2</sub>, O<sub>2</sub>-H<sub>α</sub>, and H<sub>α</sub>-N<sub>ε</sub> equaled 1.66 Å, 1.09 Å, 1.10 Å, 1.40 Å, respectively. The energy barrier for TS5' was 39.9 kcal mol<sup>-1</sup> (Figure S13†). Considering the energy barrier of the direct cyclization pathway and the water-mediated pathway, the direct cyclization pathway was more likely to occur in terms of energy (Figure 6).

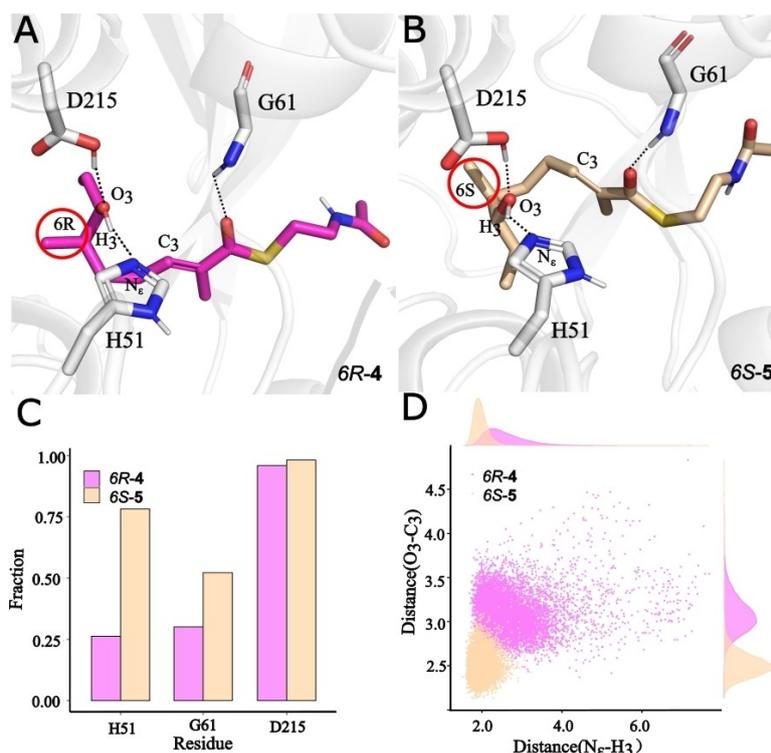
#### Mild tolerance of AmbDH3 at C6 in cyclization

HPLC-MS analyses indicated that 6R-4 could be partly cyclized by AmbDH3 to 2R6R-6, while 6S-5 could be fully cyclized to 2R6S-7, when AmbDH3 overnight incubated with 6R-4 and 6S-5, respectively, indicating its moderate tolerance at C6. To understand the reason for substrate tolerance in AmbDH3 during cyclization, two systems combined AmbDH3 with 6R-4 or 6S-5 were constructed.

Based on our proposed mechanism, two distances  $d(N_{\epsilon}-H_3)$  and  $d(O_3-C_2)$  were collected. The former represented the beginning of cyclization, and the latter represented the formation of pyran. As shown in Figure 7, the populations of distance  $d(N_{\epsilon}-H_3)$  and  $d(O_3-C_2)$  were concentrated in system 6S-5, while they were scattered in system 6R-4 (Figure 7D). Furthermore, the averages of  $d(N_{\epsilon}-H_3)$  and  $d(O_3-C_2)$  were 2.7 Å and 3.6 Å in system 6R-4, while they sharply reduced to 1.9 Å and 3.0 Å in system 6S-5 (Table S3†). By comparison, we could



**Figure 6.** The whole energy profile for the dehydration and cyclization process by QM/MM calculations. The complete pathway was colored black and the unfavorable states were colored gray.



**Figure 7.** Hydrogen bonding network in dominant cluster structure in system *6R-4* (A, colored by pink) and in system *6S-5* (B, colored by wheat); (C) Percentage of hydrogen bonding for H51, G61, and D215 in system *6R-4* and *6S-5*; (D) The distributions of two key distances  $d(N_6-H_3)$  and  $d(O_3-C_3)$ , related to the cyclization reaction in two systems.

see that *6S-5* was more prone to being cyclized by AmbDH3 than *6R-4*. The percentages overlapping both  $d(N_6-H_3) < 3 \text{ \AA}$  and  $d(O_3-C_3) < 4 \text{ \AA}$  were calculated to be 76.5% and 99.8% for system *6R-4* and system *6S-5*, respectively, indicating that in addition to *6S-5*, *6R-4* was accessible to cyclization (Table S4†).

Hydrogen bonding analysis revealed that both *6S-5* and *6R-4* could form hydrogen bonds with H51, G61, and D215. D215 and H51 greatly stabilized the substrates by fixing their tails through hydrogen bonding with  $O_3$ . The percentages of hydrogen bonding in *6S-5* system were 78.1% and 98% for H51 and D215, while they decreased to 26.0% and 96% in *6R-4* system. In addition, G61 formed a hydrogen bond with  $O_1$  and anchored the head of *6R-4* and *6S-5*. The hydrogen bonding percentage was 52.1% in *6S-5* system, while it reduced to 30% in *6R-4* system (Figure 7C). Due to these hydrogen bonds and the chiral distinction at C6, *6R-4* bent inwards and *6S-5* bent outwards (Figure 7A, 7B), which led to the difference in cyclized efficiency.

## Conclusions

AmbDH3 is well known as a unique dual-function dehydratase, responsible for both dehydration and the following pyran-forming cyclization. However, the detailed molecular mechanism for recognition and catalysis of AmbDH3 still remains unclear. Herein, MD simulations and QM/MM calculations were

combined to investigate the substrate specificity and catalytic mechanism during recognition and catalytic process.

In the first part, three enzyme-substrate complexes were constructed with *2S*-substrate (*2S6R-1*) and *2R*-substrate (*2R6R-2* and *2R6S-3*), respectively. *2R*-substrate displayed favorable conformation by forming hydrogen bonds with G61 and D215, which efficiently facilitated the dehydration process. Furthermore, compared with *2S*-substrate, *2R*-substrate had lower binding free energy, revealing more stable interactions with AmbDH3. The energy decomposition highlighted some key residues, which either stabilized the substrates by hydrogen bonding or provided hydrophobic microenvironment by forming a hydrophobic pocket. All these structural basics undertook the strict *R*-configuration selectivity of AmbDH3 at C2 of the substrate.

Next, direct ( $\beta$ -hydroxyl elimination via enol intermediate) and indirect (via a bi-hydroxyl-substrate) dehydration pathways catalyzed by AmbDH3 with natural substrate *6S-5* were proposed and the calculated energy barrier preferred the direct dehydration with  $25.2 \text{ kcal mol}^{-1}$  by QM/MM calculations. Besides, the cyclization mechanism without or with the participation of water molecule were obtained, the energy barrier was  $24.5 \text{ kcal mol}^{-1}$  in the former and  $39.9 \text{ kcal mol}^{-1}$  in the latter, indicating the hydrophobic microenvironment would be suitable for cyclization. In summary, the whole reaction pathway displayed a consecutive one-base dehydration and cyclization mechanism. The rate-determining step was the  $\beta$ -hydroxyl

elimination accompanied by the generation of a water molecule during dehydration with an energy barrier of 25.2 kcal mol<sup>-1</sup>. Our results are in agreement with the one-base dehydration mechanism,<sup>[12]</sup> where the conserved His residue acts as a base to deprotonate the substrate and also provides the proton for the formation of a water molecule.

Last, two enzyme-substrate complexes with 6R-4 and 6S-5 were constructed to uncover the mild tolerance of AmbDH3 at C6. Finally, 6S-5 stood out in stable hydrogen bonding and favorable conformations. The distance distribution of key reaction coordinates showed that both 6R-4 and 6S-5 could be cyclized, while the system of 6S-5 was much more efficient than that of 6R-4, which was consistent well with the experiments.

Overall, our results elaborately reveal the whole reaction mechanism of the unique bifunctional (dehydration and cyclization) dehydratase AmDH3, which will provide deep insights into the catalytic mechanism of dehydratases and shed light on the development of multi-functional dehydratases for generation of chiral saturated oxygen heterocycles, as well as pave the way for the application in the chemoenzymatic synthesis of chiral pyran-containing products by rationally designing dehydratase in PKS modules.

## Experimental Section

**Preparation of systems:** The crystal structure of AmbDH3 was first reported by Frank Hahn et al., in 2017 (PDB number: 5O15 and 5O16). The AmbDH3 structure of 5O15 was more complete than 5O16, and thus it was chosen as the initial structure for building systems. The protonated state of the protein was determined by the PDB2PQR web server.<sup>[15]</sup> All substrates (2S6R-1, 2R6R-2, 2R6S-3, 6R-4, and 6S-5) were optimized by Gaussian 16. Molecular docking calculations were performed using AutoDock Tools (version 1.5.6) software.<sup>[16]</sup> The protein AmbDH3 and ligand structures were first pre-treated, adding the appropriate charge and hydrogen atoms (\*.pdbqt file). The relative position of the catalytic dyad (H51 and D215) in the crystal structure was referred to determine the docking grid space. With the help of Autogrid, the atom-specific affinity maps for all ligand atom types, electrostatic, and desolvation potentials were generated. Combined with the rigid protein molecule and flexible ligands, docking proceeded. The iteration number of docking simulations was set as 200 for each protein-ligand pair using Genetic Algorithm methods. All 200 docking results were clustered and ranked by docking scores, and the single docking pose with the best energy score in the most clusters was selected as the initial structure. All substrates were optimized and calculated the electrostatic surface potential (ESP) at the level of HF/6-31G\*.<sup>[17]</sup> The bonds, angles, dihedral angles, and van der Waals radii parameters for substrates were generated by a two-step restrained electrostatic potential (RESP) charging fitting method embedded in the Multiwfn software and Antechamber package implemented in AMBER 18.<sup>[18]</sup> All systems were immersed in an octahedral box of TIP3P water, with the thickness of the water layer exceeding 10 Å.

**Molecular dynamics simulations and analysis:** Through the AMBER 18 program suite, the best docking result of all five systems was utilized to perform molecular dynamics (MD) simulations with the ff14SB force field (Figure S14†).<sup>[19]</sup> The complex was solvated to a box of TIP3P water. For achieving systems' charge neutralization, seven molecules of sodium ions were added. To avoid inappropriate atomic collisions, we performed a two-step minimization

process for the system, the first step was for all water molecules, and the second step was to avoid the rest parts' inappropriate collision. Then we progressively heated the system from 0 K to 300 K in 100 ps. After heating the system, we switched it to constant pressure and temperature mode (NPT), at this mode, we perform a 200 ps equilibration to ensure that the system was at the correct density. After equilibration, in 6R-4, and 6S-5 systems, three times 100 ns trajectories were carried out and collected for analysis. While in 2S6R-1, 2R6R-2, and 2R6S-3 systems, constraints of d(N<sub>ε</sub>-H<sub>α</sub>) were added by using harmonic potential with a force constant of 20 kcal mol<sup>-1</sup> Å<sup>-2</sup> at 2.5 Å, for making all three systems with the same distance at the beginning (Figure S15†). And then the constraints were removed, 100 ns MD simulations were performed on all these 3 systems.

The Particle Mesh Ewald (PME) method and the SHAKE algorithm were both introduced for long-range electrostatic interactions and fixing bonds and angles involving hydrogen atoms, respectively.<sup>[20]</sup> The cutoff of van der Waals interactions was set to 10.0 Å. All systems' trajectories have reached equilibrium and were analyzed by the cpptraj module in AMBER 18 (Figure S16–17†).

**Umbrella Sampling:** Umbrella sampling was applied to enhance sampling conformations in the cyclization process. For comparing the difficulty of substrate cyclization in different mutants, we constructed three mutants (WT, V173A, V173Y) with 6S-5 substrate. An additional force was applied to 6S-5 to constrain the distances d(O<sub>3</sub>-C<sub>3</sub>). The reaction coordinate was defined as the distance between O<sub>3</sub> atom and C<sub>3</sub> atom in 6S-5 substrate. We scan along the reaction coordinate from 4.9 Å to 2.4 Å, taking 0.02 Å as the step length and adding a harmonic force constant of 200 kcal mol<sup>-1</sup> Å<sup>-2</sup> to it. Each step was performed 0.1 ns. There were 125 step and 12.5 ns MD simulation in each system, totally. Then we calculated the potential mean force (PMF) of each system through weighted histogram analysis method (WHAM).

**Binding free energy calculations:** The molecular mechanics generalized Born surface area (MM-GBSA) method was used to calculate the binding free energy between substrates and AmbDH3 with a python program MMPBSA.py.<sup>[21]</sup> 150 snapshots extracted from 20–50 ns trajectories were used to calculate the binding free energies. The decomposition of the energies was utilized to identify crucial residues making contributions to the recognition progress.

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}})$$

The binding free energy was calculated by the above equation, in which  $G_{\text{complex}}$ ,  $G_{\text{protein}}$ , and  $G_{\text{ligand}}$  represent the free energy of the complex, protein, and ligand, respectively.

**QM/MM calculations:** Taking computational cost and quality into consideration, a two-layered QM/MM ONIOM scheme embedded in Gaussian 16 program was employed by us to explore the reaction mechanism of AmbDH3.<sup>[22]</sup> The dominant clustered structure that originated from the system 2R6S-3 trajectories was chosen as the initial structure. In that representative structure, substrate and residues within 4 Å of the substrate were reserved. The total atom number for the QM/MM model was 482. There were 68 atoms in the QM region, which contained the side chain of H51 and D215, as well as the whole substrate (Figure S18†). Additionally, we froze the backbone of residues in order to keep the real protein architecture in its place.

The geometry structure optimization of the transition state (TSs) and intrinsic reaction coordinate (IRC) were calculated at the ONIOM (B3LYP/6-31G\*:Amber) level. All TSs structures were confirmed by an only imaginary frequency number with the correct vibrational direction. In order to obtain more accurate energy

profiles of the reaction, we recalculated the single point energy for each minimum using basis set 6-311+G\*\*. We treated the QM and MM layers' interaction with the electrostatic embedding formalism.<sup>[23]</sup>

## Acknowledgements

The authors thank the National Key R&D Program of China (2021YFC2100600, 2019YFA0905400), the National Science Foundation of China (32070041), and the SJTU-HPC computing facility award for financial support and computational time.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

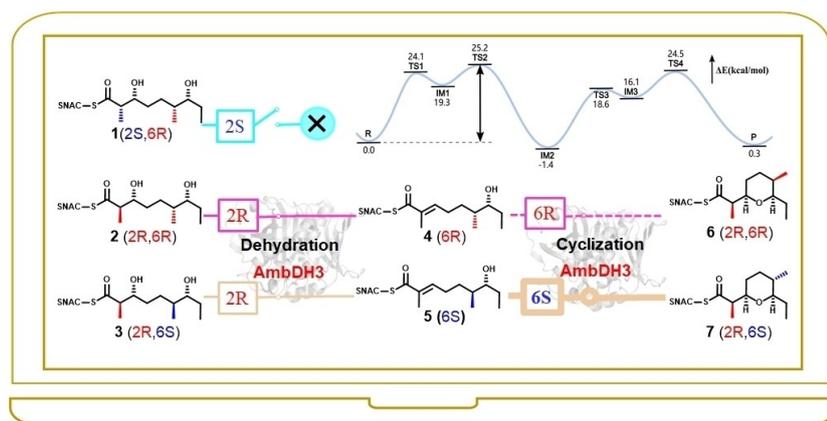
**Keywords:** cyclization · catalytic mechanism · dehydratase · MD simulation · substrate selectivity

- [1] a) E. Kalkreuter, G. J. Williams, *Curr. Opin. Microbiol.* **2018**, *45*, 140–148; b) D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2007**, *70*, 461–477; c) S. L. Robinson, J. K. Christenson, L. P. Wackett, *Nat. Prod. Rep.* **2019**, *36*, 458–475; d) M. Klaus, M. Grininger, *Nat. Prod. Rep.* **2021**, *38*, 1409–1409.
- [2] a) C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688–4716; *Angew. Chem.* **2009**, *121*, 4782–4811; b) A. T. Keatinge-Clay, *Nat. Prod. Rep.* **2012**, *29*, 1050–1073.
- [3] a) A. Keatinge-Clay, *J. Mol. Biol.* **2008**, *384*, 941–953; b) D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure* **2010**, *18*, 94–105; c) D. Gay, Y. O. You, A. Keatinge-Clay, D. E. Cane, *Biochemistry* **2013**, *52*, 8916–8928.
- [4] C. Risidian, T. Mozef, J. Wink, *Microorganisms* **2019**, *7*, 124–124.
- [5] G. Berkhan, F. Hahn, *Angew. Chem. Int. Ed.* **2014**, *53*, 14240–14244; *Angew. Chem.* **2014**, *126*, 14464–14468.
- [6] a) G. Berkhan, C. Merten, C. Holec, F. Hahn, *Angew. Chem. Int. Ed.* **2016**, *55*, 13589–13592; *Angew. Chem.* **2016**, *128*, 13787–13790; b) F. Taft, M. Brunjes, T. Knobloch, H. G. Floss, A. Kirschning, *J. Am. Chem. Soc.* **2009**, *131*, 3812–3813.
- [7] X. Q. Xie, A. Garg, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.* **2017**, *139*, 9507–9510.
- [8] L. Vetcher, H. G. Menzella, T. Kudo, T. Motoyama, L. Katz, *Antimicrob. Agents Chemother.* **2007**, *51*, 3734–3736.
- [9] a) T. Hollmann, G. Berkhan, L. Wagner, K. H. Sung, S. Kolb, H. Geise, F. Hahn, *ACS Catal.* **2020**, *10*, 4973–4982; b) L. Wagner, T. Ross, T. Hollmann, F. Hahn, *RSC Adv.* **2021**, *11*, 20248–20251.
- [10] K. H. Sung, G. Berkhan, T. Hollmann, L. Wagner, W. Blankenfeldt, F. Hahn, *Angew. Chem. Int. Ed.* **2018**, *57*, 343–347; *Angew. Chem.* **2018**, *130*, 349–353.
- [11] a) X. P. Chen, T. Shi, X. L. Wang, J. T. Wang, Q. H. Chen, L. Q. Bai, Y. L. Zhao, *ACS Catal.* **2016**, *6*, 4369–4378; b) T. Shi, L. X. Liu, W. T. Tao, S. G. Luo, S. B. Fan, X. L. Wang, L. Q. Bai, Y. L. Zhao, *ACS Catal.* **2018**, *8*, 4323–4332; c) S. A. Serapian, M. W. van der Kamp, *ACS Catal.* **2019**, *9*, 2381–2394; d) C. C. Jiang, B. B. He, R. L. Zhao, M. J. Xu, K. N. Houk, Y. L. Zhao, *ACS Catal.* **2021**, *11*, 7928–7942.
- [12] L. Liu, Q. Yu, H. Q. Zhang, W. T. Tao, R. F. Wang, L. Q. Bai, Y. L. Zhao, T. Shi, *Catal. Sci. Technol.* **2021**, *11*, 2155–2166.
- [13] F. E. Medina, R. P. P. Neves, M. J. Ramos, P. A. Fernandes, *ACS Catal.* **2018**, *8*, 10267–10278.
- [14] a) G. J. Dodge, A. Patel, K. L. Jaremko, J. A. McCammon, J. L. Smith, M. D. Burkart, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6775–6783; b) M. Leesong, B. S. Henderson, J. R. Gillig, J. M. Schwab, J. L. Smith, *Structure* **1996**, *4*, 253–264.
- [15] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *Nucleic Acids Res.* **2004**, *32*, W665–W667.
- [16] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- [17] P. W. Payne, *J. Am. Chem. Soc.* **1977**, *99*, 3787–3794.
- [18] a) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197; b) T. Lu, F. W. Chen, *J. Comput. Chem.* **2012**, *33*, 580–592.
- [19] a) J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713; b) D. Case, I. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. Cheatham, V. W. D. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, H. Gohlke, A. Götz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, P. A. Kollman, *Amber 2018*, University of California: San Francisco, CA, **2018**.
- [20] a) T. A. Darden, D. M. York, L. G. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092; b) J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [21] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg, *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
- [22] a) T. Vreven, M. J. Frisch, K. N. Kudin, H. B. Schlegel, K. Morokuma, *Mol. Phys.* **2006**, *104*, 701–714; b) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox in *Gaussian 16 Rev. C.01*, Wallingford, CT, **2016**.
- [23] G. A. Cisneros, J. P. Piquemal, T. A. Darden, *J. Phys. Chem. B* **2006**, *110*, 13682–13684.

Manuscript received: November 3, 2022

Accepted manuscript online: December 4, 2022

Version of record online: ■■■, ■■■■



The theoretical study uncovers both the structural rationale for substrate specificity of the first bifunctional dehydratase-cyclase AsmDH3 and a

complete one-base catalytic mechanism including dehydration and cyclization.

Z. Du, Y. Li, Y. Liu, Prof. T. Shi\*

1 – 10

Molecular Insights into Bifunctional Ambruticin DH3 for Substrate Specificity and Catalytic Mechanism

