

Cite this: *Catal. Sci. Technol.*, 2021,  
11, 6782

## Computational studies on the substrate specificity of an acyltransferase domain from salinomycin polyketide synthase†

Huining Ji,<sup>a</sup> Ting Shi,<sup>a</sup> <sup>a</sup> Lei Liu,<sup>a</sup> Fa Zhang,<sup>a</sup> Wentao Tao,<sup>a</sup> Qing Min,<sup>c</sup> Zixin Deng,<sup>a</sup> Linquan Bai,<sup>a</sup> Yilei Zhao <sup>\*a</sup> and Jianting Zheng <sup>\*ab</sup>

Polyketides are a large group of natural products with diverse chemical structures and biological activities. They are biosynthesized by modular polyketide synthases (PKSs) from coenzyme A (CoA) thioesters of short-chain carboxylic acids like malonyl-CoA (MCoA), methylmalonyl-CoA (MMCoA) and ethylmalonyl-CoA (EMCoA). Acyltransferase (AT) domains of modular PKSs are responsible for selecting CoA thioesters and therefore attractive targets for engineering to generate novel polyketides. Herein, molecular dynamics (MD) simulations combined with quantum mechanical/molecular-mechanical (QM/MM) calculations were conducted to dissect the substrate specificity of an AT domain from the 14th module of the salinomycin modular PKS (SalAT14), which displays a preference for its cognate substrate EMCoA over MCoA and MMCoA. Comparison of MD simulations unveiled that the hydrophobic interactions between the active site residues and the acyl groups exert a significant effect on enzyme–substrate recognition. The complex of SalAT14 and its cognate substrate EMCoA exhibited a greater tendency to stay in a conformation suitable for the reaction. QM/MM calculations demonstrated that the concerted nucleophilic attack on the thioester carbonyl group of the substrate is the rate-limiting step in the first half of transacylation. Our computational investigations revealed the structural basis of AT specificity and could potentially help the engineering of modular PKSs.

Received 14th February 2021,  
Accepted 25th August 2021

DOI: 10.1039/d1cy00284h

rsc.li/catalysis

### Introduction

Polyketides possessing remarkable structural diversity and a broad range of biological activities are a large group of pharmaceutically significant natural products. Many of them

are produced by modular polyketide synthases (mPKSs), of which each module contains several domains that exhibit disparate catalytic activities.<sup>1,2</sup> A minimal module is composed of a ketosynthase (KS) which catalyses decarboxylative Claisen condensation, an acyltransferase (AT) which selects a specific acyl-CoA precursor and an acyl carrier protein (ACP) which transfers the polyketide intermediate to each catalytic domain. Optional domains such as ketoreductase (KR), enoyl reductase (ER) and dehydratase (DH) domains may exist and catalyse the stepwise redox reaction of the  $\beta$ -keto group to control the oxidation state of the polyketide intermediate.

The AT domain serves as the essential doorman of polyketide biosynthesis. It is proposed that the transacylation performed by AT occurs through a two-stage ping-pong catalytic mechanism and a Ser–His catalytic dyad plays a pivotal role.<sup>3</sup> In the first stage, the catalytic Ser is deprotonated by the conserved His and attacks the carbonyl carbon of the substrate, resulting in the transfer of the substrate to the AT. In the second stage, the thiol group of the phosphopantetheinyl arm of the ACP domain attacks the enzyme-bound intermediate (Scheme 1).

The AT domain within each module of mPKSs recruits an extension unit required for chain extension.<sup>4</sup> Accordingly, the

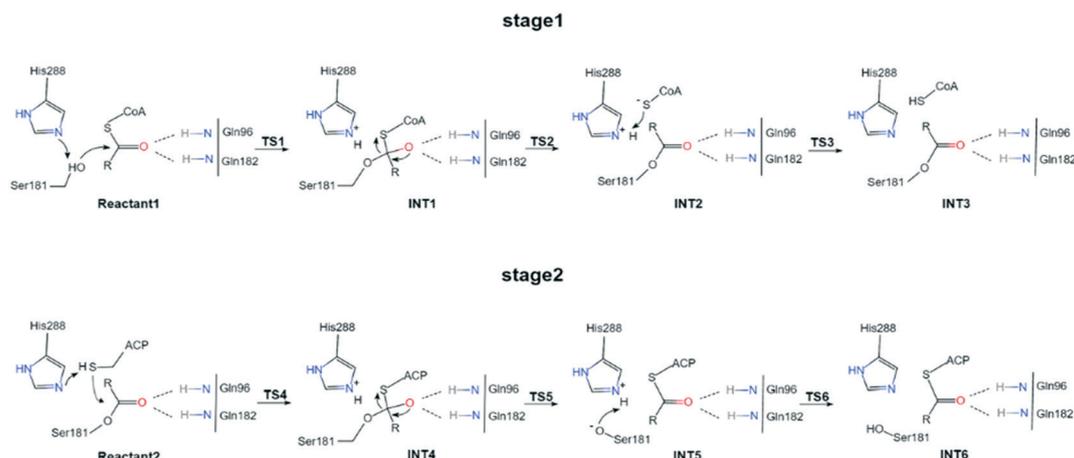
<sup>a</sup> State Key Laboratory of Microbial Metabolism, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China.

E-mail: yileizhao@sjtu.edu.cn, jtzheng@sjtu.edu.cn

<sup>b</sup> Joint International Research Laboratory of Metabolic & Developmental Sciences, Shanghai Jiao Tong University, Shanghai, China

<sup>c</sup> Pharmacy School, Hubei University of Science and Technology, HubeiXianning 437100, China

† Electronic supplementary information (ESI) available: The atom names of the three substrates; the hydrogen-bond network in the SalAT14–EMCOA system; PESs in the SalAT14–EMCOA system; indirect effects of key structures in the SalAT14–EMCOA system; critical angle in TS2 structure; RMSD of three 300 ns MD simulations; specific phenomenon in the SalAT14–MMCOA system; the overlap of SalAT14–EMCOA and SalAT14–MMCOA systems; hydrophobic interactions between MMCOA/MCOA and SalAT14 mutants; hydrophobic pockets of MMCOA with SalAT14 and its mutant; the distribution of key distances of MMCOA with the SalAT14 mutant; the structural overlaps in MCOA–SalAT14 mutants systems; hydrogen bonds between the ligands and the proteins of 300 ns simulations; energetic corrections to key structures and energy barriers in the SalAT14–EMCOA system with different basis sets (PDF). See DOI: 10.1039/d1cy00284h



**Scheme 1** Schematic representation of the proposed catalytic mechanism of SalAT14 (R = malonyl, methylmalonyl or ethylmalonyl moiety). The residues involved in the reactions are Ser181, His288, Gln96 and Gln182 in this work.

alteration of the AT substrate scope would introduce a diversified feature into the polyketide structures to produce more analogues of bioactive molecules. However, AT domains usually display strict substrate selectivity during the assembly of polyketide chains. In nature, limited building blocks are employed, in most circumstances malonyl-coenzyme A (MCoA) and methylmalonyl-CoA (MMCoA), and sometimes ethylmalonyl-CoA (EMCoA). It's challenging to shift the native substrate profile of AT, because the molecular basis for this strict specificity has not been well understood.<sup>5,6</sup>

Traditionally, AT engineering has been centred on transplanting the entire AT domains between different mPKSs to vary the final product structure.<sup>7</sup> Recently, Satoshi Yuzawa and colleagues have identified the highly conserved domain boundary to exchange the AT domains, which could contribute to sustaining enzyme activity and protein stability.<sup>8</sup> Nevertheless, even with the updated AT boundaries, the potency of these hybrid PKSs is still uncontrollable, which has often resulted in greatly reduced product titers or inactive enzymes.<sup>9</sup> The poor catalytic efficiencies compared with their wild-type counterparts may result from the impaired protein folding and non-native protein-protein interactions.<sup>10</sup> A reasonable alternative is converting the conserved motif such as YASH which is common in methylmalonyl-only ATs to the motifs corresponding to other substrates.<sup>11</sup> However, it is disappointing that this method may also lead to enzyme inactivation or lower yields of the desired products.<sup>12,13</sup> An increasingly popular approach to inverting AT substrate specificity and maintaining the catalytic activity involves introducing several mutations in or around the active site.<sup>11,12,14</sup> This strategy usually relies on high-throughput screening of mutations from active site saturation libraries,<sup>15</sup> but the reasons for the decrease in production of altered polyketides still remain unknown. Moreover, it is a sensible approach to utilize the structural information of AT-substrate complexes as the guidance for mutagenesis in the active site.<sup>12,16</sup> This strategy can often become better when combined with computational

simulations, which allow people to explore the mechanism of substrate specificity and to predict the incorporation of unusual extender units.<sup>14,17</sup> In fact, an *in silico* based engineering method that allows the identification of favourable positions and a virtual screening of possible variants has been successfully applied in the expansion of the substrate scope of amine transaminase.<sup>18</sup> It is worth remarking that the calculated prediction for the incorporation of unusual substrates is of great significance, not only to understand the fundamental substrate selectivity but also as a guide to experiments.

Salinomycin is a significant polyether polyketide with antibacterial and anticoccidial activities.<sup>19</sup> Six MCOAs, six MMCOAs, and three EMCOAs are necessary for the formation of salinomycin.<sup>20</sup> The AT domain from the 14th module of the salinomycin mPKS (SalAT14) utilizes EMCOA as the extension unit, while SalAT2 and SalAT8 are specific for MCoA and MMCoA, respectively. We have solved the crystal structures of these AT domains and successfully switched the substrate specificity of SalAT14.<sup>12</sup> However, the mechanism of the substrate recognition of AT domains remains to be elucidated. Herein, our computational study uncovered the fundamental reaction pathway for SalAT14 in the first half of transacylation *via* utilizing molecular dynamics (MD) simulations as well as quantum mechanics/molecular mechanics (QM/MM) methods. Based on these, it is found that the appropriate positioning adopted by the substrates and the maintenance of critical interactions within the active site are vital to substrate recognition. The knowledge gathered in this work would instigate further explorations of AT engineering.

## Materials and methods

### MD simulations

The X-ray crystal structure of SalAT14 (PDB entry: 6IYT)<sup>12</sup> was utilized as the initial structure. AutoDock software was used to provide the structures of AT-acyl-CoA complexes.<sup>21</sup> SalAT14 mutants with switched specificity for MMCoA (V285Y/F210V/

V220M triple mutant) and MCoA (V285H/S287F/Q182I/F210M quadruple mutant) have been generated in our previous report.<sup>12</sup> The structures of the mutant enzymes were generated by employing mutations on the structure of wild type SalAT14. The mutated residues adopted conformations according to the Dunbrack library.<sup>22</sup> The H++ web server was utilized to assign the protonation state of residues of the SalAT14 domain at pH 7.0 followed by visual inspection.<sup>23</sup> We performed a conformational optimization for acyl-CoA extension units with the Gaussian 16 program at the level of HF/6-31G(d) and calculated their electrostatic surface potential charge.<sup>24</sup> A restrained electrostatic potential (RESP)<sup>25</sup> fitting method was employed to fit on the substrate by using the antechamber program. The parmchk2 program generated the missing parameters such as the dihedral angle and bond information. A cubic box of TIP3P water molecules was used to solvate the complexes with a water thickness of the external water layer exceeding 10 Å. The neutral simulation systems were created by adding sodium ions.

MD simulations were carried out under the AMBER ff03.r1 force field and using the PMEMD.cuda in AMBER 18 suite.<sup>26</sup> The top structures of each complex from AutoDock were similar and two of them were used for MD simulations. Long-range electrostatic interactions were accounted by the Particle Mesh Ewald (PME) method.<sup>27</sup> And the SHAKE algorithm was utilized to fix the angles as well as bonds involving hydrogen atoms.<sup>28</sup> A cutoff of 10 Å was set for van der Waals interactions. Minimization and heating were performed to the solvated systems, with the system temperature gradually being increased from 0 to 300 K over 50 ps. Next, the systems were transformed to constant pressure and temperature (NPT) and subjected to a 50 ps equilibration. Finally, 50 ns production simulations in the absence of any restraint were conducted under NPT conditions. Each simulation was repeated six times with a different random number for adequate sampling, and one of them was extended to 300 ns to evaluate the stability of the critical distances.

The C $\alpha$  atoms of the complex were used to calculate the root mean-square deviation (RMSD) of the systems. The widely accepted average linkage clustering algorithms were utilized to carry out RMSD-based clustering and a representative outcome of the analysis is summarized in Table S5.† Characteristic structures were abstracted to represent the change of the substrate conformations. CPPTRAJ<sup>29</sup> was used to conduct hydrogen-bonding analysis (the acceptor–donor–hydrogen angle is >135° and the donor–acceptor distance is <3.0 Å) of the MD simulation trajectories and VMD was employed to carry out visual inspection.<sup>30</sup> The hydrophobic interaction between two hydrophobic groups was defined when the mass-center distance  $\leq 6.5$  Å. Additionally, the solvent-accessible surface area (SASA) between the side-chains of the hydrophobic residues (Ile149, Phe210, Val220 and Val285) and the acyl group of the substrate was also calculated to quantitatively compare the hydrophobic interactions.

## QM/MM calculations

The initial structure of the QM/MM calculations was the representative snapshot of the dominant cluster obtained from MD simulations. All of the QM calculations were conducted with the M06-2x/6-31G(d) method. The MM region was under the AMBER Parm99 force field. And QM/MM calculations which were at the ONIOM (M06-2x/6-31G(d):Amber) level were conducted by utilizing a two-layered ONIOM method<sup>31,32</sup> in the Gaussian16 program.

The QM (high level) region consisted of the atoms that participated in the acetyl transfer reaction including the full Ser181 and Gln182, the side chain of Gln153, Arg206, Ser287 and His288, a portion of the backbone of Gln96 and Gly95, the acyl moiety, the sulphur atom and the two adjacent CH<sub>2</sub> groups of the substrate according to a previous study.<sup>3</sup> Additionally, H<sub>2</sub>O583 and H<sub>2</sub>O677 which were crystalline water molecules and played a significant role in stabilizing the substrate were retained in the QM layer. A total of 115 atoms were included in the QM region of the EMCOA system (112 atoms in the system of MMCOA and 109 atoms in the system of MCOA), while the MM (low level) region accounted for the remaining 12 281 atoms (12 347 atoms in the system of MMCOA and 12 210 atoms in the system of MCOA). Besides, the electronic embedding method that allows the QM wave function to be polarized by the MM region with scaled partial atomic charges of MM atoms was employed to calculate the electrostatic interactions between the QM and MM regions. Single-point energy calculations for SalAT14–EMCOA were implemented on the basis of the optimized structures by the usage of larger basis sets, including 6-311+G(d), 6-311+G(d,p), and 6-311+G(2df,2p) (Table S4†).

## Results and discussion

To decipher the catalytic mechanism and the substrate specificity of SalAT14, three complex systems (SalAT14–EMCOA, SalAT14–MMCOA and SalAT14–MCOA) were constructed. Six molecular dynamic (MD) simulations of 50 ns were carried out with one of them being extended to 300 ns. Following MD simulations, QM/MM calculations were performed to explore the transfer of the acyl moiety from CoA to SalAT14 and to obtain the energy barriers. The CoA substrates were used in all MD simulations, whereas only the acyl groups are shown in the figures for clarity. The findings for the SalAT14–EMCOA were presented in detail and the results obtained with the three substrates were compared. We also analysed the SalAT14 mutants to expand our understanding of the substrate specificity.

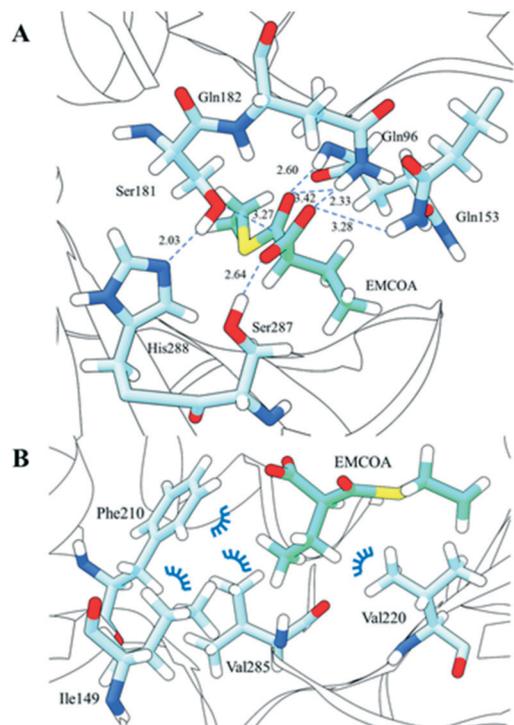
### SalAT14–EMCOA binding mode

The MD simulations reveal that the natural substrate (EMCOA) is oriented in a conformation suitable for reaction in the SalAT14 active site. The crucial residues of SalAT14 include the catalytic dyad (His288 and Ser181), the residues involved in an oxyanion hole (Gln182 and Gln96), and the

residues in contact with the substrate carboxylate anions (Gln153 and Ser287). According to previous studies,<sup>3,33</sup> the catalytic His first deprotonates the catalytic Ser, and the resulting Ser-O<sup>-</sup> with enhanced nucleophilicity attacks the carbonyl carbon of the substrate. In all simulations, the hydroxyl H<sub>γ</sub> atom on the Ser181 side chain of SalAT14 is oriented towards the N<sub>ε</sub> atom of His288 with an average distance of 2.31 ± 0.58 Å. Simultaneously, the Ser181 O<sub>γ</sub> atom is close to the thioester carbon of the substrate with an average distance of 3.61 ± 0.82 Å and ready for the nucleophilic attack. Table S1† summarizes the hydrogen-bonding network within the active sites in the 300 ns MD simulations (the first 10 ns simulations are excluded in the analysis and will be explained hereinafter). The thioester carbonyl oxygen of the substrate stays in the oxyanion hole formed by the side chain amide of Gln182 and the backbone amine of Gln96 (Fig. 1A). The carboxylate anions of EMCOA form strong ionic interactions with the side hydroxyl of Ser287, and the side chain amine of Gln153 and Gln182 (Fig. 1A). These interactions obviously help to position and stabilize the carboxylate of the substrate. Thus, the Gln182 residue played two roles: it promotes the nucleophilic attack of the substrate and assists in maintaining EMCoA in a

position close to Ser181. The experimental data demonstrate<sup>12</sup> that their mutations would facilitate altering the specificity of SalAT14. The Arg206 residue has been proposed to play a significant role in stabilizing the substrates.<sup>3,17,34,35</sup> However, it doesn't directly interact with the substrate in the simulations, but exerts an indirect effect by stabilizing H<sub>2</sub>O583 and H<sub>2</sub>O677 that form hydrogen bonds with the substrates (Fig. S2 and Table S1†). The hydrogen-bond network of these water molecules bound by Gln153, Gln182 and Arg206 likely stabilizes the substrate and aids in creating a polar environment in the active site that may favour substrate binding.

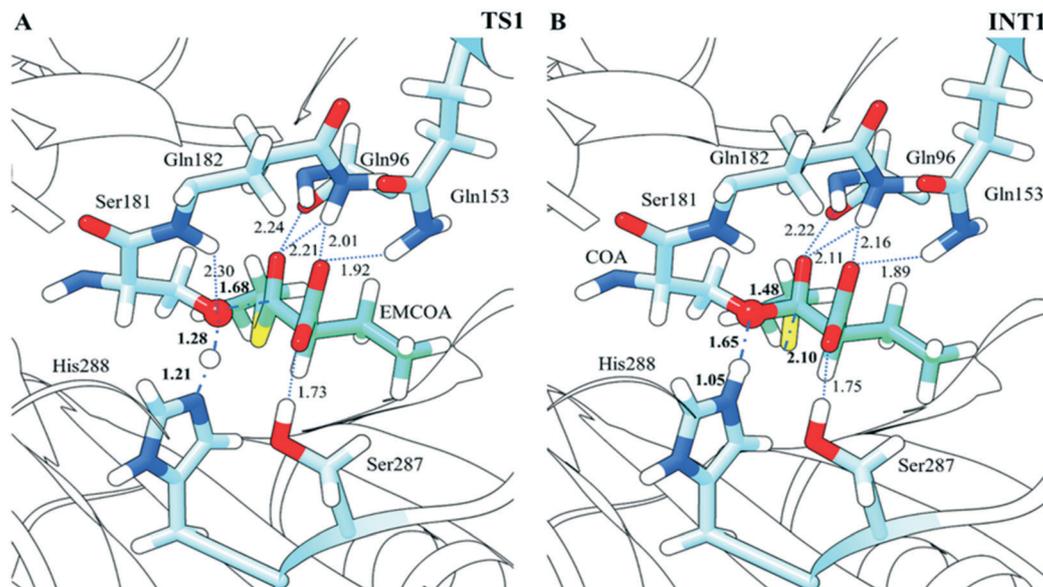
As shown in Fig. 1B, Ile149, Phe210, Val220 and Val285 form a hydrophobic environment to accommodate the ethyl group of EMCOA. These residues define an important hydrophobic pocket accommodating the substrate and control the substrate specificity. Mutation analysis of SalAT14 has verified the involvement of these residue in selecting the substrate.<sup>12</sup> These essential interactions assist in properly positioning EMCOA for reaction in the SalAT14 active site. Thus, the MD results support the hypothesis that His288 could form close contact with Ser181 to facilitate the proton transfer from the latter residue and the occurrence of nucleophilic attack.



**Fig. 1** Representative structure for SalAT14-EMCOA in the reactant state: (A) key interactions involved with the catalytic Ser181 and His288, the oxyanion hole (Gln182 and Gln96) interacting with EMCOA and the residues (Gln153 and Ser287) which formed strong ionic interactions with the substrate; (B) the hydrophobic pocket with EMCOA, the dark blue “eyelashes” represent the hydrophobic interaction. The substrate EMCOA is cyan and the residues are blue. EMCOA only shows the acyl moiety plus the sulfur atom and the two adjacent CH<sub>2</sub> groups. The distances are given in Å.

#### QM/MM results for SalAT14-EMCOA, step 1: nucleophilic attack on the thioester carbonyl group

Starting from the SalAT14-EMCOA binding structure as the initial model, the combined QM/MM (M06-2x/6-31G(d); Amber) calculations were carried out to simulate the deprotonation of the catalytic Ser181 and the nucleophilic attack by the hydroxyl oxygen (O<sub>γ</sub>) on the thioester carbon of EMCOA.<sup>3</sup> The deprotonation of Ser181 enhances its nucleophilicity and promotes the attack (Scheme 1). The structural descriptions of the first transition state (TS1) and the first intermediate (INT1) of the SalAT14-EMCOA complex are displayed in Fig. 2. This process involves the formation of a new covalent bond between the hydroxyl oxygen (O<sub>γ</sub>) of Ser181 and the thioester carbon of EMCOA through a concerted step. The dipole-dipole H<sub>γ</sub>-N<sub>ε2</sub> interaction between Ser181 and His288 (1.21 Å) at the TS1 state is changed to an ion-dipole H<sub>γ</sub>-O<sub>γ</sub> interaction (1.05 Å) in the INT1 structure. The transition structure (TS1) is characterized by a relatively short C-O bond length (1.68 Å), indicating a late transition state close to the tetrahedral intermediate (INT1) (Fig. S3†). Meanwhile, the thioester carbonyl oxygen of EMCOA is confined in the oxyanion hole constituted by the side chain amide of Gln182 and the backbone amine of Gln96 (Fig. 2). The oxyanion hole plays a pivotal role in expelling the electron concentration from the thioester carbon to the carbonyl oxygen, facilitating the formation of the scissile bond highly prone to suffer from nucleophilic attack. Previous work made by Pedro and Sousa<sup>3</sup> has indicated that the oxyanion holes could lower the activation energy of step 1 which is crucial for the steadiness of the carbonyl group of



**Fig. 2** Optimized structures of TS1 and INT1 (A and B) obtained with SalAT14 complexed with EMCOA. As the O<sub>γ</sub> of Ser181 attacks EMCOA, the H<sub>γ</sub> moves toward the N<sub>ε2</sub> of His288 in a concerted, but asynchronous pathway. The substrate EMOA is cyan and the residues are blue. The distances are given in Å. See also Fig. S3<sup>†</sup>

the substrate when it passes on to the catalytic Ser. Hence, these two residues play significant roles in the formation of TS1 and INT1. Interestingly, in the process of the deprotonation of Ser181, the adjacent Gln182 constitutes a hydrogen bond with the O<sub>γ</sub> atom of Ser181 (2.30 Å). This interaction displays an additional stabilization on Ser181, but has an adverse effect on Ser181 to execute the nucleophilic attack on EMCOA.

Furthermore, the strong ionic interactions of the carboxylate anions of the substrate deserve attention. During the first step (Fig. 2), the Ser287 hydroxyl hydrogen, the Gln153 amine and the Gln182 amine form strong ionic interaction with the carboxylate anion of ethylmalonyl-CoA. According to the previous study of MAT,<sup>3</sup> SpnD-AT,<sup>16</sup> AsmAT3 (ref. 36) and so on, the conserved Arg serves as a crucial residue for stabilizing the carboxylate anion of the substrate. As mentioned above, the corresponding Arg206 has an indirect effect on EMCOA through H<sub>2</sub>O583 and H<sub>2</sub>O677 (Fig. S4<sup>†</sup>). Hence, during step1, Gln153, Gln182, Arg206, Ser287, H<sub>2</sub>O583 and H<sub>2</sub>O677 seem to play a pivotal role in positioning the carboxylate of the substrate and maintaining the steadiness of the TS1 as well as INT1 states *via* interactions with the ethylmalonyl moiety when it transfers from CoA to SalAT14.

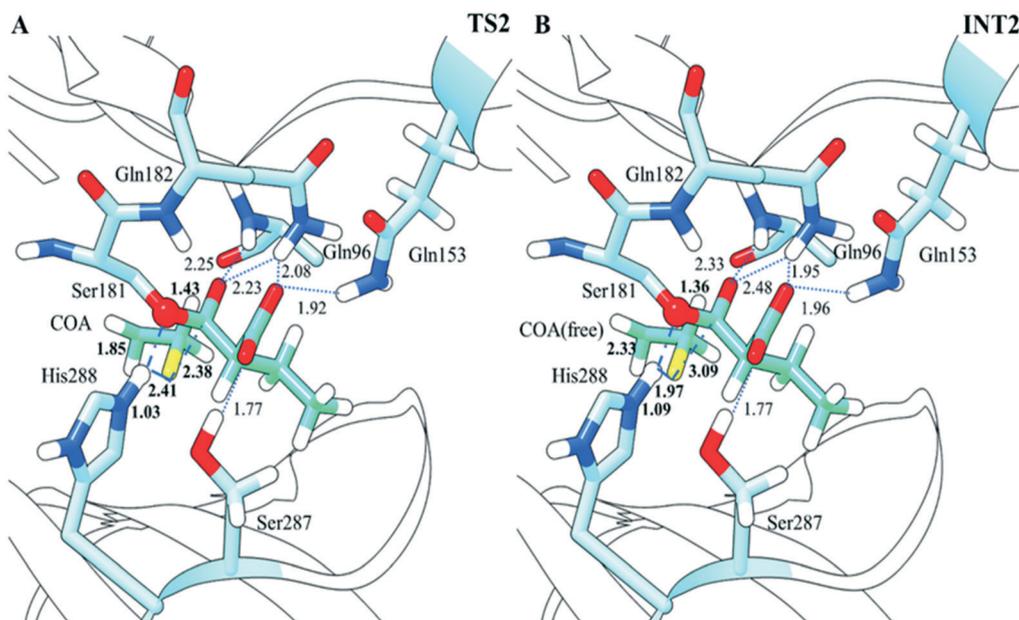
### QM/MM results for SalAT14-EMCOA, step 2: tetrahedral intermediate breakdown & step 3: formation of intermediate 3

Following step 1, the distance between the O<sub>γ</sub> of Ser181 and the carbonyl of the substrate becomes shorter (1.43 Å in TS2 structure and 1.36 Å in INT2 structure), with the interaction involving them being more powerful. The sulphur leaves the carbonyl carbon of the substrate, which leads to the breakdown

of the tetrahedral intermediate (Fig. 3). Additionally, the imidazole ring of His288 suffers a rotation and the distance between the H<sub>γ</sub> and the sulphur shortens (2.51 Å in TS2 structure and 1.97 Å in INT2 structure). The angle of S-C-O in TS2 is 105.44° (Fig. S5<sup>†</sup>), close to 107° as preferred in the trajectory of carbonyl additions studied by B:urgi and Dunitz.<sup>37</sup> Obviously, the rotation of His288 enables the H<sub>γ</sub> to approach the sulphur. In a previous study,<sup>3</sup> the structural difference between INT1 and INT2 was too huge to be ignored and the step2 found in our work has never been reported before.

At TS2 and INT2 states, the carbonyl oxygen continues to be steady because of the side chain amine of Gln182 (2.14 Å in TS2 structure and 2.48 Å in INT2 structure) and the backbone amine of Gln96 (2.23 Å in TS2 structure and 2.33 Å in INT2 structure). As previously mentioned, Gln153, Gln182, Ser287, H<sub>2</sub>O583 and H<sub>2</sub>O677 remain responsible for the positioning and stabilization of the ethylmalonyl moiety (Fig. 3 and S4<sup>†</sup>). Above all, these vital interactions remain throughout the TS2 and INT2 states.

After the formation of INT2, protonated His288 undergoes a deprotonation caused by the negative S atom of the free CoA group (Fig. 4). As the S-H<sub>γ</sub> distance shortens (1.76 Å), the ion-dipole interaction formed in INT1 and INT2 breaks at the TS3 state, with His288 adopting a new position. Finally, the H<sub>γ</sub> forms a bond with S (1.35 Å) in the INT3 structure. Besides, compared with the TS3 state, the H<sub>γ</sub> is closer to the O<sub>γ</sub> of Ser181 (Fig. 4). The oxyanion hole and strong ionic interactions involving the carboxylate of the substrate remain stable throughout TS3 and INT3 (Fig. 4 and S4<sup>†</sup>). Once step 3 finished, the protonated CoA is about to move away from the active site pocket and the phosphopantetheine group of ACP comes on stage, triggering the second stage of the proposed SalAT14 catalytic reaction.



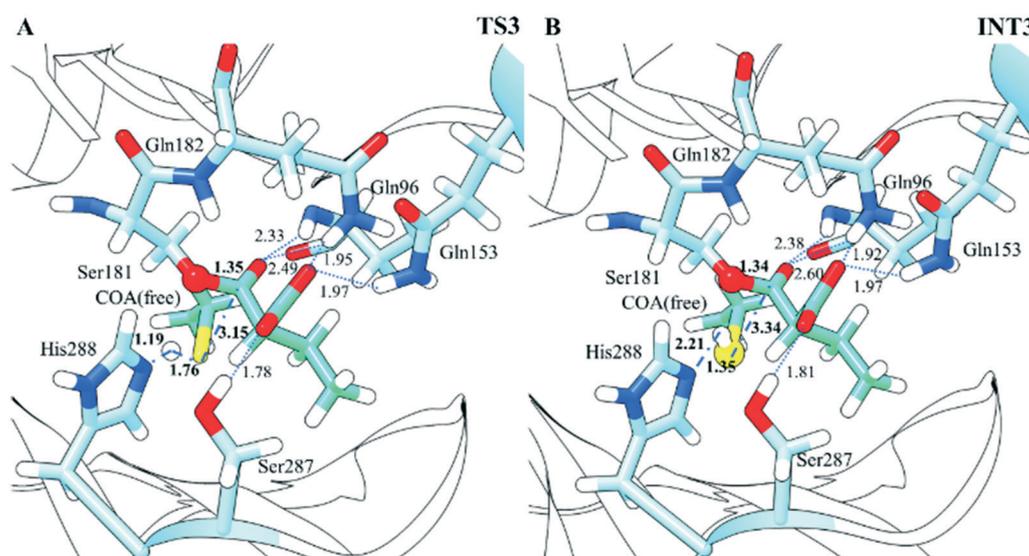
**Fig. 3** Optimized structures of TS2 and INT2 (A and B) obtained with SalAT14 complexed with EMCOA. After the nucleophilic attack, the negative thiol group of the CoA moves toward the H<sub>2</sub>, attracted by His288 and breaks down from the tetrahedral intermediate. The substrate EMOA is cyan and the residues are blue. The distances are given in Å. See also Fig. S3 and S5.†

Notably, as depicted in Fig. 5 (see also Fig. S3†), our calculations revealed that the energy barrier of step1 was apparently higher than that of step2 and step3 in SalAT14-EMCOA, which suggested that step 1 was the rate-limiting step in the first half of transacylation.

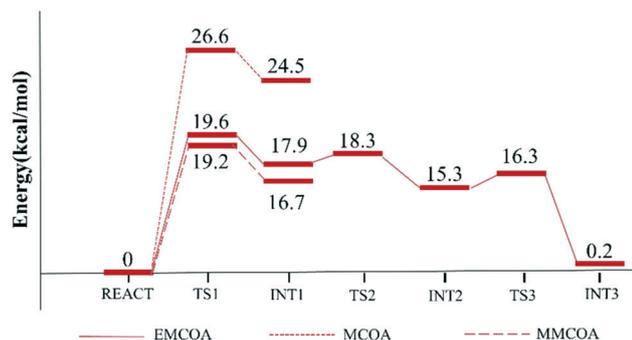
### Comparison of substrate specificity

To better probe the substrate specificity of SalAT14, we also conducted QM/MM calculations to obtain the energy barriers

for MMCoA and MCoA. Since the above calculations suggested that the concerted step (step 1) was the rate-limiting step in the first half of transacylation, we only calculated the energy barriers of the first step for SalAT14-MMCoA and SalAT14-MCoA. As shown in Fig. 5, compared with SalAT14-EMCoA, the energy barrier of SalAT14-MCoA is significantly higher, but that of SalAT14-MMCoA is slightly lower. In fact, as observed experimentally,<sup>12</sup> SalAT14 shows weak catalytic activities towards MMCoA and MCoA due to their large values of  $K_m$ . The little difference in activation



**Fig. 4** Optimized structures of TS3 and INT3 (A and B) obtained with SalAT14 complexed with EMCOA. The negative thiol group of the CoA successfully deprotonates the N<sub>2</sub> of His288 and INT3 takes shape. The substrate EMOA is cyan and the residues are blue. The distances are given in Å. See also Fig. S3.†



**Fig. 5** Energy profiles for the acyl transfer reaction from CoA to SalAT14 of the three substrates at the M06-2x/6-31G(d):AMBER level calculated with Gaussian 16 rev B.01. ONIOM QM/MM has been extensively utilized in the enzymatic studies and the critical atoms of the active site are treated with density functional (DFT) theory, while the remaining part is at the molecular mechanics level. See also Fig. S3†

barriers between MMCoA and EMCoA suggests that other factors affect the reactivity and substrate specificity of SalAT14. For example, the process of the substrate transfer from AT to ACP may also play a significant role in specificity and needs to be explored in the future. Herein, we not only compared the substrate specificity in the thermodynamic aspect, but probed the substrate binding dynamically.

The three substrates (EMCoA, MMCoA, MCoA) are very similar and the only difference among them are the side chains of the acyl groups. To investigate their binding modes, we first calculated the RMSD values of the binding pocket residues, and the acyl and the pantetheine of the substrate that reside in the pocket. The adenosine nucleotide of the substrate is located outside the protein, exposed to the surrounding solvent, and therefore not included in the analysis. As shown in Fig. S6A,† after nearly 10 ns of simulations, the binding pockets (contain substrates) have reached equilibria. So, the following analysis would exclude the first 10 ns simulations. During their MD trajectories, the pantetheines of the three substrates show a stable and similar binding mode and interact with the Gln96, Ser222, Asn250, Arg277 and Arg391 residues (Fig. S6B†). Thus, our analysis focused on the acyl groups of the substrates.

The acyl group binding poses of SalAT14-EMCoA, SalAT14-MMCoA and SalAT14-MCoA are significantly different throughout the MD trajectories. We utilize two distances to evaluate the reliability of the catalytically competent state in SalAT14-EMCoA, SalAT14-MMCoA and SalAT14-MCoA. One is the distance between the  $N_{\epsilon}$  atom of His288 and the  $H_{\gamma}$  atom of Ser181, demonstrating the deprotonation of the catalytic Ser181. The other distance is defined between the  $O_{\gamma}$  atom of Ser181 and the thioester carbon of the substrate, indicating the nucleophilic attack. It has been shown by Pedro and Sousa<sup>3</sup> that the catalytically competent state adopts an “active” conformer, where the former distance is inferior to 3 Å and the latter is under 3.5 Å. Hence, structures with restraints of both  $d(N_{\epsilon}\text{-HG}) \leq 3$  Å and  $d(\text{CS1-OG})/d(\text{CM1-OG}) \leq 3.5$  Å will be a prerequisite of

the “active” conformation. Therefore, the two distances are monitored in the MD simulations and used as criteria to evaluate the formation of the catalytically productive binding poses in the three complexes. As shown in Fig. 6, it's clear that the SalAT14-EMCoA complex contains the most catalytically productive structures. Undoubtedly, it's difficult for SalAT14-MMCoA and SalAT14-MCoA to form a structure favourable to the reaction. The average  $d(N_{\epsilon}\text{-HG})$  distance in all MD simulations is  $2.31 \pm 0.58$  Å for SalAT14-EMCoA, and less than  $3.30 \pm 0.87$  Å and  $2.79 \pm 0.79$  Å in SalAT14-MMCoA and SalAT14-MCoA, respectively. It should be noted that the  $H_{\gamma}$  atom of Ser181 would interact with the carboxyl group of MMCoA and the average distance is  $2.57 \pm 0.91$  Å, explaining the long  $d(N_{\epsilon}\text{-HG})$  distance in SalAT14-MMCoA (Table S1 and Fig. S7†). This phenomenon is dominant in SalAT14-MMCoA, accompanied by the  $N_{\epsilon}$  atom of His288 forming a hydrogen bond with Ser287 (Table S1 and Fig. S7†). The average  $d(\text{CM1-OG})$  distance in SalAT14-MCoA is  $3.96 \pm 0.59$  Å, larger than  $3.61 \pm 0.82$  Å in SalAT14-EMCoA but less than  $4.40 \pm 0.44$  Å in SalAT14-MMCoA. In both SalAT14-MMCoA and SalAT14-MCoA, the long  $d(\text{CS1-OG})/d(\text{CM1-OG})$  distance has an adverse effect on the formation of the “active” state. As mentioned above, in the catalytically competent state, the  $d(N_{\epsilon}\text{-HG})$  distance and the  $d(\text{CS1-OG})/d(\text{CM1-OG})$  distance should be close enough simultaneously, so that the deprotonation of the catalytic Ser181 and the nucleophilic attack on the substrate could happen concertedly. The population of both  $d(N_{\epsilon}\text{-HG}) \leq 3.0$  Å and  $d(\text{CS1-OG})/d(\text{CM1-OG}) \leq 3.5$  Å in SalAT14-EMCoA is about 60.77%, conspicuously larger than 5.33% and 2.14% in SalAT14-MMCoA and SalAT14-MCoA, respectively. In conclusion, compared with SalAT14-MMCoA and SalAT14-MCoA, SalAT14-EMCoA is more apt to form a catalytically competent state, which would also sustain a relatively longer time.

The acyl group binding poses are different according to the two distances ( $d(N_{\epsilon}\text{-HG})$  and  $d(\text{CS1-OG})/d(\text{CM1-OG})$ ) that are important for forming the “active” conformation. The only difference among the three substrates is the length of the side chain, which is the ethyl group for EMCoA and the methyl group for MMCoA (MCoA lacks the side chain). This discovery prompted us to dissect the strict substrate specificity of SalAT14. Hence, we analysed the hydrophobic and the hydrogen-bonding interactions between SalAT14 and the three substrates. The stabilizing effects of the crucial oxyanion hole on the substrates are observed in all three complexes, but the existence time is quite different (Table S1 and Fig. S1†). We have observed that the backbone amine of Gln96 also interacts with the carboxyl group of the substrate in SalAT14-MMCoA (47.15%). However, this interaction is rarely observed in SalAT14-EMCoA. The hydrogen bonds stabilizing the carboxylate anions of the substrate could play a significant role in substrate recognition<sup>3</sup> and exist in the three complexes, but the hydrogen bond donors in the three systems are slightly disparate (Table S1†). The existence time of key hydrogen bonds in SalAT14-EMCoA is the longest in

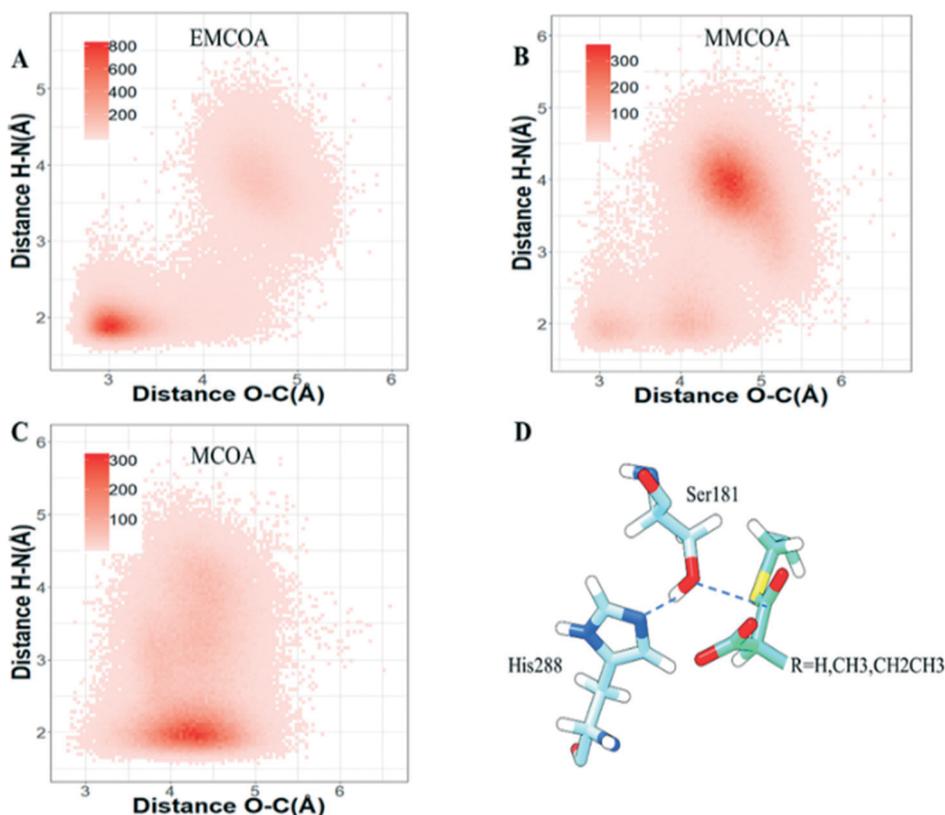


Fig. 6 Conformer populations with  $d(N_e-HG)$  and  $d(CS1-OG)/d(CM1-OG)$  distances obtained from 300 ns MD simulations in SalAT14-EMCOA, SalAT14-MMCOA and SalAT14-MCOA (A-C); (D) schematic representation of the  $d(N_e-HG)$  and  $d(CS1-OG)/d(CM1-OG)$  distances.

the three systems. According to the simulations, Ile149, Phe210, Val220 and Val285 form hydrophobic interactions with the substrates, but the strength in the three systems is quite different (Fig. 7). To better compare the hydrophobic

interactions in the three complexes quantitatively, the solvent-accessible surface area (SASA) between the hydrophobic side-chains and the acyl group of the substrate was calculated. During the 300 ns MD simulations, the

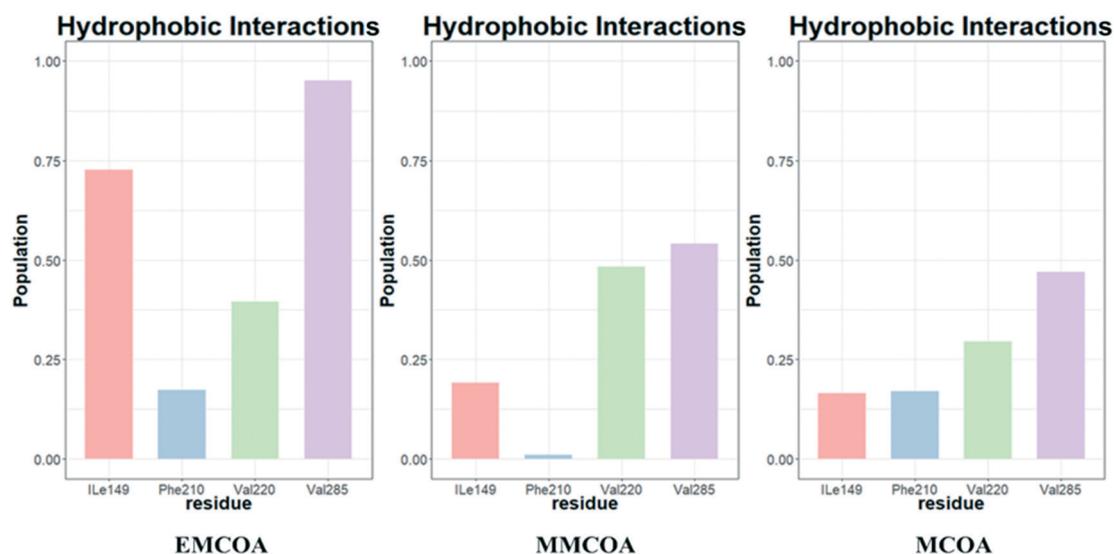


Fig. 7 Hydrophobic interactions between SalAT14 and substrates (EMCOA, MMCOA and MCOA). The SalAT14-EMCOA system has the most effects.

average SASA in SalAT14–EMCoA is 35.32 Å<sup>2</sup>, larger than that in SalAT14–MMCoA (21.55 Å<sup>2</sup>), but smaller than that in SalAT14–MCOA (48.89 Å<sup>2</sup>). Lacking the side chain, the malonyl group shows an extremely unstable binding pose during the simulations, which may account for the low activity of SalAT14 for MCoA.

### MD simulations on the SalAT14 mutants

To advance our understanding of the substrate specificity of SalAT14 and the formation of the “active” conformation, we analysed the SalAT14 mutants that have been engineered to show altered specificity for MMCoA and MCoA in our previous research.<sup>12</sup> The side chain of MMCOA which is methyl instead of ethyl is located in the hydrophobic pocket just like that of EMCOA, but the position of the carboxyl group is very different (Fig. S8†), leading to a weaker hydrogen-bonding network and longer  $d(N_{\epsilon}-HG)$  &  $d(CS1-OG)$  distances. Our previous *in vitro* enzymatic assays suggest that the SalAT14 (V285Y/F210V/V220M) triple mutant shows switched substrate specificity for MMCOA.<sup>12</sup> We carried out MD simulations for the triple mutant. The results demonstrated that the strength of the hydrophobic interactions between the triple mutant and MMCOA is enhanced (Fig. S9†). The SASA between the MM group and the surrounding residues is increased from 36.16 Å<sup>2</sup> in the wild-type enzyme to 53.83 Å<sup>2</sup> in the triple mutant. The residues involved in hydrophobic interactions change from Val220 and Val285 to Ile149 and Val210 (Fig. S9†). We speculate that the combined mutation of V285Y/V220M assists in altering the hydrophobic interactions and F210V may reduce the steric hindrance for the approximation of the substrate. Consequently, the location of the side chain of MMCOA has moved forward (Fig. S10†), strengthening the hydrogen-bonding network (Table S2†) and promoting the formation of the “active” state (Fig. S11†).

Due to the lack of a side chain, malonyl is very flexible in the active site pocket. The MD simulations for SalAT14–MCOA showed very inconsistent trends, which were in accord with its flexibility in the binding mode. The SalAT14 (V285H/S287F/Q182I/F210M) quadruple mutant shows altered specificity for MCOA in our previous enzymatic assays.<sup>12</sup> We conducted MD simulations for the quadruple mutant and the triple mutants including SalAT14 (V285H/S287F/Q182I), SalAT14 (S287F/Q182I/F210M) and SalAT14 (V285H/S287F/F210M). These mutational residues (especially Phe287) restrict the rotation freedom of the malonyl group of the substrate. Notably, V285H and F210M mainly reduce the steric hindrance for S287F. Without F210M, Phe287 might approach the substrate so closely that the carboxyl of MMCOA would bend and interact with the catalytic Ser181 (Fig. S12†). Without V285H, Phe287 would be close to the substrate, which exerts an adverse effect on the distance between the O<sub>γ</sub> atom of Ser181 and the thioester carbon of MCOA (Fig. S13†). In the presence of V285H/S287F/F210M, Gln182 would form strong hydrogen bonds with the carboxyl

of MCOA, leading to a long  $d(CM1-OG)$  distance (Fig. S14†). Hence, the four mutations are necessary and mutually reinforcing. With all these mutations, the malonyl moiety could adopt a suitable orientation at the active site and the population of the “active” conformation ( $d(N_{\epsilon}-HG) \leq 3$  Å and  $d(CM1-OG) \leq 3.5$  Å) is up to 22.28% (Fig. S15†). From Table S3,† we observed that the hydrogen-bonding interactions between SalAT14 (V285H/S287F/Q182I/F210M) and MCOA have increased, aiding in positioning and stabilizing the substrate. In the SalAT14 quadruple mutant, the subtle reinforcement of steric hindrance likely accounts for the alteration of substrate specificity.

The analysis of these mutants indicates that hydrophobic interactions collaborate with hydrogen-bond interactions to position the substrate and are essential for forming the conformation suitable for reaction. Meanwhile, the steric hindrance may exert a previously unimagined effect in the active site pocket. As we analysed above, the H<sub>γ</sub> atom of Ser181 would have an interaction with the carboxyl group of the substrate in SalAT14–MMCOA. Actually, this phenomenon could be observed in both SalAT14–EMCOA and SalAT14–MCOA. It prompts us that the sophisticated hydrophobic interactions and the hydrogen-bond interactions jointly act on adjusting the position of the malonyl moiety of the substrate, especially the carboxyl group. It's significant to avoid the close contact between the H<sub>γ</sub> atom of Ser181 and the carboxyl group of the substrate. Excessive displacement of the substrate is observed in this conformation and results in the long  $d(CS1-OG)/d(CM1-OG)$  distance. Therefore, the interactions between the malonyl moiety and active sites must be optimized to obtain the appropriate position in altering AT substrate specificity.

## Conclusion

There are great efforts on varying the structures of polyketides in the pursuit of creating therapeutically active molecules. AT domains of modular PKSs are responsible for introducing structural complexity of the resulting polyketides by selecting and incorporating alternative substrates. Thus, the AT domains that exhibit narrow substrate specificity are attractive engineering targets. Our work sheds light on the mechanism for extension unit selection of SalAT14 through computational approaches. Employing MD and QM/MM methods, this study tried to unravel the molecular basis of the transfer of acyl moieties from the CoA to SalAT14. We find that in comparison with SalAT14–MMCOA and SalAT14–MCOA, the “active” conformations are more prone to form in SalAT14–EMCOA. In addition, the hydrophobic interactions combined with the hydrogen-bonding network are pivotal to the substrate positioning and the formation of the “active” conformers. The computational results presented here indicate that the acyl transfer follows a similar way to the mechanism raised by Pedro and Sousa<sup>3</sup> and uncover additional perception concerning this reaction. In the first step, the O<sub>γ</sub> from Ser181 launches the nucleophilic attack on the substrate, occurring with the

deprotonation of the Ser181 hydroxyl group. Following this concerted step, the H<sub>γ</sub> connecting with His288's N<sub>ε2</sub> gradually approaches the negative free-thiol group of the CoA, with the O<sub>γ</sub> of Ser181 and the carbonyl of the substrate becoming shorter. Finally, the newly protonated His288 residue is deprotonated by the thiol group of the CoA which would leave the active site. The first transition state (TS1) is the rate-determining step. Arg206–H<sub>2</sub>O583–H<sub>2</sub>O677, Gln153, Gln182 and Ser287 form hydrogen bonds with the carbonyl group of the substrate and contribute significantly to the position and the stabilization of the acyl group throughout the catalysis. The findings also show that the requisite oxyanion hole is formed by the Gln96 and Gln182 residues. Our computational results demonstrate that the ethylmalonyl transfer catalysed by SalAT14 is more favourable than malonyl and methylmalonyl. And the concerted nucleophilic attack on the substrate is the rate-limiting step in the first half of transacylation. The knowledge gathered in our work elucidates the catalytic mechanism of SalAT14 and different effects dictating AT's substrate specificity, which would pave the way for the design of the AT domain.

## Author contributions

T. S., Z. D., L. B., Y. Z., Q. M. and J. Z. conceived and designed the investigation. H. J., L. L., F. Z. and W. T. performed calculations and analyses. H. J. and J. Z. wrote the paper.

## Conflicts of interest

The authors declare that they have no conflicts of interest.

## Acknowledgements

This work was supported by the National Key R&D Program of China (2019YFA0905400, 2020YFA0907900) and National Natural Science Foundation of China (31770068, 32070040, 31800045).

## References

- H. Jenke-Kodama and E. Dittmann, *Nat. Prod. Rep.*, 2009, **26**, 874–883.
- A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
- P. Paiva, S. F. Sousa, M. J. Ramos and P. A. Fernandes, *ACS Catal.*, 2018, **8**, 4860–4872.
- C. Khosla, R. S. Gokhale, J. R. Jacobsen and D. E. Cane, *Annu. Rev. Biochem.*, 1999, **68**, 219–253.
- B. J. Dunn and C. Khosla, *J. R. Soc., Interface*, 2013, **10**, 20130297.
- Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11124–11129.
- M. Oliynyk, M. J. Brown, J. Cortes, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1996, **3**, 833–839.
- S. Yuzawa, K. Deng, G. Wang, E. E. Baidoo, T. R. Northen, P. D. Adams, L. Katz and J. D. Keasling, *ACS Synth. Biol.*, 2017, **6**, 139–147.
- L. B. Pickens, Y. Tang and Y. H. Chooi, *Annu. Rev. Chem. Biomol. Eng.*, 2011, **2**, 211–236.
- C. P. Ridley, H. Y. Lee and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4595–4600.
- H. Petkovic, A. Sandmann, I. R. Challis, H. J. Hecht, B. Silakowski, L. Low, N. Beeston, E. Kuscer, J. Garcia-Bernardo, P. F. Leadlay, S. G. Kendrew, B. Wilkinson and R. Muller, *Org. Biomol. Chem.*, 2008, **6**, 500–506.
- F. Zhang, T. Shi, H. Ji, I. Ali, S. Huang, Z. Deng, Q. Min, L. Bai, Y. Zhao and J. Zheng, *Biochemistry*, 2019, **58**, 2978–2986.
- F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortes, B. A. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 489–494.
- E. Kalkreuter, J. M. CroweTipton, A. N. Lowell, D. H. Sherman and G. J. Williams, *J. Am. Chem. Soc.*, 2019, **141**, 1961–1969.
- I. Koryakina, C. Kasey, J. B. McArthur, A. N. Lowell, J. A. Chemler, S. Li, D. A. Hansen, D. H. Sherman and G. J. Williams, *ACS Chem. Biol.*, 2017, **12**, 114–123.
- Y. Li, W. Zhang, H. Zhang, W. Tian, L. Wu, S. Wang, M. Zheng, J. Zhang, C. Sun, Z. Deng, Y. Sun, X. Qu and J. Zhou, *Angew. Chem., Int. Ed.*, 2018, **57**, 5823–5827.
- K. Bravo-Rodriguez, A. F. Ismail-Ali, S. Klopries, S. Kushnir, S. Ismail, E. K. Fansa, A. Wittinghofer, F. Schulz and E. Sanchez-Garcia, *ChemBioChem*, 2014, **15**, 1991–1997.
- M. Voss, D. Das, M. Genz, A. Kumar, N. Kulkarni, J. Kustoscz, P. Kumar, U. T. Bornscheuer and M. Höhne, *ACS Catal.*, 2018, **8**, 11524–11533.
- C. Jiang, H. Wang, Q. Kang, J. Liu and L. Bai, *Appl. Environ. Microbiol.*, 2012, **78**, 994–1003.
- M. E. Yurkovich, P. A. Tyrakis, H. Hong, Y. Sun, M. Samborsky, K. Kamiya and P. F. Leadlay, *ChemBioChem*, 2012, **13**, 66–71.
- G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.
- M. V. Shapovalov and R. L. Dunbrack, Jr., *Structure*, 2011, **19**, 844–858.
- R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537–W541.
- 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, D. Williams-Young, 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 and D. J. Fox, *Gaussian 16*, avx2, Gaussian, Inc., Wallingford CT, 2016.

- 25 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, *J. Am. Chem. Soc.*, 1996, **118**, 2309–2309.
- 26 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. I. Cheatham, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. A. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, University of California, San Francisco, 2018.
- 27 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1992, **98**, 10089–10092.
- 28 J.-P. Ryckaert, G. Ciccotti and H. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 29 D. R. Roe and T. E. Cheatham, 3rd, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.
- 30 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38, 27–38.
- 31 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *J. Mol. Struct.: THEOCHEM*, 1999, **461–462**, 1–21.
- 32 T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. Montgomery, K. Morokuma and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 815–826.
- 33 S. Smith and S. C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.
- 34 S. Poust, I. Yoon, P. D. Adams, L. Katz, C. J. Petzold and J. D. Keasling, *PLoS One*, 2014, **9**, e109421.
- 35 V. S. Rangan and S. Smith, *J. Biol. Chem.*, 1997, **272**, 11975–11978.
- 36 F. Zhang, H. Ji, I. Ali, Z. Deng, L. Bai and J. Zheng, *ChemBioChem*, 2020, **21**, 1309–1314.
- 37 H. B. Burgi, J. D. Dunitz, J. M. Lehn and G. Wipff, *Tetrahedron*, 1974, **30**, 1563–1572.