PAPER

Check for updates

Theoretical study on substrate recognition and catalytic mechanisms of gephyronic acid dehydratase DH1†

Lei Liu,‡ Qian Yu,‡ Haoqing Zhang, Wentao Tao, Rufan Wang, Linquan Bai, Yi-Lei Zhao [ID] and Ting Shi [ID] *

Dehydratase (DH), a domain in polyketide synthase (PKS) modules, can catalyze the dehydration of β-hydroxy to an α,β-unsaturated acyl intermediate. As the first dual-function (dehydratase/isomerase) DH domain accessible in the PDB database, the gephyronic acid (GphF) DH1 domain from the PKS biosynthetic pathway attracts great attention from researchers. However, the mechanisms of dehydration and isomerization in type I PKS still remain unclear. In this study, MD simulations and QM/MM calculations were combined to elucidate the molecular mechanism of GphF DH1. The results indicated that GphF DH1 had better recognition effect towards the (2R,3R)-substrate and preferred forming the α–β double bond in advance to the β–γ double bond directly. By calculating the binding energy, some key residues near the active pocket were highlighted. Umbrella sampling results showed that non-methylated substrates could form intra-molecular hydrogen bonds more easily than α-methyl substrates. The QM/MM calculations with the M062X/6-311+G**//M062X/6-31G* method and SMD solvation correction supported the one-base dehydration and one-base isomerization mechanism with energy barriers of 27.0 kcal mol^{-1} and 17.2 kcal mol^{-1}, respectively. These results can encourage future studies for the comprehensive understanding of the catalytic mechanism of PKS DHs and for the rational design of typical DHs.

Introduction

Polyketide synthases (PKSs) deploy a modular biosynthetic strategy to produce diverse bioactive molecules, most of which have important medicinal applications such as antibiotics, anti-cancer agents, and immunosuppressants.[1–4]

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, China. E-mail: tshi@sjtu.edu.cn; Fax: +86 21 34207347; Tel: +86 21 34207347

† Electronic supplementary information(ESI) available: The schematic diagram of the whole system (large model, gray) and the truncated system (small model, blue) for QM/MM calculations; RMSD values of 3 × 50 ns in sub3, sub4, sub5, sub2-A and sub2-B systems; the distributions of $d(O_1–O_{D1898})$ and $d(N_{εH1735}–H_α/H_γ)$ in sub9-A and sub9-B systems; the energy profiles for the transfer of $H_α/H_γ$ to $N_{εH1735}$ in sub9-A and sub9-B; a hydrogen bond in the sub3 system; the oxyanion hole function performed by residue Gly1745; a hydrogen bond between the O3 atom and NH group of V1746 in the sub3 system; hydrophobic interactions between substrates and GphF DH1 in sub2 and sub3 systems; RMSD values of 3 × 50 ns MD simulations in the sub9 system; the structures of TS1 and IM in Path1 calculated by small model QM/MM calculations; energy profiles of $C_β–O_1$ bond breaking and structures in the sub9 system for Path2; distances and angles in key structures, ONIOM calculation energies and energetic corrections at different levels with the M062X method in the sub9 system; distances in key structures and energetic corrections at different levels with the M062X method the in sub10 system. See DOI: 10.1039/d0cy01776k

‡ L. L and Q. Y. contributed equally to this work.

To assemble an intact polyketide chain, PKS modules minimally include an acyltransferase (AT) responsible for loading a specific chain extension unit, an acyl carrier protein (ACP) with a growing polyketide chain covalently tethered, and a ketosynthase (KS) that catalyzes the chain elongation reaction. Additionally, variable combinations of other domains, such as ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), and methyltransferase (MT), can decorate the polyketide chain to expand the product pool.

One or more DH domains are organized in PKSs, responsible for the dehydration of β-hydroxy to an α,β-unsaturated acyl intermediate. Besides the typical dehydration function, several DH domains have additional catalytic activities such as double-bond isomerization,[5] pyran formation[6,7] and methyl group epimerization.[8] These special functions set the stage for greater chemical diversities in products. In addition, compounds with either α–β double bonds or β–γ double bonds are very common in polyketides, such as rhizoxin[9] and bacillaene.[10] Therefore, DHs are notable for their catalytic activities in polyketide biosynthesis.

As the literature states, PKS DH domains display characteristic α + β double hotdog folds with a largely hydrophobic substrate binding pocket, in which an universally conserved dyad of His–Asp/Glu residues is positioned.[11–13] Inspired largely by the observed spatial

relationship between the substrate and the His–Asp/Glu dyads, a base–acid dehydration mechanism has been generally accepted.[13–15] It is thought that His serves as the general base to remove the α-H of the substrate, and Asp or Glu can act as the general acid to protonate the β-hydroxyl, thus promoting the elimination of a water molecule. This mechanism was clarified by QM/MM calculations in 2018.[16] However, in 2019, the study by Cane et al. supported a simple one-base mechanism that only the His residue acted as the base to promote the elimination of a water molecule at the conserved active site.[17]

In 2018, the crystal structures of the first dual-function (dehydratase/isomerase) gephyronic acid DH domain (GphF DH1) from the PKS biosynthetic pathway were reported,[18] making it possible to explore the atomic mechanisms of DH for dehydration and isomerization. What's more, the GphF DH1 reactivity with NAC-linked thioesters of all predicted natural diastereomers (four potential α-methyl-β-hydroxy substrates) was detected by LC-MS analysis, and the (2R,3R)-substrate was predicted to be the natural substrate for GphF DH1. Meanwhile, GphF DH1 showed tolerance to methylated and non-methylated substrates (Fig. 1A). It was proposed that dehydration took place firstly and then isomerization, as shown in Fig. 1B. However, why does GphF DH1 have tolerance to different diastereomers and analogs? What are the molecular mechanisms of dehydration and isomerization? These questions remain unanswered.

In this study, high-resolution models of the GphF DH1 domain with different substrates were employed to explore the substrate recognition and catalytic mechanism by combining MD simulations and QM/MM calculations. The conformational characteristics that facilitate the formation of α–β and β–γ double bonds were investigated. Binding free energies were calculated to find out key residues that played a critical role in substrate recognition. And umbrella sampling was utilized to detect the intra-molecular hydrogen bond in α-methyl and non-methylated substrates. Finally, QM/MM calculations were carried out to support the one-base dehydration and one-base isomerization mechanism.

Our study provides detailed mechanistic insights into this unique dual-function dehydratase GphF DH1.

Materials and methods

Setup of systems

In 2018, Smith et al. reported three crystal structures of GphF DH1 (PDB number: 6mbf, 6mbh and 6mbg).[18] Since there were more missing structures in the wild type (PDB number: 6mbf) and the P1711L mutation had no effect on the activity of this enzyme, mutant P1711L (PDB number: 6mbg) was chosen as the initial structure and the missing six residues (SKAALL) in the C-terminal were modelled from the structure (PDB number: 6mbg) by using PyMOL software. The protonation states of the titratable residues (such as His, Glu, and Asp) in GphF DH1 were determined (pH: 7.5, force field: Amber) using the PDB2PQR web server.[19] Although the pK_a of Asp1898 in the catalytic dyad was 6.92, considering that Asp1898 would form water with β-hydroxyl in the catalytic reaction, it was defined as Ash, and the rest of the Asp had a negative charge. The pK_a of Glu1750 was 8.93, defined as Glh. For histidine, residues defined as Hie were: 1703, 1801, 1876, 1893 1914, 1939, 1943, 1954 and residues defined as Hid included 1735, 1817, 1865, 1911, 1950. Next, molecular docking was performed by utilizing the LibDock module in Discovery Studio 3.5. For substrates 2, 3, 4, 5 and 9, the conformations with a high score and substrate's orientation consistent with the crystal structure (PDB number: 5tz6 (ref. 20)) were selected. For sub10, the conformation with the highest score was selected as the initial structure. For the preparation of substrate parameters, after the optimization at the level of pm3 (ref. 21) and HF/6-31G(d),[22] the electrostatic surface potential (EPS) was calculated. A two-step restrained electrostatic potential (RESP)[23] charge fitting method was applied to generate the bonds, angles, dihedral angles, and van der Waals radii parameters for the substrate using the Antechamber package. Through tleap, the system was solvated in an octahedral box of TIP3P water. Due to the whole system being electrically neutral, there was no need to add chlorine or sodium ions to balance the simulation system.
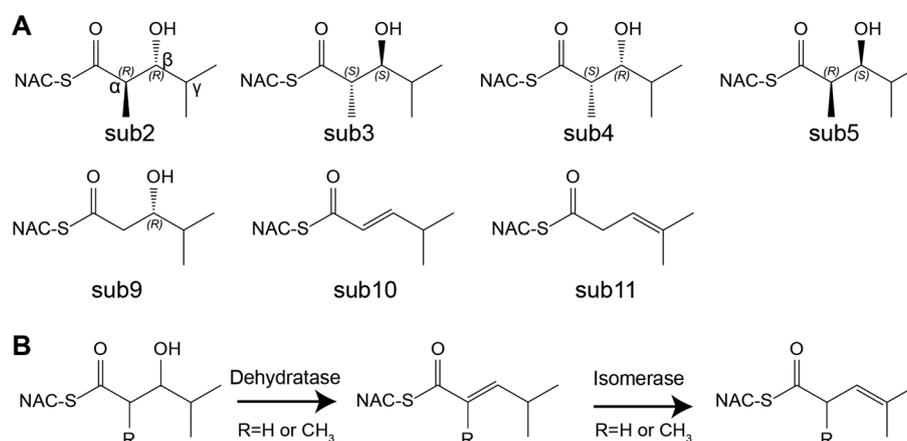
Fig. 1 (A) Chemical structures of substrates reported in ref. 18. (B) Process of dehydration and isomerization catalyzed by GphF DH1.

2156 | Catal. Sci. Technol., 2021, **11**, 2155–2166

This journal is © The Royal Society of Chemistry 2021

Molecular dynamics simulations

Classical MD simulations were carried out on the prepared structures of six systems utilizing the AMBER14 (ref. 24) program suite with ff03.r1 force field. In the MD simulations, the particle mesh Ewald (PME)[25] method was employed for long-range electrostatic interactions, the SHAKE algorithm[26] was used to fix bonds and angles involving hydrogen atoms, and the non-bonded cutoff was set to 10.0 Å. To remove atomic collisions, a two-step energy minimization was performed, first for the water molecules, and then the rest of the overall system. After that, the system was gradually heated from 0 K to 300 K in 50 ps. Next, the system was switched to constant pressure and temperature (NPT), and equilibrated for 50 ps to adjust the system to the correct density. After an additional 1 ns of NPT simulation, constraints of $d(O_{D1898}-O_1)$ and $d(N_{\epsilon H1735}-H_\alpha)$ or $d(O_{D1898}-O_1)$ and $d(N_{\epsilon H1735}-H_\gamma)$ were added by using a harmonic vibrational potential with a force constant of 20 kcal mol^{-1} Å^{-2} at 3 Å and 1.8 Å, respectively, as shown in Fig. 2. Then the constraints were removed, and the systems were re-equilibrated with the same procedure. Finally, taking the equilibrated structure as the initial structure, 50 ns trajectories were collected three times for further analysis.

Umbrella sampling

In sub2 and sub9 systems, umbrella sampling was used to enhance sampling. To ensure the pre-reaction state conformations, two additional forces were applied to substrates to constrain the distances $d(O_{D1898}-O_1)$ and $d(N_{\epsilon 1735}-H_\alpha/H_\gamma)$. The distance of the H atom and O_2 atom was defined as a reaction coordinate and scanned closer, taking 0.03 Å as the step length from 4.5 Å to 1.8 Å by adding a harmonic force constant of 200 kcal mol^{-1} Å^{-2} to it. 0.1 ns MD simulations were carried out in each bin, for a total of 9 ns MD simulations. Then the potential of mean force (PMF) of the systems was computed *via* the weighted histogram analysis method (WHAM).[27]

Binding free energy calculations using MM-GBSA and MM-PBSA

The molecular mechanics generalized Born surface area (MM-GBSA) method and molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method were used to calculate the binding free energy between substrates and the
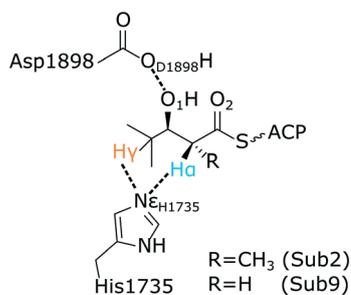
DH enzyme with a python program MMPBSA.py.[28] The last 40 ns frames of each trajectory were extracted to calculate the binding free energy according to the following equation:

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

Here, $G_{complex}$, $G_{protein}$ and G_{ligand} represent the free energy of the protein–ligand complex, protein and ligand, respectively.

After the completion of MM-GBSA and MM-PBSA calculations, the calculated free energy was decomposed into specific residue contributions to identify crucial amino acid residues that play an important role in the interaction between substrates and GphF DH1.

QM/MM calculations of reaction profiles

In this study, QM/MM calculations with two models were performed by using a two-layered ONIOM method[29,30] implemented in the Gaussian 09 program.[31] Two mechanisms were proposed for the dehydration. Path1, it is initiated by the deprotonation of the α-carbon in the substrate, concomitantly with β-elimination of the β-hydroxyl. Path2, the elimination of the β-hydroxyl takes place firstly, and then the deprotonation proceeds. Herein, the large model was used to obtain an approximate structure of the transition state for Path1, and then the small model was used to calculate the accurate transition state and perform intrinsic reaction coordinate calculations (Fig. S1†). For the large model, the conformation was extracted with a 4.5 Å external water layer, which was close to the average of the most dominant cluster obtained from MD trajectories and contained the whole enzyme. The QM region included the side chain of His1735, Tyr1856, Asp1898 and Gln1902, the full residues of Leu1744 and Gly1745, the carbonyl group of Val1743, and a stable water molecule. The atom number in the QM region was 113, and the atom number of the whole system is 9293 in total. The small model was truncated from the large model, consisting of the substrate and the residues within 4.5 Å of the substrate, 30 residues and 10 water molecules near the catalytic center. The QM region consisted of the side chain of His1735 and Asp1898, and a stable water molecule. The whole atom number in the small model was 539. In addition, the backbone of all residues was frozen to maintain the real protein architecture. What's more, three replicas of the small model were chosen to improve the accuracy of the energy profile. It should be noted that for Path2 the energy barrier is much higher for the elimination of β-hydroxyl in the large model, meaning the dehydration reaction could not proceed *via* Path2, so accurate and time-consuming small model calculations are not necessary, which focuses on the precise search of the transition state and calculations of intrinsic reaction coordinates.

For isomerization calculations, the re-equilibrated structure was extracted to carry out QM/MM calculations in the sub10 system. The small model included 30 residues, 5 water molecules and sub10, 589 atoms in total. The QM

Fig. 2 Substrate and key residues with constraint distances in computational models.

region contained the side chain of His1735, Tyr1856, Asp1898, Gln1902, the whole residues Leu1744, Gly1745 and the carbonyl group of Val1743, 107 atoms in total in both small model and large model QM/MM calculations. Similarly, the large model was used to obtain an approximate structure of the transition state and the small model was used to obtain the accurate transition state and other stable states.

Single point optimization, linear scanning, transition state optimization and characterization of critical points by frequency, and intrinsic reaction coordinate (IRC) calculations were carried out at the ONIOM (M062X[32]/6-31G*:[33] Amber) level. These TSs were confirmed by vibrational frequency calculations, resulting in a single imaginary frequency with the correct transition vectors assigned. Finally, in order to improve the accuracy of the calculations, single-point energy calculations were performed on the optimized structures using a larger basis set and solvation model based on density (SMD)[34] solvation correction.

Results and discussion

Substrate specificity of GphF DH1

PKS DHs generally have strictly chiral selectivity on substrates. For example, RifDH10 catalyzes the stereospecific *syn* dehydration of the model substrate (2S,3S)-2-methyl-3-hydroxypentanoyl-RifACP10, but does not dehydrate any of the other three diastereomers.[11]

To explain the stereospecificity of GphF DH1, four models with different chiral substrate candidates (sub2, sub3, sub4, sub5) were constructed. Based on the results of molecular docking, the distances of O_1-O_{D1898}, which represents the protonation of the β-hydroxyl oxygen and the elimination of a water molecule, were restrained to 3.0 Å. Also, the distances between the H_α and the N_ϵ of H1735 were constrained to 1.8 Å, indicating the removal of H_α from the substrate. Then these restrains were removed and 50 ns MD simulations were performed three times in each system. The RMSD values indicated that the four systems reached the equilibrium (Fig. S2†).

According to our previous studies,[35,36] analyses of conformations obtained from MD simulations with key reaction coordinates, especially the pre-reaction state (PRS), can be used to evaluate the recognition and interaction between the substrate and the enzyme. Here the distances $d(N_{\epsilon H1735}-H_\alpha)$ and $d(O_1-O_{D1898})$ were chosen and analyzed. As shown in Fig. 3A, the population of the two distances was concentrated in the sub2 system. The average of $d(O_1-O_{D1898})$ was 2.7 Å and that of $d(N_{\epsilon H1735}-H_\alpha)$ was 2.9 Å (Fig. 3B). These conformations demonstrated that the O_1 atom could form stable hydrogen bond interaction with D1898 and the H_α atom was very close to the N_ϵ atom, which could facilitate the dehydration of GphF DH1. However, in other systems, either $d(O_1-O_{D1898})$ or $d(N_{\epsilon H1735}-H_\alpha)$ was unsatisfactory. For example, the average of $d(N_{\epsilon H1735}-H_\alpha)$ was observed to be 4.4 Å in the sub3 system, much larger than that in the sub2 system. For the sub4 system, a proper $d(N_{\epsilon H1735}-H_\alpha)$ was
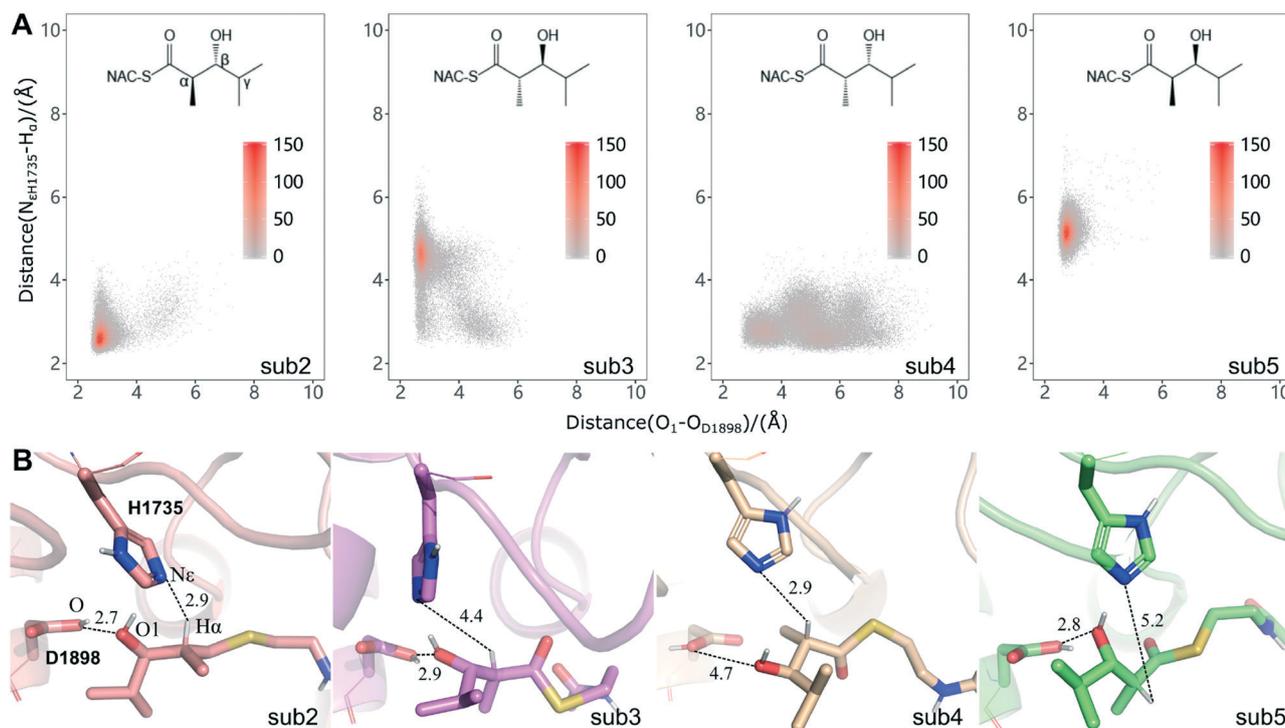
Fig. 3 (A) The distribution of two key distances $d(O_1-O_{D1898})$ and $d(N_{\epsilon H1735}-H_\alpha)$ in four systems (sub2, sub3, sub4, sub5); (B) the four representative structures from their dominant clusters obtained from MD simulations.

obtained as 2.9 Å, while the improper $d(O_1-O_{D1898})$ was observed at 4.7 Å. In the sub5 system, $d(N_{\varepsilon H1735}-H_\alpha)$ even increased to 5.2 Å. Nearly zero proportion of the PRS was noticed in the sub5 system, which was well consistent with the experimental result that GphF DH1 cannot catalyze the dehydration reaction in the sub5 system.[18] To sum up, the largest proportion of the PRS in the sub2 system shows the great substrate specificity of the enzyme, indicating that the natural substrate of GphF DH1 is likely the (2*R*,3*R*)-substrate. This is consistent with experimental observations.[18]

α–β double bond and β–γ double bond

According to previous studies, two possible mechanisms can be used to explain the formation of a β–γ double bond: (1) enzymes directly catalyze the formation of a β–γ double bond; (2) enzymes catalyze the formation of an α–β double bond firstly, and then isomerize it to a β–γ double bond. To clarify the mechanism of β–γ double bond formation in GphF DH1, we constructed two models with the favorable substrate of GphF DH1 as sub2-A and sub2-B. With the same strategy mentioned above, the distances of O_1-O_{D1898} were constrained to 3.0 Å, and the distances of $N_\varepsilon-H_\alpha$ or $N_\varepsilon-sH_\gamma$ were constrained to 1.8 Å, indicating the removal of H_α or H_γ from sub2. After equilibration, the constrains were revoked and 50 ns MD simulations were performed three times in both systems.

According to the RMSD values shown in Fig. S3,† two systems achieved their equilibrium quickly. Although a stable hydrogen bond $(O_1-H\cdots O_{D1898})$ could be observed in both systems, there were some differences in the distribution of $d(O_1-O_{D1898})$. As shown in Fig. 4A, the average distance in the sub2-A model was 2.7 Å, but it enlarged to 3.2 Å in the sub2-B model, suggesting the stronger interaction between the β-hydroxyl oxygen of sub2 and D1898 of GphF DH1 in the sub2-A model. Then, we evaluated the proton transfer by
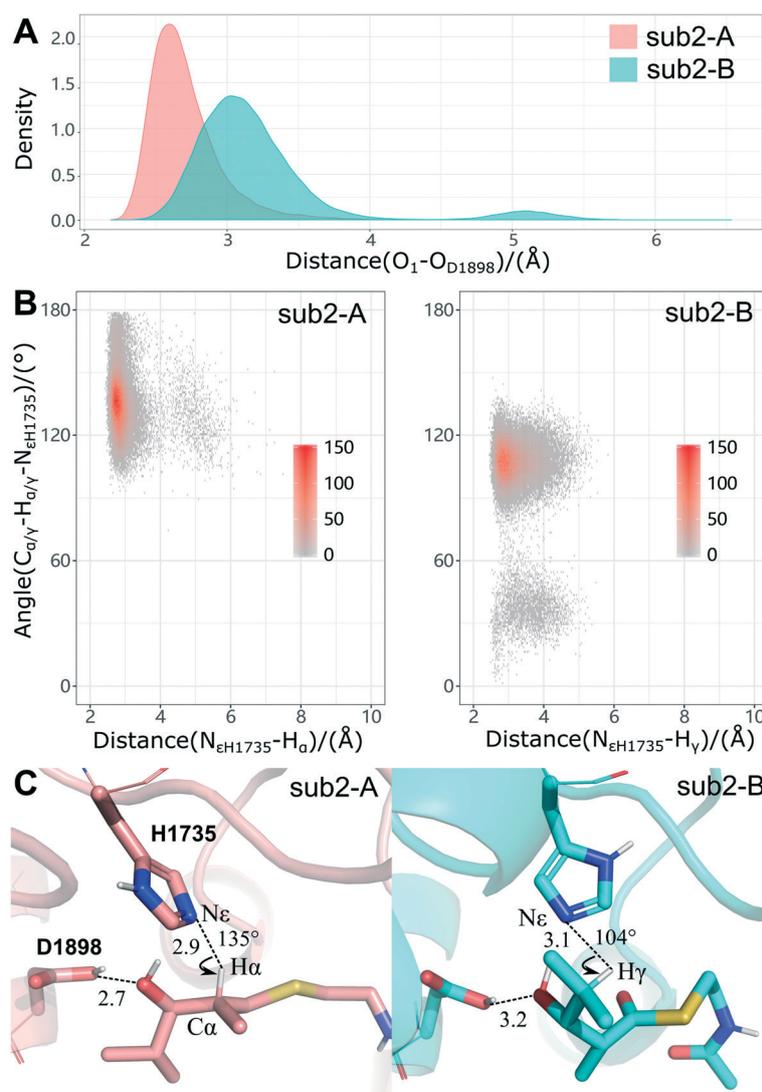
Fig. 4 Comparison of the sub2-A model and sub2-B model. (A) The distribution of $d(O_1-O_{D1898})$ in two models; (B) the distributions of $d(N_{\varepsilon H1735}-H_\alpha/H_\gamma)$ and angle $(C_\alpha-H_\alpha-N_{\varepsilon H1735}/C_\gamma-H_\gamma-N_{\varepsilon H1735})$ in two models; (C) the most representative structures of two models in MD simulations.

measuring the distance of $N_{\varepsilon H1735}$–H_α/H_γ and the angle of C_α/C_γ–H_α/H_γ–$N_{\varepsilon H1735}$ (Fig. 4B). It can be seen that the distances of $N_{\varepsilon H1735}$–H_α/H_γ were similar, the mean was 2.9 Å in model sub2-A and 3.1 Å in the sub2-B model, meaning the slight advantage in transfer of H_α in the sub2-A model. More importantly, the average angle of C_α–H_α–$N_{\varepsilon H1735}$ (135°) in the sub2-A model was more favorable for the proton transfer than that (C_γ–H_γ–$N_{\varepsilon H1735}$ is 104°) in the sub2-B model.

A similar phenomenon was also observed in sub9-A and sub9-B models. As shown in Fig. S4,† the distributions of $d(O_1$–$O_{D1898})$ and $d(N_{\varepsilon H1735}$–$H_\alpha/H_\gamma)$ indicated the superiority of α–β double bond formation (sub9-A). Moreover, three frames, obtained from sub9-A and sub9-B, respectively, were used to calculate the reaction energy profile by the QM/MM method. It was found that the energy barriers of the first step were quite different in these two pathways. As shown in Fig. S5,† the energy barrier of transferring H_γ to $N_{\varepsilon H1735}$ (more than 40 kcal mol^{-1}) was obviously higher than that of transferring H_α to $N_{\varepsilon H1735}$ (less than 15 kcal mol^{-1}).

Taken together, besides the superiority of the stable hydrogen bond between O_1 and O_{D1898}, a closer distance of $N_{\varepsilon H1735}$–H_α, a proper angle of C_α–H_α–$N_{\varepsilon H1735}$, and the lower energy barrier of transferring H_α to $N_{\varepsilon H1735}$ made it easier for GphF DH1 to form an α–β double bond than a β–γ double bond.

Key residues near the active pocket

MM-GBSA and MM-PBSA methods were utilized to evaluate the binding energies of four systems. As shown in Table 1, the binding energy in the sub3 system was found to be the lowest. The average energy was −35.6 kcal mol^{-1} in MM-GBSA and −29.5 kcal mol^{-1} in MM-PBSA, suggesting that GphF DH1 had stronger affinity to sub3 than other substrates. However, the close interaction with GphF DH1 did not improve the dehydration of sub3, because most conformations were not conducive to transferring proton H_α with the inappropriate distance of $N_{\varepsilon H1735}$–H_α, which was generally larger than 4.0 Å in MD simulations.

Table 1 The binding energy calculated by MM-GBSA and MM-PBSA methods in four systems

Complex	MMGBSA (kcal mol^{-1})	Average	Std. err. of mean	MMPBSA (kcal mol^{-1})	Average	Std. err. of mean
sub2	−30.3	−30.2	0.9	−24.2	−21.9	1.3
	−32.0			−23.9		
	−28.7			−20.1		
sub3	−35.4	−35.6	0.1	−29.0	−29.5	0.2
	−35.6			−29.7		
	−35.6			−29.7		
sub4	−28.1	−30.5	1.3	−16.8	−18.7	1.6
	−31.2			−17.6		
	−32.3			−21.8		
sub5	−29.7	−31.0	0.7	−20.4	−21.3	1.0
	−31.9			−20.3		
	−31.4			−23.3		

To uncover the origin of affinity and the role of key residues in recognition and catalysis, the energy contribution of individual amino acids was calculated in each system. Thirteen amino acids were obtained near the active pocket (energy contributions ≥0.5 kcal mol^{-1}) and they all were supposed to play important roles in the recognition and catalysis process (Fig. 5A).

Firstly, as the catalytic dyad, the function of His1735 and Asp1898 in the catalytic process is crucial, although in EI, sometimes, only His acts as the catalytic base, and Asp is usually replaced with Asn[37] (Fig. 5B).

Secondly, Val1742, replaced by L51 in EyrDH4,[13] promotes the binding of all substrates in our systems, in which the carbonyl of the Val1742 backbone can form hydrogen bonds with the key residue His1735, thus contributing to α-deprotonation.

Thirdly, in most DHs, the position of L1744 in a loop region is occupied by Pro, which bears a small side chain. In GphF DH1 and EI, the residue usually is Leu or Val, which restricts the pocket and adjusts the substrate with its large side chain group.

Fourthly, three relatively conserved amino acids Gly1745, Val1746, and Thr1747 were observed. They are located at the α helix and Gly1745 can stabilize the substrate conformation by forming a hydrogen bond with the O_2 atom of the substrate (Fig. S6†). Previous experiments indicated that the absence of the hydrogen bond between the substrate and Gly888 in the mammalian DH (corresponding to Gly1745) resulted in unfavorable α-deprotonation.[16] Besides, it should be noted that Val1746 is beneficial for binding with sub3, but against sub2, owing to a hydrogen bond between the O_3 atom and N atom of Val1746's main chain in the sub3 system (Fig. S7†).

Fifthly, the mutation experiment of Phe3746 in PIK DH2 corresponding to Tyr1856 in GphF DH1 has proved that Tyr1856 can affect the binding of substrates,[38] while the mutation of Y1856F can affect the isomerization reaction of substrates.[18]

Finally, residues 1918–1921, as well as Phe1905, are five large hydrophobic amino acids that reside in the active pocket and provide a relatively hydrophobic environment for substrates (Fig. 4B and S8†). These results highlighted the key residues near the active pocket, which might dramatically influence the substrate recognition and the catalytic activity of GphF DH1.

Effect of α-methyl in sub2

Some modules harbor an integrated *S*-adenosylmethionine (SAM)-dependent *C*-methyltransferase (*C*-MT) domain that can install 2-methyl groups. 98% *trans*-AT PKSs harbor one to ten *C*-MTs, while embedded *C*-MTs are considerably less common in *cis*-AT PKS systems, with 5% harboring one or two *C*-MTs.[39] Interestingly, among *cis*-AT PKS systems, only gephyronic acid synthase contains as many as seven *C*-MTs.[40] Experimental results showed that GphF DH1 could catalyze α-methyl sub2 and non-methylated sub9 to produce
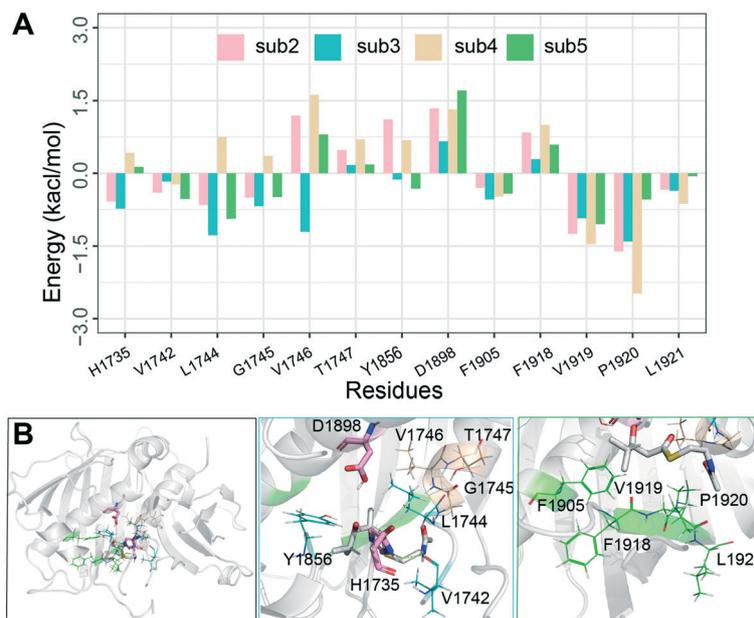
Fig. 5 (A) Energy decomposition of residues; (B) position of some key residues. The left is an overview of the whole structure. The middle shows the positions of key residues in an α helix, catalytic dyad colored pink, and Val1742, Leu1744, and Tyr1856 colored blue. The right shows five large hydrophobic residues in the active pocket colored green.

dehydration products and isomeric products, which indicated that bifunctional GphF DH1 was tolerant to methyl and non-methylated substrates.[18]

In order to explain the function of α-methyl during the dehydration reaction, the comparison of sub2 and sub9 systems was performed. RMSD values (Fig. S9†) showed that
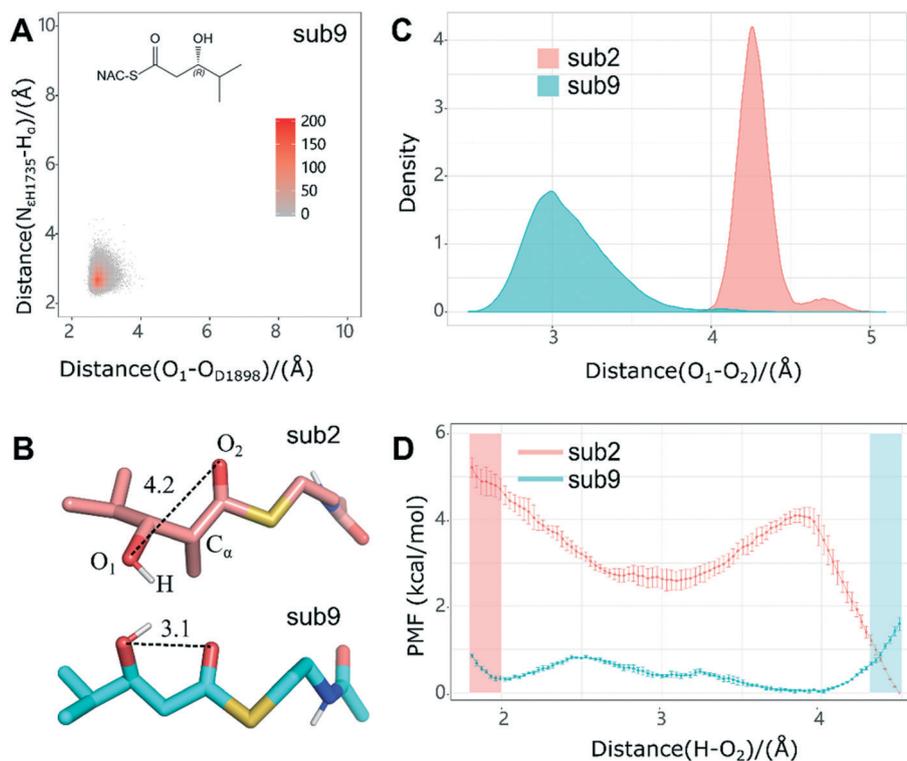
Fig. 6 (A) The distribution of two key distances $d(O_1–O_{D1898})$ and $d(N_{\varepsilon H1735}–H_\alpha)$ in the sub9 system; (B) structures of sub2 and sub9 in the GphF DH1 pocket; (C) the distribution of distance $d(O_1–O_2)$ in substrates; (D) the potential of mean force (PMF) of sub2 and sub9 systems obtained by umbrella sampling.

This journal is © The Royal Society of Chemistry 2021

Catal. Sci. Technol., 2021, **11**, 2155–2166 | **2161**

the sub9 system reached its equilibrium. According to Fig. 6A, the distribution of key distances $d(O_1–O_{D1898})$ and $d(N_{\varepsilon H1735}–H_\alpha)$ in the sub9 system was dense, leading to a similar ratio of PRS to the sub2 system in Fig. 3A. Thus, we proposed that GphF DH1 had no significant differences in recognition and catalysis of these two substrates. By analyzing the structural feature of sub2 and sub9 systems (Fig. 6B and C), it was found that there was an intra-molecular hydrogen bond in sub9, where the average of $d(O_1–O_2)$ was 3.1 Å, while owing to the large 2-methyl group in sub2, it was difficult to form a similar intra-molecular hydrogen bond, where the average of $d(O_1–O_2)$ was 4.2 Å. By umbrella sampling (Fig. 6D), we found that it needed to overcome at least 4.0 kcal mol^{-1} energy barrier to form an intra-molecular hydrogen bond for sub2, while the PMF was gentle in the sub9 system and the conformations with the intra-molecular hydrogen bond were favorable in energy. In conclusion, the α-methyl group plays an important role in adjusting the conformation of substrates in the GphF DH1 active pocket.

The mechanism of dehydration

A general mechanism has been put forward that the dehydration is initiated by the deprotonation of the α-carbon of the substrate, concomitantly with β-elimination of the β-hydroxyl, as shown in Fig. 7A Path1. In 2018, QM/MM calculations of dehydratase domains from the mammalian fatty acid synthase proved this proposal.[16] But according to the knowledge of organic chemistry, carbenium ions have been regarded as intermediates of the catalytic reactions on zeolites,[41,42] which means the elimination of the β-hydroxyl firstly, and then the deprotonation proceeds, as shown in Fig. 7A Path2. The two mechanisms are mainly different in the sequence of the removal of the H_α atom and the hydroxyl group. Although they both seem reasonable, we prefer the Path1 mechanism because of the lower energy barrier obtained by QM/MM calculations than that of Path2 in the GphF DH1 system.

To uncover the molecular mechanism of GphF DH1, small model QM/MM calculations and large model QM/MM calculations were carried out on GphF DH1. Based on the study that the intra-molecular hydrogen bond contributes to the decrease of the energy barrier of dehydration,[16] the sub9 system was chosen to investigate the mechanism of GphF DH1 via QM/MM calculations. To improve the calculation accuracy, three replicas of small model QM/MM calculations were utilized at the ONIOM (M062X/6-31G*: Amber) level and single point calculations were performed at the ONIOM (M062X/6-311+G**: Amber) level with SMD solvation correction. The energy profile of the reaction mechanism was obtained by averaging three replicas. For the first step, the elimination of H_α to form an enol intermediate, 17.7 kcal mol^{-1} energy barrier was required, as shown in Fig. 7B. In the representative structure from three replicas, the first transition state (TS1) was observed when the distance of H_α–
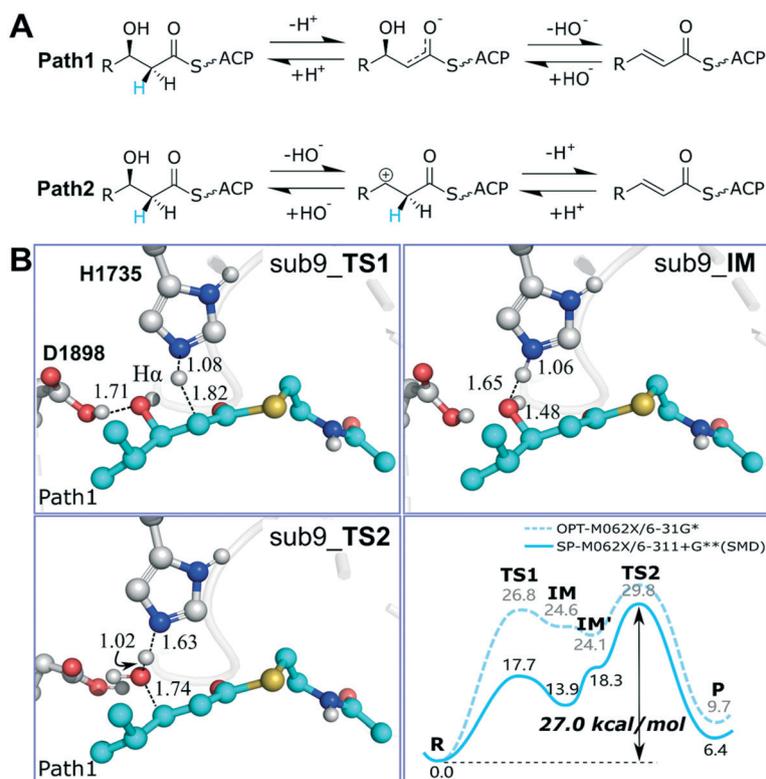
Fig. 7 (A) Two possible dehydration pathways; (B) representative optimized structures of transition states (TS1 and TS2) and the intermediate (IM) with the average energy profile of three replicas for Path1 calculated by small model QM/MM calculations.

$N_{\varepsilon H1735}$ was located at 1.08 Å and that of H_α–C_α was located at 1.82 Å (Table S1†). The imaginary frequency was $133.03i$ cm^{-1}, and it was characterized by an antisymmetric stretching of C_α–H_α and H_α–N_ε bonds. Although the distance between the hydrogen atom (H_α) and the nitrogen of His1735 (N_ε) changed little (from 1.08 to 1.06 Å), the distances of C_α–N_ε and O_2–N_ε changed obviously (2.85 and 3.01 Å in TS1 and 3.11 and 2.70 Å in IM, Fig. S10†), which means His1735 moved away from substrate C_α to approach substrate O_2 along with the proton transfer. After enol rearrangement, β-hydroxyl elimination took place, and a water molecule was generated with a reaction energy barrier of 13.1 kcal mol^{-1}. The vibrational analysis was performed on the structure of TS2 in a similar way, and an imaginary frequency of $271.16i$ cm^{-1} was observed, which corresponds to the stretching of $C_\beta\cdots O_1H$ and $H_\alpha\cdots O_1H$, indicating the formation of a water molecule. In TS2, the distance of C_β–O_1 was located at 1.74 Å and the distance of O_1–H_α was located at 1.02 Å. The whole energy barrier of the dehydration through enol intermediate was calculated to be 27.0 kcal mol^{-1}. We found that the energy barriers of β-hydroxyl elimination, as the rate-determining step, were 22.2, 21.2 and 23.5 kcal mol^{-1} in QM calculations, much lower than 27.0 kcal mol^{-1} in ONIOM calculations, indicating the large influence of the MM environment (Table S2†). Besides, considering its low catalytic efficiency (only 28% conversion to the α–β double bond intermediate), we think the relatively high energy barrier is basically consistent with the experiments.

As mentioned in the study by Smith et al.,[18] in FAS DH the His878-imidazole acts as a base to deprotonate the C_α of the β-hydroxyacyl (HAC) substrate and the β-elimination of the β-hydroxyl of HAC proceeds with late protonation of the leaving hydroxide by the Asp1033 carboxylic group, forming a water molecule. They supported the acid/base catalytic mechanism. However, in our calculations, the negatively charged carboxylic group of the active site Asp1898 oriented and interacted with the hydroxyl group of the substrate in the favored conformation for the elimination of water instead of serving as a general acid. Our results are in agreement with the one-base dehydration mechanism suggested by Cane et al.[17] that only a conserved active site His residue acts as the base to deprotonate the substrate, and His also provides the proton for the formation of a water molecule.

The large model was used to explore the feasibility of Path2 by performing QM/MM calculations at the ONIOM (M062X/6-31G*: Amber) level. The distance of C_β–O_1 was scanned from 1.43 Å to 2.03 Å. Surprisingly, the energy barrier reached about 30 kcal mol^{-1} when the distance of C_β–O_1 was 1.83 Å, and it even increased to more than 70 kcal mol^{-1} when the distance of C_β–O_1 was 2.03 Å (Fig. S11 and S12†). It seemed that the carbenium ion and the dissociated hydroxy could not remain stable in this system. Therefore, Path2 seemed infeasible in GphF DH1. In conclusion, Path1 may be the favorable dehydration reaction pathway for GphF DH1, and the energy barrier was calculated to be 27.0 kcal
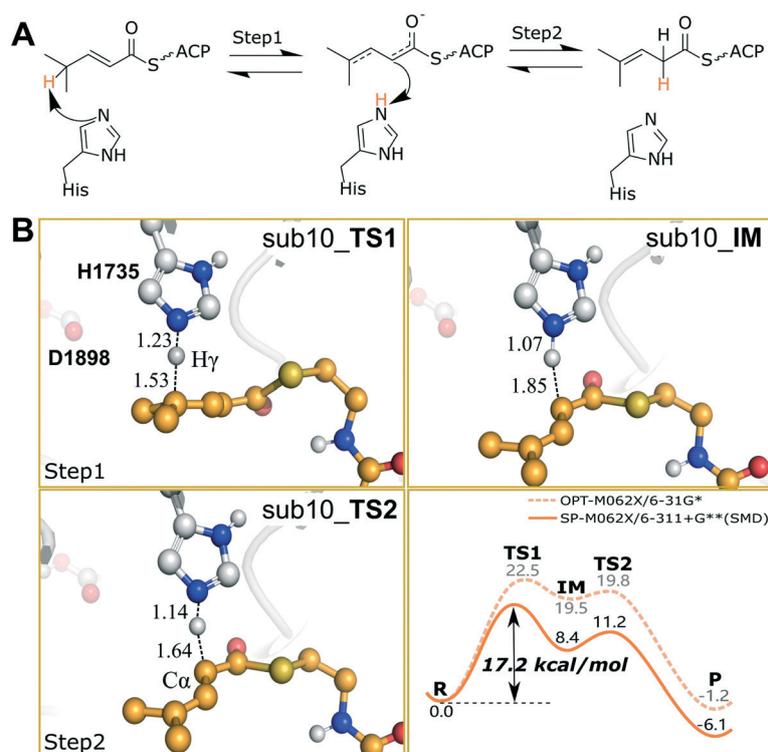
Fig. 8 (A) One possible isomerization pathway; (B) optimized structures of transition states (TS1 and TS2) and the intermediate (IM) with the energy profile calculated by small model QM/MM calculations.

This journal is © The Royal Society of Chemistry 2021

Catal. Sci. Technol., 2021, 11, 2155–2166 | 2163

mol^{-1}, when larger basis sets (6-311+G**) and SMD solvent correction were considered (Table S3†).

The mechanism of isomerization

Double bond isomerization can increase the structural diversity of polyketide compounds and unsaturated fatty acids. Besides as a dehydratase, GphF DH1 is also reported as an isomerase. However, the mechanism of isomerization remains largely elusive. In 2014, the first crystal structure of an enoyl-isomerase was obtained. The catalytic histidine was demonstrated to independently shuttle a proton between the γ- and α-positions of the intermediate.[36] Therefore, we proposed that only catalytic histidine was responsible for proton transfer between the γ- and α-positions in GphF DH1, as shown in Fig. 8A.

In the present study, the dehydration product sub10 was used to build the system. By constraining the distance of H_γ and $N_{\varepsilon H1735}$, and re-equilibrating the system, an appropriate conformation for the proposed isomerization mechanism was obtained. Optimizations were performed in the small model, and then the single point calculations with SMD solvent correction were performed at the ONIOM (M062X/6-311+G**: Amber) level.

For step 1, the elimination of H_γ to form the enol intermediate, the energy barrier was calculated to be 17.2 kcal mol^{-1}. As shown in Fig. 8B, TS1 was located at $d(N_{\varepsilon H1735}–H_\gamma) = 1.23$ Å, $d(C_\gamma–H_\gamma) = 1.53$ Å. The angle of $N_{\varepsilon H1735}–H_\gamma–C_\gamma$ enlarged to 167.2° from 133.5° in the reactant (Table S4†), and the imaginary frequency was 1117.96i cm^{-1}. After optimization, the distance of H_γ and C_α was observed to be 1.85 Å. For step 2, re-donating the proton occured easily with an energy barrier of 2.8 kcal/mol, and the energy of the product was −6.1 kcal mol^{-1}, indicating that the isomerization process was reversible. TS2 was observed at $d(N_{\varepsilon H1735}–H_\gamma) = 1.14$ Å and $d(C_\alpha–H_\gamma) = 1.64$ Å, and the imaginary frequency is 406.97i cm^{-1}.

In conclusion, the energy barrier for the whole process is 17.2 kcal mol^{-1} with the larger basis set (6-311+G**) and SMD solvent correction considered (Table S5†), and the elimination of γ-H is the rate-limiting step.

Conclusions

Employing a combined method of MD simulations and QM/MM calculations, this study explored the substrate recognition of four chiral candidates and the catalytic mechanism of the GphF DH1 domain at the atomic level. Four enzyme–substrate complex systems with sub2, sub3, sub4 and sub5, respectively, were constructed and sub2 was proposed to be the most favorable substrate in the recognition process due to its high proportion of PRS structures obtained from MD simulations. Comparing the sub2-A system with the sub2-B system, more proper conformations were observed for α–β double bond formation other than the β–γ double bond, indicating that GphF DH1 preferred to form an α–β double bond in advance. Then, the binding energies were calculated via MM-GB/PBSA

methods and thirteen key residues in substrate recognition and catalysis reaction were highlighted. Based on the PMF calculations by umbrella sampling, it was found that the non-methyl substrate formed the intra-molecular hydrogen bond more easily than the α-methyl substrate. In addition, the potential energy profiles were obtained and the critical structures along the reaction pathway were located with the ONIOM (M062X/6-31G*: Amber) method. The single point calculations were performed at the ONIOM (M062X/6-311+G**: Amber) level with SMD solvation correction. The energy barrier of the dehydration reaction in the sub9 system was calculated to be 27.0 kcal mol^{-1}, and the energy barrier of the isomerization reaction in the sub10 system was obtained to be 17.2 kcal mol^{-1}, which supported one-base dehydration and one-base isomerization mechanisms. Overall, the results elucidate the molecular mechanism of this unique dual-function dehydratase GphF DH1 and encourage future studies for the full comprehension of the catalytic mechanism of PKS DHs.

Author Contributions

T. S., Y.-L. Z. and L. B. conceived and designed the investigation. L. L., Q. Y., H. Z., W. T. and R. W. performed the calculations and analyses. T. S., Q. Y. and L. L. wrote the paper.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgements

References

1 S. L. Robinson, J. K. Christenson and L. P. Wackett, Biosynthesis and chemical diversity of beta-lactone natural products, *Nat. Prod. Rep.*, 2019, **36**(3), 458–475.

2 M. Klaus and M. Grininger, Engineering strategies for rational polyketide synthase design, *Nat. Prod. Rep.*, 2018, **35**(10), 1070–1081.

3 E. Kalkreuter and G. J. Williams, Engineering enzymatic assembly lines for the production of new antimicrobials, *Curr. Opin. Microbiol.*, 2018, **45**, 140–148.

4 G. J. Williams, Engineering polyketide synthases and nonribosomal peptide synthetases, *Curr. Opin. Struct. Biol.*, 2013, **23**(4), 603–612.

5 G. Berkhan, C. Merten, C. Holec and F. Hahn, The interplay between a multifunctional dehydratase domain and a C-methyltransferase effects olefin shift in ambruticin biosynthesis, *Angew. Chem., Int. Ed.*, 2016, **55**(43), 13589–13592.

6 K. H. Sung, G. Berkhan, T. Hollmann, L. Wagner, W. Blankenfeldt and F. Hahn, Insights into the dual Activity of a Bifunctional Dehydratase-Cyclase Domain, *Angew. Chem., Int. Ed.*, 2017, **57**(1), 343–347.

7 G. Berkhan and F. Hahn, A dehydratase domain in ambruticin biosynthesis displays additional activity as a pyran-forming cyclase, *Angew. Chem., Int. Ed.*, 2014, **53**(51), 14240–14244.

8 X. Xie, A. Garg, C. Khosla and D. E. Cane, Elucidation of the cryptic methyl group epimerase activity of dehydratase domains from modular polyketide synthases using a tandem modules epimerase assay, *J. Am. Chem. Soc.*, 2017, **139**(28), 9507–9510.

9 L. P. Partida-Martinez and C. Hertweck, Pathogenic fungus harbours endosymbiotic bacteria for toxin production, *Nature*, 2005, **437**(7060), 884–888.

10 R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh and J. Clardy, The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**(5), 1506–1509.

11 D. Gay, Y.-O. You, A. Keatinge-Clay and D. E. Cane, Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase, *Biochemistry*, 2013, **52**(49), 8916–8928.

12 D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway, *Structure*, 2010, **18**(1), 94–105.

13 A. Keatinge-Clay, Crystal structure of the erythromycin polyketide synthase dehydratase, *J. Mol. Biol.*, 2008, **384**(4), 941–953.

14 G. J. Dodge, A. Patel, K. L. Jaremko, J. A. McCammon, J. L. Smith and M. D. Burkart, Structural and dynamical rationale for fatty acid unsaturation in Escherichia coli, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**(14), 6775–6783.

15 M. Leesong, B. S. Henderson, J. R. Gillig, J. M. Schwab and J. L. Smith, Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site, *Structure*, 1996, **4**(3), 253–264.

16 F. E. Medina, R. P. P. Neves, M. J. Ramos and P. A. Fernandes, QM/MM study of the reaction mechanism of the dehydratase domain from mammalian fatty acid synthase, *ACS Catal.*, 2018, **8**(11), 10267–10278.

17 X. Xie and D. E. Cane, pH-rate profiles establish that polyketide synthase dehydratase domains utilize a single-base mechanism, *Org. Biomol. Chem.*, 2018, **16**(47), 9165–9170.

18 G. J. Dodge, D. Ronnow, R. E. Taylor and J. L. Smith, Molecular basis for olefin rearrangement in the gephyronic acid polyketide synthase, *ACS Chem. Biol.*, 2018, **13**(9), 2699–2707.

19 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, *Nucleic Acids Res.*, 2004, **32**(suppl_2), W665–W667.

20 W. D. Fiers, G. J. Dodge, D. H. Sherman, J. L. Smith and C. C. Aldrich, Vinylogous dehydration by a polyketide dehydratase domain in curacin biosynthesis, *J. Am. Chem. Soc.*, 2016, **138**(49), 16024–16036.

21 E. Anders, R. Koch and P. Freunscht, Optimization and application of lithium parameters for PM3, *J. Comput. Chem.*, 1993, **14**(11), 1301–1312.

22 H. Stoll, G. Wagenblast and H. Preuss, The Hartree-Fock theory of local regions in molecules, *J. Am. Chem. Soc.*, 1978, **100**(24), 7742–7743.

23 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 and P. A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.*, 1995, **117**(19), 5179–5197.

24 R. Salomon-Ferrer, D. A. Case and R. C. Walker, An overview of the Amber biomolecular simulation package, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**(2), 198–210.

25 T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N· log(N) method for Ewald sums in large systems, *J. Chem. Phys.*, 1993, **98**(12), 10089–10092.

26 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Chem. Phys.*, 1977, **23**(3), 327–341.

27 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method, *J. Comput. Chem.*, 2010, **13**(8), 1011–1021.

28 B. R. Miller, T. D. Mcgee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, MMPBSA.py: An efficient program for end-state free energy calculations, *J. Chem. Theory Comput.*, 2012, **8**(9), 3314–3321.

29 T. Vreven, M. J. Frisch, K. N. Kudin, H. B. Schlegel and K. Morokuma, Geometry optimization with QM/MM methods II: Explicit quadratic coupling, *Mol. Phys.*, 2006, **104**(5–7), 701–714.

30 T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. Montgomery, K. Morokuma and M. J. Frisch, Combining quantum mechanics methods with molecular mechanics methods in ONIOM, *J. Chem. Theory Comput.*, 2006, **2**(3), 815–826.

31 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P.

This journal is © The Royal Society of Chemistry 2021

Catal. Sci. Technol., 2021, **11**, 2155–2166 | **2165**

Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian, Inc., Wallingford, CT, 2009.

32 Y. Zhao and D. G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.*, 2008, **120**(1), 215–241.

33 V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern and L. A. Curtiss, 6-31G* basis set for third-row atoms, *J. Comput. Chem.*, 2001, **22**(9), 976–984.

34 A. V. Marenich, C. J. Cramer and D. G. Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions, *J. Phys. Chem. B*, 2009, **113**(18), 6378–6396.

35 T. Shi, L. Liu, W. Tao, S. Luo, S. Fan, X.-L. Wang, L. Bai and Y.-L. Zhao, Theoretical studies on the catalytic mechanism and substrate diversity for macrocyclization of pikromycin thioesterase, *ACS Catal.*, 2018, **8**(5), 4323–4332.

36 X.-P. Chen, T. Shi, X.-L. Wang, J. Wang, Q. Chen, L. Bai and Y.-L. Zhao, Theoretical studies on the mechanism of thioesterase-catalyzed macrocyclization in erythromycin biosynthesis, *ACS Catal.*, 2016, **6**(7), 4369–4378.

37 D. C. Gay, P. J. Spear and A. T. Keatingeclay, A Double-Hotdog with a new trick: Structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase, *ACS Chem. Biol.*, 2014, **9**(10), 2374–2381.

38 Y. Li, G. J. Dodge, W. D. Fiers, R. A. Fecik, J. L. Smith and C. C. Aldrich, Functional characterization of a dehydratase domain from the pikromycin polyketide synthase, *J. Am. Chem. Soc.*, 2015, **137**(22), 7003–7006.

39 X. Xie, C. Khosla and D. E. Cane, Elucidation of the stereospecificity of C-methyltransferases from trans-AT polyketide synthases, *J. Am. Chem. Soc.*, 2017, **139**(17), 6102–6105.

40 J. Young, D. C. Stevens, R. Carmichael, J. Tan, S. Rachid, C. N. Boddy, R. Müller and R. E. Taylor, Elucidation of gephyronic acid biosynthetic pathway revealed unexpected SAM-dependent methylations, *J. Nat. Prod.*, 2013, **76**(12), 2269–2276.

41 J. N. Kondo, D. Nishioka, H. Yamazaki, J. Kubota, K. Domen and T. Tatsumi, Activation energies for the reaction of ethoxy species to ethene over zeolites, *J. Phys. Chem. C*, 2010, **114**(47), 20107–20113.

42 M. R. Nimlos, S. J. Blanksby, G. B. Ellison and R. J. Evans, Enhancement of 1,2-dehydration of alcohols by alkali cations and protons: a model for dehydration of carbohydrates, *J. Anal. Appl. Pyrolysis*, 2003, **66**(1), 3–27.

2166 | *Catal. Sci. Technol.*, 2021, **11**, 2155–2166

This journal is © The Royal Society of Chemistry 2021