Article

Biosynthesis of Vanillin by Rational Design of Enoyl-CoA Hydratase/Lyase

Qi Ye [1,†], Weizhuo Xu [2,†], Yanan He [1], Hao Li [1], Fan Zhao [1], Jinghai Zhang [1,*] and Yongbo Song [1,*]

[1] School of Life Sciences and Biopharmaceuticals, Shenyang Pharmaceutical University, 103 Wenhua Road, Shenhe District, Shenyang 110016, China; yeokqi@163.com (Q.Y.); heyanan987@outlook.com (Y.H.); 15694142939@163.com (H.L.); zhaofan@syphu.edu.cn (F.Z.)

[2] School of Functional Food and Wine, Shenyang Pharmaceutical University, 103 Wenhua Road, Shenhe District, Shenyang 110016, China; weizhuo.xu@syphu.edu.cn

[*] Correspondence: zhangjinghai@syphu.edu.cn (J.Z.); songyongbo@syphu.edu.cn (Y.S.); Tel.: +86-024-43520923 (Y.S.)

[†] These authors contributed equally to this work.

Abstract: Vanillin holds significant importance as a flavoring agent in various industries, including food, pharmaceuticals, and cosmetics. The CoA-dependent pathway for the biosynthesis of vanillin from ferulic acid involved feruloyl-CoA synthase (Fcs) and enoyl-CoA hydratase/lyase (Ech). In this research, the Fcs and Ech were derived from *Streptomyces* sp. strain V-1. The sequence conservation and structural features of Ech were analyzed by computational techniques including sequence alignment and molecular dynamics simulation. After detailed study for the major binding modes and key amino acid residues between Ech and substrates, a series of mutations (F74W, A130G, A130G/T132S, R147Q, Q255R, ΔT90, ΔTGPEIL, ΔN1-11, ΔC260-287) were obtained by rational design. Finally, the yield of vanillin produced by these mutants was verified by whole-cell catalysis. The results indicated that three mutants, F74W, Q147R, and ΔN1-11, showed higher yields than wild-type Ech. Molecular dynamics simulations and residue energy decomposition identified the basic residues K37, R38, K561, and R564 as the key residues affecting the free energy of binding between Ech and feruloyl-coenzyme A (FCA). The large changes in electrostatic interacting and polar solvating energies caused by the mutations may lead to decreased enzyme activity. This study provides important theoretical guidance as well as experimental data for the biosynthetic pathway of vanillin.

Keywords: vanillin; biosynthesis; rational design; enoyl-CoA hydratase/lyase; molecular dynamics

1. Introduction

Vanillin (4-hydroxy-3-methoxybenzaldehyde, CAS 121-33-5) is one of the most important flavoring agents in the world. Presently, the industrial vanillin product could be manufactured in three routes. Firstly, the classical natural vanillin can be extracted from natural vanilla beans [1]. Secondly, various chemical synthesis processes could be applied to generate vanillin from diversified sources, such as lignin and cresols [2]. For the complement/alternation of the unstable source and high costs of the natural vanilla pods, a third route which had been explored for years, and researchers had attempted multiple pathways to generate vanillin by biosynthesis with different raw materials, such as eugenols, isoeugenol and ferulic acids [3–5].

As early as 1977, Tadasa had reported the vanillin biotransformation from eugenols by *Corynebacterium* sp. [6]. Later on, more studies had disclosed the same biotransformation results with different microbes [7–10]. More extensive works indicated that isoeugenol could also be used as a substrate to generate vanillin [8,11,12], and enzymatic processes were also pointed out [13,14]. When ferulic acid was used as starting materials, multiple microbes had been tested and provided valuable genetic and enzymatic results for the vanillin

biosynthesis pathways. The pathways could be classified into CoA-independent [15] and CoA-dependent (Scheme 1).
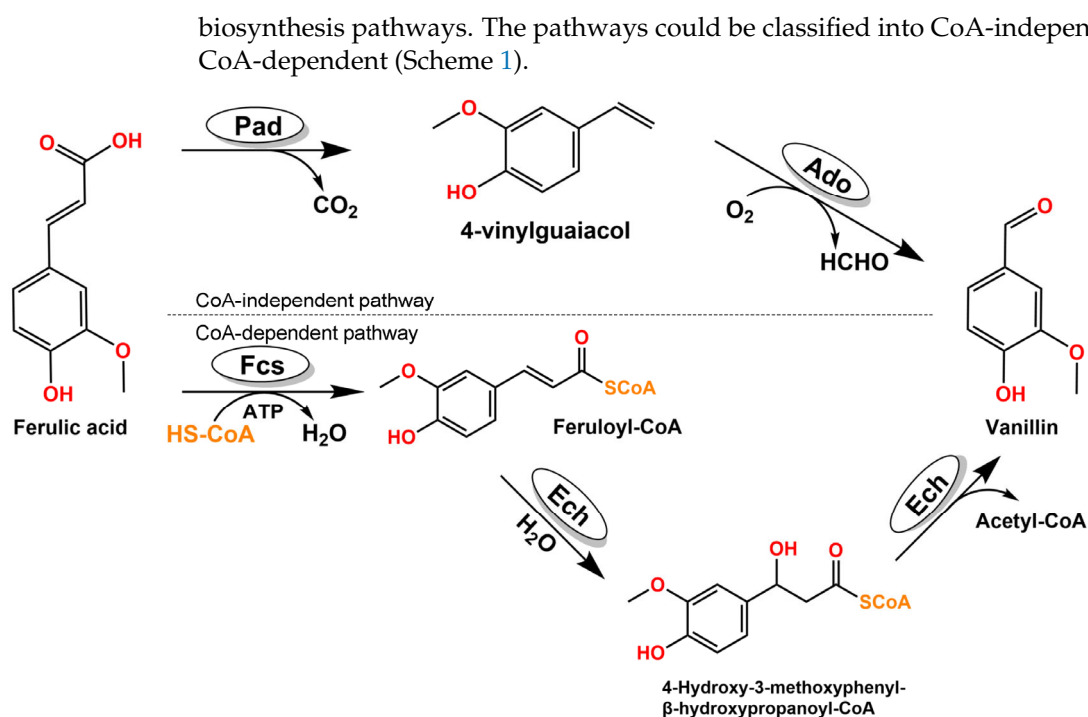
Scheme 1. CoA-independent pathway and CoA-dependent pathway for biosynthesis of vanillin from ferulic acid.

In the CoA-independent pathway, the phenolic acid decarboxylase (Pad) and aromatic dioxygenase (Ado) could transform the ferulic acid into vanillin without coenzyme A [15]. This pathway also converts many other renewable lignin-related aromatics into valuable chemicals and pharmaceutical intermediates, thus enabling value-added utilization of lignin. In the CoA-dependent pathway, a gene encoding an enoyl-CoA hydratase/lyase (Ech) enzyme for the hydration and non-oxidative cleavage of feruloyl-CoA to vanillin had been isolated and characterized in 1998 [16]. The gene encoding feruloyl-CoA synthase (Fcs) was studied in *Pseudomonas* sp. HR199 (DSM7063) [17]. By using these two above genes, vanillin was been generated in engineered *E. coli* [18], *Rhodococcus* strains [10], *Pseudomonas* strains [19,20], *Pediococcus* strains [21], *Bacillus* strains [22,23] and *Streptomecyes* strains [24,25]. However, few reports have been disclosed for these enzymes rational design to improve their activities, resulting in the catalytic activity of these enzymes gradually becoming a bottleneck in the vanillin production pathway. Recently, with the development of computer-based technologies in the biocatalysis field, enzyme rational design by bioinformatics tools can effectively improve the catalytic activity of the enzyme.

In the study, due to the lack of reliable structural data for Fcs, only the Ech from *Streptomyces* sp. strain V-1 was selected as the enzyme to be modified. A various of bioinformatics techniques, including multiple sequence alignment, sequence conservation analysis, homology modeling, molecular docking and molecular dynamics simulation to clarify the molecular mechanism of the enzyme-substrate complex, and nine mutants (F74W, A130G, A130G/T132S, R147Q, Q255R, ΔT90, ΔTGPEIL, ΔN1-11, ΔC260-287) were ultimately generated by rational design to enhance production of vanillin. Whole-cell catalysis was performed for wild-type and all mutants using the pET-Duet-1 vector with *fcs* gene and above mutants to verify the enzyme activity for transforming ferulic acid to vanillin. Among all the mutants, F74W, R147Q and ΔC260-287 displayed the best enzyme activities. It is believed that these experimental results could provide valuable ideas and design guide for the biosynthesis of vanillin in the future.

2. Results

2.1. Sequence and Conservation Analysis of Ech

The targeted Ech sequence was selected from *Streptomyces* sp. strain V-1 (accession number: S5LPF1) in UniProt database. The multiple sequence alignment (MSA) of Ech were performed by retrieving 250 sequences which exhibit more than 60% identity with the targeted sequence from UniProt database. The sequence logo diagram (Figure S1—Supplementary Material) indicated several conserved regions, which were considered related to binding and catalysis between Ech enzyme and substrate feruloyl-CoA. To elucidate the essential residues which interacted with the enzyme, the substrate feruloyl-CoA could be categorized into two parts, i.e., the feruloyl group and the CoA group according to its structure. As shown in Figure 1, residues of Y83 and G160 (orange), the binding site for the feruloyl group, displayed the highly consistency in Ech enzymes, indicating their important role in the catalytic functions. Meanwhile, the proximity of feruloyl group close to the catalytic center should also be considered. The residues of R38, A40, A76–L80, W125, F127–G129, S151, E152 and W155 (green) binding with the CoA moiety displayed complete conservation, except E72. It could be recognized that these binding residues were crucial for the stability of the CoA group, and it identified that the CoA group was used to bind and stabilize the substrate. Within the binding pocket of Ech, there are both highly and partially conserved residues, so it is reasonable and practical to design and modify these conserved residues to increase the enzymatic catalytic activities.
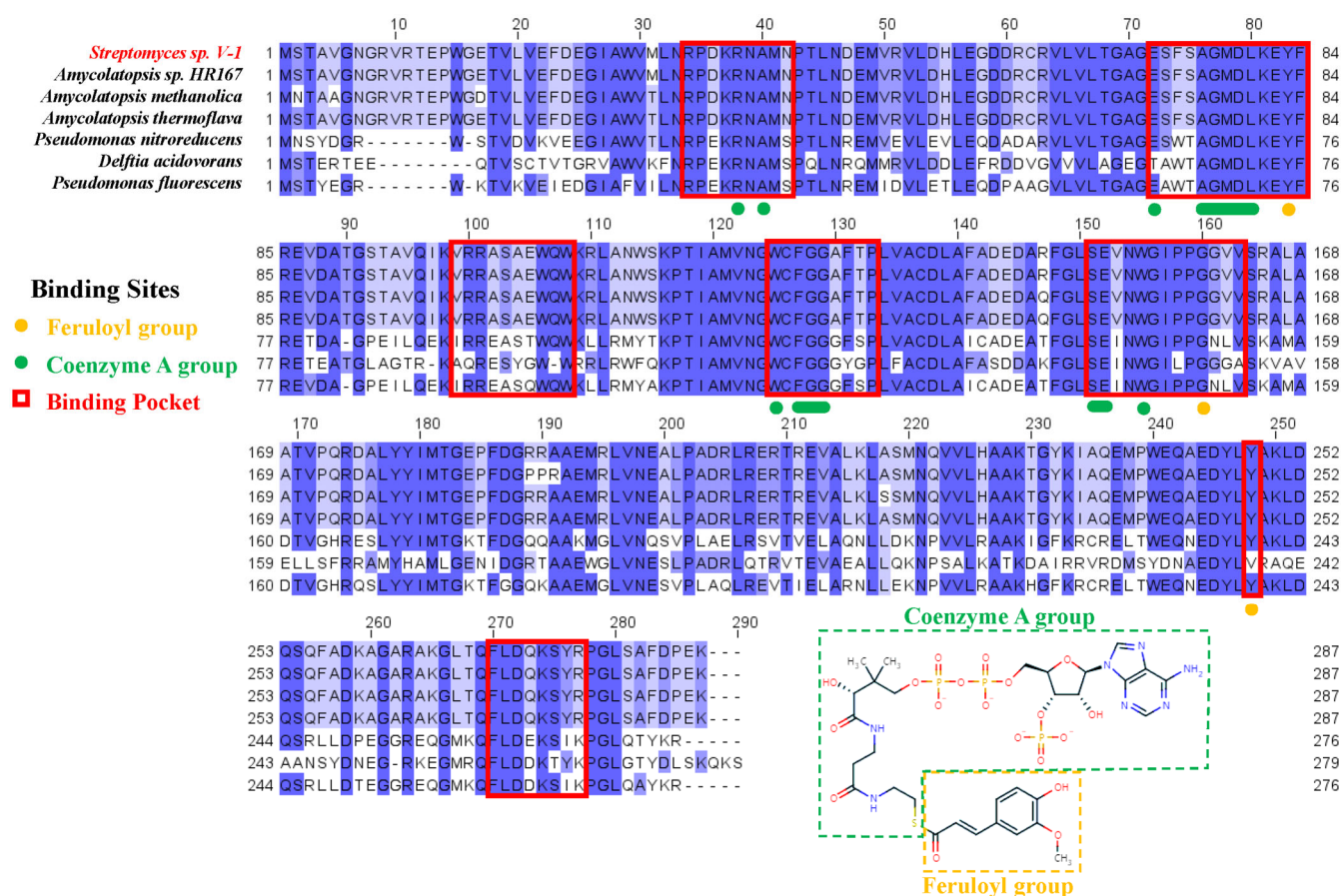

Figure 1. Sequence alignment of Ech/Hchl from different species. Binding sites for feruloyl moiety and CoA moiety are annotated using orange and green, respectively, and red boxes represent the binding pocket regions.

2.2. Structural Features of Ech and Its Binding Mode with FCA

In order to further understand the structure and binding mode with the substrate feruloyl-coenzyme A (FCA), the Ech monomer 3D structure was established based on the hexamer crystal data of hydroxycinnamoyl-CoA hydratase lyase (Hchl from *Pseudomonas fluorescens*, PDB: 2VSS) (Figure S2a—Supplementary Material). To keep the intact catalytic pocket, the Ech monomer was compared with the hexamer model, and generated the 3D structure of the Ech dimer (Figure S2b—Supplementary Material). In this research, unless specified, the structure model of Ech was the dimer structure with Chain A and Chain B.

The electrostatic potentials were calculated using Adaptive Poisson-Boltzmann Solver (APBS) to visualize the electrostatic potential surface, which indicated that Chain A and Chain B of the Ech dimer together form a long and narrow cavity as the binding pocket (Figure 2a). The entrance of the binding pocket presented a higher positive potential distribution, with the basic amino acid residues on both sides (Chain A: K37, R38; Chain B: K274, R277), which is favorable for the binding of negatively charged CoA moiety. Meanwhile, to further understand the statistical potential distribution, the binding pockets of Chain A and Chain B were disassembled. It could be found that polar uncharged residues (Y83, S103, G129, S151 and W155) and M78 were distributed in the middle of the binding pocket of Chain A, which is fundamental for the binding of substrate FCA. Acidic amino acid residue E152 of Chain A is located in the catalytic center in the binding pocket, which was indispensable for the catalytic reaction. These results indicated that the binding pocket in Chain A constituted the major region for the whole binding pocket (Figure 2b). Compared with Chain A, the binding pocket of Chain B was composed of only several residues, such as the K274, R277 at and G279 near the entrance, and polar uncharged Y248 located innermost of the pocket, which was believed to stabilize the terminal of substrate FCA (Figure 2c).
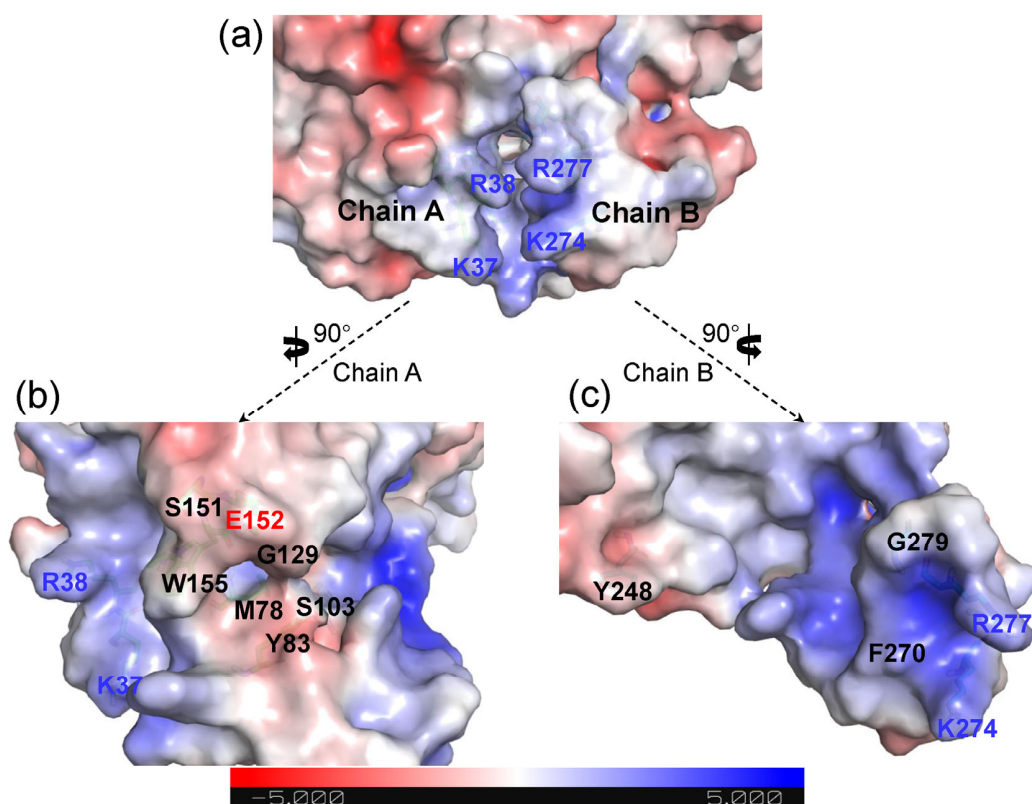

Figure 2. APBS Electrostatics potential surface visualization of the binding pocket of Ech dimer (**a**). Chain A (**b**) and Chain B (**c**) are shown separately.

The Ech-FCA complex was generated by molecular docking, with the binding mode of Ech and FCA shown in Figure 3. The binding energy of −8.5 kcal/mol generated by AutoDock Vina indicated a stable complex structure between Ech and FCA. Similarly, the

binding sites in the binding pocket could be categorized into two moieties, i.e., the feruloyl binding moiety and the CoA binding moiety, corresponding to the feruloyl group and the CoA group mentioned above. The CoA binding moiety consists of quantities of hydrogen bonds and salt bridges, which is the starting point of catalytic reaction. Among these hydrophilic and charged residues, basic residues R38 and K274 (Chain B) could form stable salt bridge with the phosphor group of FCA, which is the structural element for the stable binding between FCA and Ech. At the same time, some polar residues such as S151, W155 and G279 (Chain B) as well as residues A76, M78 and L80 could bind with the CoA binding moiety by hydrogen bonds, further strengthening the binding extent of FCA. The feruloyl binding moiety involved in the catalytic reactions, the carbonyl group in feruloyl binding moiety could establish the hydrogen bonds with M78 and G129, and Y248 (Chain B) could also establish the hydrogen bonds with the phenolic hydroxyl group in the feruloyl binding moiety. The most important E152 is located above the FCA, which maintains the proper distance (4.44 Å) between the substrate and reactive site, to guarantee the initiation of the catalytic reaction. All these docking results are highly consistent with the sequence analysis, so it is practical to start the rational design on these docking results.
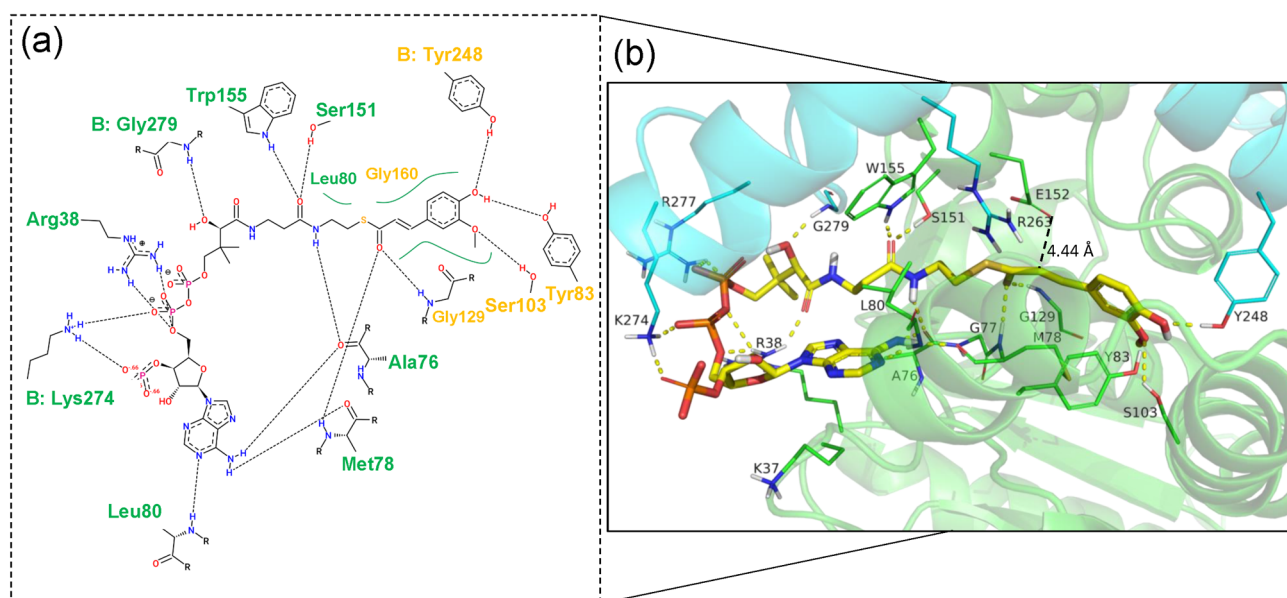

Figure 3. The binding mode of Ech and FCA in 2D (**a**) and 3D (**b**). The residues in the feruloyl binding moiety (orange) and the CoA binding moiety (green) are labeled in 2D binding mode. Key residues and FCA are shown as sticks, and Ech structural is shown as a cartoon in 3D binding mode.

2.3. Rational Design for Ech

Based upon the above profiling for the catalytic mechanism, the primary candidate region for rational design focuses on the feruloyl binding moiety. Conservation analysis indicated that A130 and T132 are moderately conserved, which made these two loci prone to become smaller residues like Gly and Ser, so that a single mutant A130G and a double mutant A130G/T132S were designed, to enlarge the cavity for the binding pocket and strengthen the flexibility to increase the catalytic efficiency. When the Ech from *Streptomyces* sp. V-1 was performed homologous alignment with other genera, some other candidate residues could be chosen and substituted to optimize the Ech catalytic activity. Hence, F74 was mutated to Trp, R147 was mutated to Gln, and Q255 was mutated to Arg based on sequence alignment (Figure S3—Supplementary Material).

Subsequently, a 100 ns molecular dynamics (MD) simulation was performed towards the Ech-FCA complex in three independent replicas to identify the residues with unstable effects on the Ech-FCA complex. Due to the irregular structure of residues 1–14 at the N-terminus, residues 1–14 were removed at the beginning of the simulation to ensure the

acceptable and validity of the initial model. RMSD and RMSF during MD simulation were shown in Figure 4a,b. It could be read from the first 50 ns stage, that the RMSD indicated a relatively stable condition, while in the next 50 ns the average RMSD was 5.64 Å which indicated some unstable factors in the wild-type (WT) of Ech. The RMSF result also verified this proposition. The RMSF indicated the most significant fluctuation region located in the 232–287 from N terminal (Chain A), which sited at the junctions of the polymer. However, residues 519–574 (232–287 in Chain B) indicated relatively smaller fluctuations. It could be speculated that the residues 232–287 in Chain B established tight interactions with Chain A and substrate FCA. Also, a relatively large RMSF was found near T90. Hence, two deletion mutants (ΔT90 and ΔTGPEIL) were designed. ΔT90 means that the T90 residue was deleted, and ΔTGPEIL means that the residues 91–95 were replaced by GPEIL besides the T90 deletion. Additionally, to decrease the influence of residues from both N and C terminals, two truncated mutants were designed as ΔN1-11 and ΔC260-287. All these designed mutants and their descriptions are listed in Table 1, and their position in protein can be found in Figure 4c.
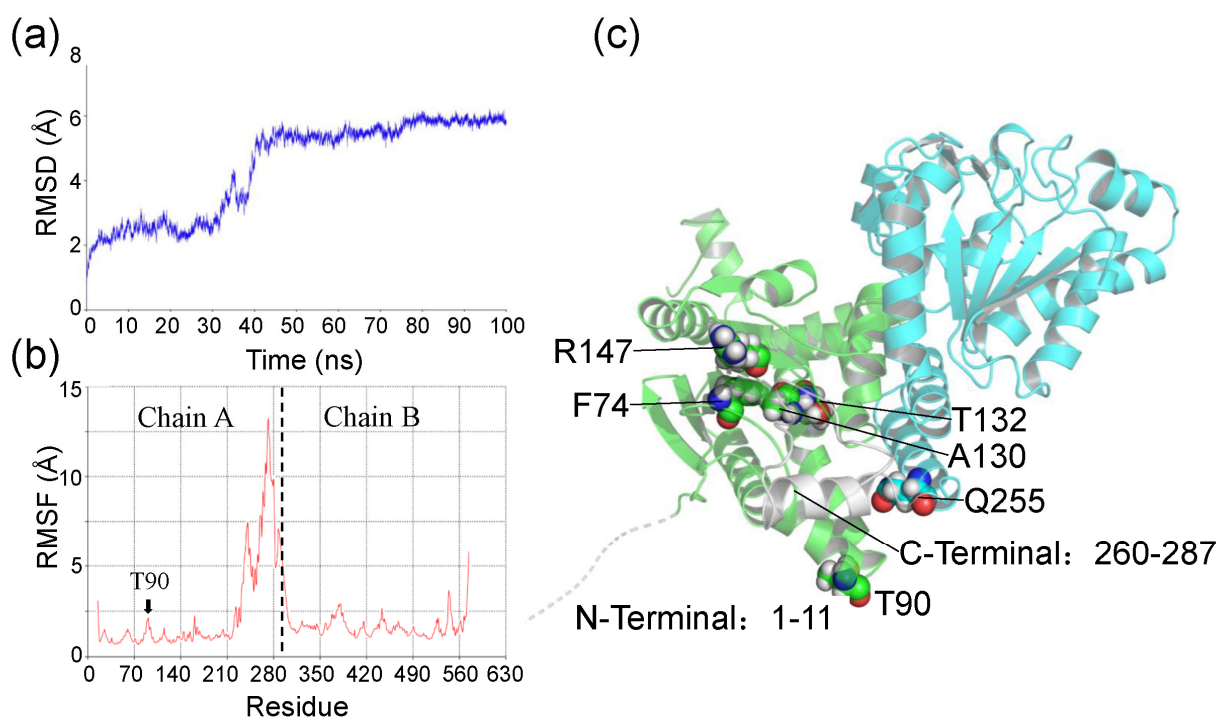

Figure 4. (**a**) Location of designed mutations in Ech dimer. The Ech dimer structure is shown as a cartoon, while the atoms of the design mutations are shown as spheres. (**b**) RMSD curves and (**c**) RMSF curves of wild-type Ech-FCA complex in MD simulations. Chain A and Chain B in Ech dimer are separated in RMSF curves by dotted lines.

Table 1. Designed mutations and related descriptions for the Ech.

Mutation Type	Mutant	Description
Single	A130G	Changing the size and flexibility of the active cavity
Double	A130G/T132S	Changing the size and flexibility of the active cavity
Single	R147Q	Substitution from *Amycolatopsis thermoflava*
Single	F74W	Substitution from *Pseudomonas fluorescens*
Single	Q255R	Substitution from *Pseudomonas fluorescens*
Deletion	ΔT90	Reducing the flexibility of the structure
Deletion & Multiple	ΔTGPEIL	Reducing the flexibility of the structure
Truncation	ΔN1-11	N-Terminal truncation
Truncation	ΔC260-287	C-Terminal truncation

2.4. Molecular Dynamics Simulation and Binding Free Energy Analysis for Ech and Its Mutants

To further understand the dynamic influence of different mutations on Ech, as well as the structural differences between mutants and wild-type of Ech, 100 ns molecular dynamics (MD) simulations were performed on all mutants. To ensure the credibility of the simulation experiments, the parametric conditions of the MD were consistent with the WT as described above. All mutants were constructed based on the initial model. The molecular mechanisms of each mutant would be reassessed through the analysis of molecular dynamics simulations. The RMSD of three independent MD simulations confirmed that all systems had reached equilibrium during the 100 ns MD simulations (Figures S4a and S5—Supplementary Material). The RMSD histograms showed the changes and distribution of the conformations of the system, and the WT appeared to have different conformational distributions (Figure S4b—Supplementary Material), suggesting that the conformation of WT was unstable, thus affecting the catalytic activity. Compared to the WT, most mutations (such as F74W and Q255R) reduce the conformational distribution of the structure to some extent, indicating that appropriate mutations can stabilize the overall structure of Ech. The specific structural features and locations of the mutants are shown in Figures 5 and S6 in the Supplementary Material. To explore the effect of mutations on catalytic activity in more detail, the MD simulation results together with the experimental validation will be discussed in the Section 3.
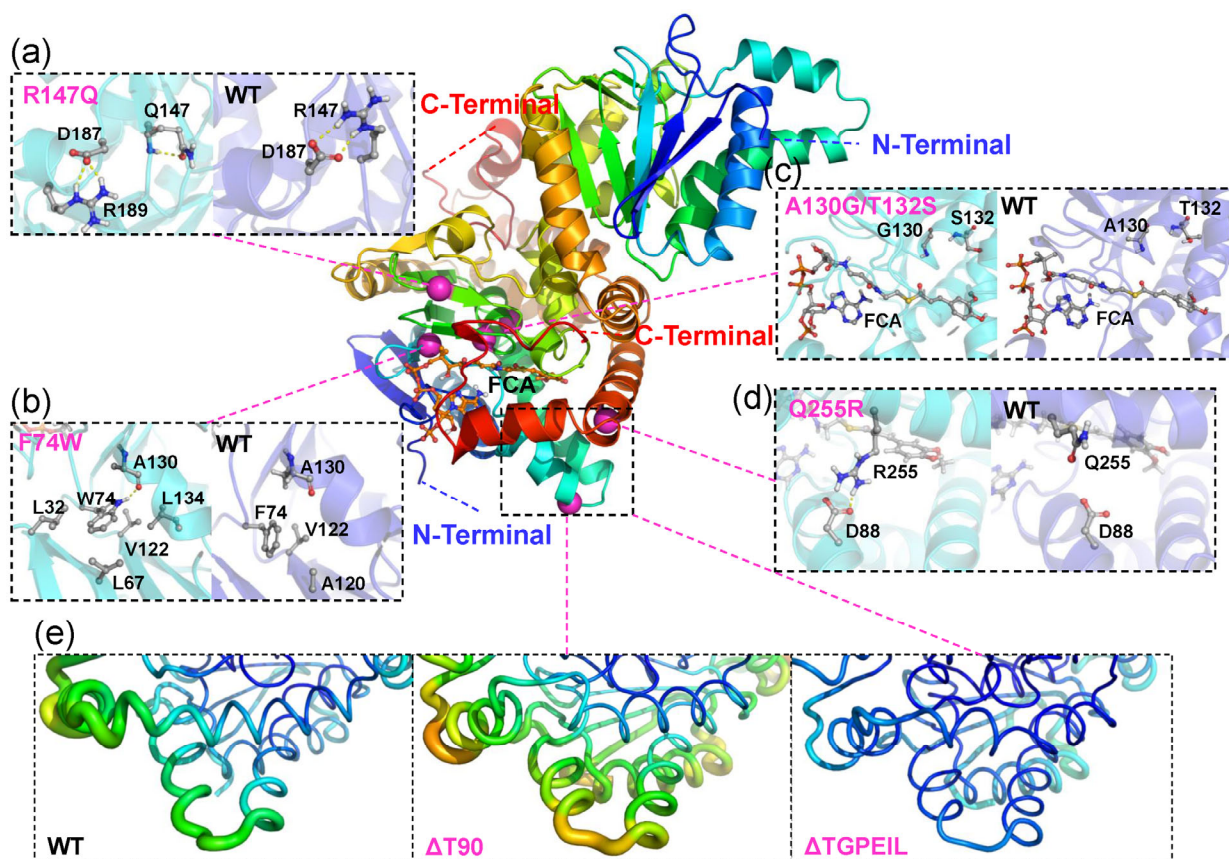

Figure 5. Structural features and location of Ech mutations. The Ech dimer is shown as a cartoon, while the Cα atoms of the mutations are shown as purple spheres. Structural effects of R147Q (**a**), F74W (**b**), A130G/T132S (**c**) and Q255R (**d**) mutants (cyan) compared to WT (blue) in MD simulations. Key residues and FCA are shown in ball and stick representations. (**e**) The conformational flexibility of WT and ΔT90 and ΔTGPEIL mutants is displayed using predicted B-factors in MD simulations.

MM/GBSA were performed to calculate the system binding free energy for wild-type Ech and mutants. It could be read from Table 2 that the wild-type Ech exhibited a

binding free energy of -102.34 kcal/mol, indicating a high binding between WT and FCA. Mutants A130G, A130G/T132S and Q255R displayed lower energies compared with WT, indicating their higher combination with the substrate FCA. While the binding free energies of F74W, ΔT90 and ΔTGPEIL were slightly reduced, and R147Q remained almost constant. Furthermore, compared to WT, mutants F74W and R147Q showed a decrease in electrostatic interaction energies (ΔE_{EL}) of 75.34 kcal/mol and 35.27 kcal/mol, respectively, while polar solvation energies (ΔE_{GB}) increased by 78.54 kcal/mol and 24.11 kcal/mol, respectively. On the other hand, the other mutants (except ΔTGPEIL) almost always showed large energy changes (exceeding 100 kcal/mol) in both the electrostatic interaction energy and the polar solvation energy, suggesting that the mutations significantly altered the interaction mode between Ech and FCA. The binding affinity of the truncation ΔC260-287 to FCA was drastically reduced, suggesting that the partially truncated amino acid residues at the C-terminal end play a very crucial role in the binding of the FCA, and that the significant reduction in the polar solvation energy is indicative of an increase in the surface area of the substrate exposed to the solvent environment.

Table 2. MM/GBSA binding free energy for wild-type Ech and its mutants.

System	ΔE_{VDW}	ΔE_{EL}	ΔE_{GB}	ΔE_{SURF}	ΔE_{Total}
WT	-113.41	76.80	-51.39	-14.34	-102.34
A130G	-102.45	-84.87	93.89	-13.28	-106.71
A130G/T132S	-105.73	-104.10	112.78	-13.27	-110.32
F74W	-108.98	1.46	27.15	-13.70	-94.07
R147Q	-98.27	35.37	-27.28	-12.81	-102.99
Q255R	-98.50	-181.67	186.42	-13.27	-107.02
ΔT90	-101.89	-42.82	59.27	-13.49	-98.93
ΔTGPEIL	-108.30	67.65	-43.93	-13.79	-98.37
ΔC260-287	-77.13	298.71	-287.68	-9.94	-76.04

Note: The unit of binding free energy is kcal/mol; ΔE_{VDW} and ΔE_{EL}: represent the van der Waals interaction energy and electrostatic interaction energy between molecules, which are an important part of the non-covalent interaction between the enzyme and the substrate; ΔE_{GB} and ΔE_{SURF} respectively represent polar solvation energy and nonpolar solvation energy, which is an energy term describing the solvation effect around the enzyme; ΔE_{Total} represents the total binding free energy between the enzyme and the substrate in the system, indicating the binding between the enzyme and the substrate strength.

Then, it is necessary to perform the per-residue energy decompositions on the total binding free energy of the system into the energy contribution of each residue, so as to elucidate/identify the residues that significantly affect the system binding free energy (Figure S7—Supplementary Material), and providing valuable data to support further discussion of the effect of each residue. All residues that exhibited energy contribution are listed in Table S1 of the Supplementary Material.

2.5. Experimental Verification of Wild-Type Ech and Its Mutants

Wild-type Ech and all mutants were constructed and introduced into the pET-Duet-1 expression vectors, and the ferulic acid was supplemented as substrate. After 24 h of biotransformation, the resulting vanillin was tested by GC and the productivity was plotted in Figure 6.

As shown in Figure 6, the single mutants F74W and R147Q with residue substitutions both showed an approximately 2-fold increase in activity over WT, whereas the Q255R mutant lost all catalytic activity. In addition, the N-terminal truncation ΔN1-11 showed an increase in catalytic activity, suggesting that the disordered motif at the N-terminus has an effect on the catalytic activity of the enzyme. In contrast, the C-terminal truncation (ΔC260-287) showed the opposite effect, indicating that the residues in the C-terminal were necessary for the catalytic activity of Ech. Unfortunately, neither the A130G nor the A130G/T132S mutants near the active site showed higher catalytic activity, which may indicate a potentially critical role of A130 for catalytic activity. Lastly, the two deletion

mutants ΔT90 and ΔTGPEIL also show reduced catalytic activity, indicating that a deletion mutation at site 90 is not a favorable option.
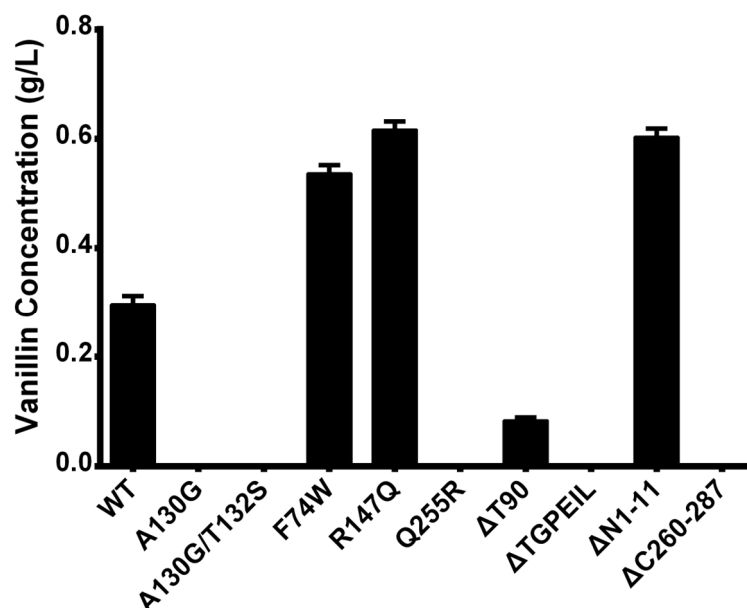

Figure 6. Productivities for the wild-type and mutants of Ech.

Among these mutants, there are only three positive mutants displayed better productivity, one with decreased performance, and other with no products generated. The detailed analysis will be discussed subsequently.

3. Discussion

Directed evolution is classical method for modifying the target enzyme activities for years [26–28]. Rational designs are strongly dependent on the enzyme structure and bioinformatics methods. In recent years, the molecular mechanisms of a large number of enzymes have been elucidated by molecular dynamics simulation methods [29–31], thus enabling the rational design of enzymes. In this study, Ech, a key enzyme in the biosynthesis pathway of vanillin, was selected to improve its catalytic activity through rational design. The sequence and structure features of Ech and its binding mode to FCA were investigated by bioinformatics techniques such as sequence comparison, molecular docking and molecular dynamics simulation. Nine mutants (F74W, A130G, A130G/T132S, R147Q, Q255R, ΔT90, ΔTGPEIL, ΔN1-11 and ΔC260-287) were rationally designed based on sequence and structure guidance, of which three mutants (F74W, R147Q and ΔN1-11) were eventually confirmed to have enhanced activity through experimental validation. Here, all mutants were divided into four groups according to the type of design, and the main factors affecting and improving the catalytic activity of Ech are discussed by comparing the computational data with the experimental results.

For the group with residues substituted from *Pseudomonas fluorescens* and *Amycolatopsis thermoflava*, F74W and R147Q exhibited positive vanillin production, indicating the higher Ech activity, whereas the other mutant Q255R from *Pseudomonas fluorescens* was completely inactive. Among them, both mutants F74W and R147Q displayed more than twice Ech activity than WT. Despite F74 and R147 being located some distance away from the active pocket, and not interacting with the substrate FCA directly, there are some indescribable effects which may increase the Ech activity. Also, F74W could interact with many more residues to establish the hydrogen bond and hydrophobic interactions, including A130, L32, L67, V122 and L134 (Figure 5b). Comparably, the F74 residue in the WT only interact with A120 and V122 with hydrophobic interaction. However, it seems that R147Q did not increase the Ech activity by strengthening the autologous interactions. The simulation results indicated that due to the elimination of autologous charge, the intrinsic salt bridge

which connected the guanidyl of R147 to the carboxyl of charged D187 could be changed to connect to another basic R189, with hydrogen bond generated in the residue itself (Figure 5a). These changes seemed beneficial for the Ech catalytic activity towards FCA. Meanwhile, F74W displayed a lower binding free energy (−94.07 kcal/mol) than the WT, demonstrating the reduction of electrostatic interaction energy and increase of polar solvation energy. Despite the fact that R147Q displayed some electrostatic interaction energy reduction, the total binding energy (−102.99 kcal/mol) is similar to the WT. However, the residue Gln, which substitutes the Arg, demonstrated a smaller side chain, which may cause the substrate some directional and/or positional changes so as to increase the Ech catalytic activity. Compared with the WT, V162/V163 and G161/V162 energy contributions were identified in Chain A of F74W and R147Q (Table S1—Supplementary Material). As for the inactive mutant Q255R, the simulated conformation showed that Q255R formed a new salt bridge interaction with residue D88 on the other α-helix (Figure 5d), which appeared to promote structural stabilization. Further investigation revealed, however, that this interaction was not beneficial for the catalytic reaction. Since Q255 is close to the release channel of the product, an overly rigid structure may result in a difficult release of the product, leading to a decrease in catalytic activity.

For the group that modified the size and flexibility of the active cavity, mutants A130G and A130G/T132S generated little vanillin production. This unexpected result suggested that the A130 was essential for the catalytic activity, even a minor modification could inactivate this role. This unfavorable result may be partially due to these two residues being too close to the catalytic center (Figure 5c). Also, A130G demonstrated similar binding free energy distribution with A130G/T132S. Compared with WT, both mutants displayed significantly lower binding free energy and higher polar solvation energy, indicating the prominent changes in the interaction between the FCA and Ech, so as to deactivate the enzyme. Also, A130G lost the A76 interaction in Chain A, and L567 interaction in Chain B. Despite the W125/F127 provided better energy contribution in A130G/T132S, only the R564 in Chain B kept original energy (Table S1—Supplementary Material). These changes finally rendered the deactivation of Ech.

For the group with reducing structure flexibility, ΔT90 exhibited lower vanillin production, indicating lower Ech activity. Both the deletion/multiple mutants ΔT90 and ΔTGPEIL exhibited the configuration changes for the vanillin release tunnel. The B-factor color (Figure 5e) indicated more significant flexibility changes for the ΔT90, while the ΔTGPEIL greatly decreased the flexibility. This is crucial so that ΔT90 still demonstrated some Ech activity but ΔTGPEIL did not. The binding free energy of ΔT90 and ΔTGPEIL is similar to WT, with lower binding strength. Partially due to the conformation changes rendered by the mutation, the substrate position in the Ech could be changed so as to hamper the catalytic activity. Compared with the WT, ΔT90 and ΔTGPEIL almost kept all the essential residues, and it could be calculated that the W125 and F127 in Chain A exhibited more energy contribution (Table S1—Supplementary Material). So, the activity of Ech decrease and loss may be attributed to the regional structure change.

For the group with N and C terminal truncation, ΔN1-11 also exhibited positive vanillin production, indicating higher Ech activity. This result indicated that the N terminal unordered motif is important for the entire enzyme stability, the elimination of these unorder motifs could greatly increase the Ech stability. On the contrary, the C terminal truncation ΔC260-287 could not catalyze the vanillin production, indicating that this C terminal is indispensable for the enzyme activity. As shown in Figure S6 in the Supplementary Material, the loss of those truncated residues renders the significant conformation changes, especially in the CoA binding moiety, which made the substrate FCA unable to keep a stable configuration during the catalytic reaction process, so as to deactivate the enzyme. Meanwhile, for the ΔC260-287, it could be found from the energy decomposition that all the residues that could provide energy contribution were truncated in Chain B (Figure S7 and Table S1—Supplementary Material), so the binding between Ech and FCA were quite unstable, that the catalytic activity were greatly hampered.

In summary, the residues that affected the binding free energy between Ech and FCA could be mainly categorized into two regions, such as K37, R38, A76, M78, L80, W125 and F127 in Chain A, and F557, F561 and R564 in Chain B. Among these residues, the basic residues K37, R38, K561 and R564 demonstrated higher energy contribution in most systems, which indicated their important roles in the binding of Ech and FCA. According to the previous simulation, A76, M78, and L80 could form a stable binding region, which could impose the FCA in the binding pocket. Additionally, non-polar residues W125 and F127 could contribute to the hydrophobic interaction with FCA, so as to strengthen the binding of FCA.

In this study, our rational design strategy is reasonable and verified by the experiment results. Three mutants, F74W, R147Q, and ΔN1-11, displayed a positive increase in the production of vanillin, which provided the basis for rational engineering of the enzyme to improve the vanillin biosynthesis. The MD simulation results could partially explain the unfavorable results with mutation location and stability by hydrogen bonds, hydrophobic interactions etc. However, there are still some unresolved issues, e.g., for the mutants A130G and A130G/T132S, it is still unclear which specific effects of residues A130 and T132 on the catalytic activity of the enzyme. Although there is evidence to suggest that the neighboring residue G129, as an important residue in the formation of the oxyanion hole (OAH) in active sites, may be affected by mutations in neighboring residues, how small changes affect the catalytic activity of Ech could be probably unobservable with classical MD simulation methods. Perhaps quantum mechanics (QM) or quantum mechanics/molecular mechanics (QM/MM) methods could be used in future studies to further investigate the effect of mutations on the catalytic mechanism of Ech.

From this analysis, it could be concluded that the effect of enzyme stability is a crucial factor to be considered when improving the catalytic activity of enzyme catalysis through rational design. Therefore, it is necessary to balance the catalytic activity with the enzyme stability. This rational design strategy could also enlighten other studies for enzyme optimization in the future, and further work was desired to accomplish better results.

4. Materials and Methods

4.1. Bioinformatics Analysis

4.1.1. Multiple Sequence Alignment and Conservation Analysis

Two key enzymes of the CoA-dependent pathway of vanillin biosynthesis from ferulic acid from *Streptomyces* sp. V-1 were collected from the UniProt database. One is the ferulic acid synthase (Fcs, accession number: S5M744), and the other is the enoyl-CoA hydratase/lyase (Ech, accession number: S5LPF1). Sequence alignments were performed towards Ech, sequences with more than 60 percent homology were retrieved, and the ClusterW [32] algorithm was used to calculate the multiple sequence alignment. The Jalview2.11.2.6 [33] software was applied for the homologous alignments; the sequence log was generated by the online website WebLogo3 [34] (https://weblogo.threeplusone.com (accessed on 14 March 2023)).

4.1.2. Homologous Modeling and Molecular Docking

By retrieving Ech sequence (UniProt: S5LPF1) from Uniprot, Blast results indicated the optimum template for Ech modeling was hydroxycinnamoyl-CoA hydratase lyase (Hchl, PDB ID: 2VSS). Modeller9.23 [35] was utilized to construct the homologous model. By listing the resultant Discrete Optimized Protein Energy (DOPE) value and Molpdf rate, a final 3D model was screened and determined. As the entire active pockets were involved in two subunits, an Ech dimer model was then constructed. The Ech and substrate FCA complex was constructed by AutoDock Vina 1.1.2 [36]. The docking box was set as $23 \times 14 \times 8$ with 0.375 grid spacing. All data were calculated three times to generate the best results.

4.1.3. Molecular Dynamics (MD) Simulation and Binding Free Energy Calculations

All the Molecular dynamics (MD) simulations of the enzyme-substrate complex were performed by AMBER20 [37] package. The force field parameter was set as AMBER-ff14SB [38] for protein, and the general AMBER force field (GAFF) [39] was applied to FCA. The bond angle parameters and resp charge of the small molecule were calculated by Gaussian09 [40]. The topology and coordinate files of the enzyme-substrate complex were obtained by the Tleap module from AMBER20. The whole system was dissolved in a cubic TIP3P water box with periodic boundaries, and Na^+ and Cl^- were used to maintain the electric neutrality of the system. Gradient minimization is employed in the initial system to perform energy minimization so as to eliminate unfavorable contacts and structures. Then, the system would be heated slowly in NVT system by from 0 K to 300 K by a 10 kcal/(mol·Å^2) constraint force, and balanced in NTP system for 300 ps, to keep the temperature and pressure stabilized at 300 K and 1 atm (standard atmosphere). Finally, three independent 100 ns production MD simulations (total of 300 ns) was performed in a well-balanced system with a 2.0 fs time step, and all enzyme-substrate complex systems were calculated under the same parameters and conditions. The particle-mesh Ewald (PME) [41] algorithm were used for system energy minimization and MD simulation process with a cutoff value set at 10 Å, so as to calculate the long-range electrostatic interactions throughout the whole simulation. The SHAKE [42] algorithm was applied to constrain all covalent bonds with hydrogen and to control the temperature change during the MD simulation under Langevin dynamics [43] with a collision frequency γ of 1.0.

The molecular mechanics/generalized Born surface area (MM/GBSA) method [44] was applied in analyzing the molecular interaction between enzyme and substrate. The Python script MMPBSA.py in AMBER20 was used for the calculation of the binding free energies and the per-residue energy decomposition of Ech and FCA. A total of 2500 snapshots were extracted from the last 50 ns of the MD simulation to calculate the binding free energy of MM/GBSA. The most representative conformation for each complex was selected using a clustering algorithm [45]. In addition, an energy decomposition of each residue was performed to evaluate the energy contribution of each residue in the system.

4.2. Experimental Verification

4.2.1. Strains, Plasmids and Reagents

Synthetic *fcs* and *ech* gene sequences were optimized, manufactured and cloned into *E. coli* BL21(DE3) with *BamHI*/*HindIII* and *BglII*/*XhoI* (TaKaRa Bio Inc., Dalian, China). The recombinant stains were used for expression and whole-cell biocatalysts. The rationally designed mutant gene sequences were amplified by PCR with the original *ech* primer pairs (Table 3). All Ech mutants were verified by sequencing services provided from Genscript (Nanjing, China). All reagents used herein are analytical and chromatographic grade.

Table 3. The original Ech primers pairs.

Primers	Primers Sequence (5′→3′)
Bgl II-*ech*-F	GGAAGATCTATGAGCACAGCGGTCGGCAACGG
Xho I-*ech*-R	CCGCTCGAGCTACTTCTCCGGGTCGAAGGCGCTCAGG

4.2.2. Whole Cell Catalysis

40 μL of WT and mutant strains were inoculated into 4 mL LB media supplemented with 50 ug/mL Amp resistance. Cells were cultured at 37 °C overnight with 220 rpm. Then 1% inoculation was transferred into a fresh LB media with the same resistance, and cultured for 4–6 h with the same condition until the OD_{600} accumulated to 0.6~0.8. Next, the final concentration of 0.2 mM of IPTG was supplemented, and cells were cultured at 16 °C for 16 h to induce the enzyme expression. After the induction stage, cells were harvested by centrifugation at 3500 rpm for 10 min, and washed twice with 10 mL sterilized PBS solutions. Finally, 10 mL cells were transferred into the 50 mL conical flask and final

concentration of 5 mM ferulic acid was supplemented, and cultured at 30 °C, 200 rpm for 24 h.

4.2.3. Product Identification and Analysis

The reactions were stopped by adding 20 μL of 6 M HCl, and an equal volume of ethyl acetate was fully mixed for 1 min to perform the extraction. Then, the mixtures were centrifugated at 1200 rpm for 10 min. The supernatant was used for the GC analysis. GC analysis was performed on Varian 456 Gas chromatograph equipped with HP-5 capillary column (30 m × 0.320 mm × 0.25 μm). The detection conditions were set 2 min at 140 °C, then increased to 220 °C at 15 °C/min and kept at 220 °C for 2 min. The injector temperature was set at 250 °C, the FID detector temperature was set at 280 °C, and the injection volume was 5 μL.

5. Conclusions

In this study, Ech was rationally designed based on sequence and structural features, and a series of mutants, including F74W, A130G, A130G/T132S, R147Q, Q255R, ΔT90, ΔN1-11, ΔC260-287, and ΔTGPEIL, were obtained, among which the mutants F74W, R147Q, and ΔN1-11 showed an approximately 2-fold increase in activity compared with the WT. The MD simulation results indicate that the basic residues K37, R38, K561, and R564 play critical roles in the binding of Ech and FCA. Therefore maintaining the stability of these residues is essential for the catalytic activity of the enzyme. In addition, the significant changes in the electrostatic interaction energy and polar solvation energy caused by the mutations could further reduce the enzyme activity. Our study provides valuable theoretical support and a reliable basis for the rational design of enzymes and the industrialization of biosynthesized vanillin. In the future, we will further improve and explore the molecular mechanisms affecting the catalytic activity of Ech to enhance the catalytic activity of Ech and thus promote the development of the vanillin industry.

References

1. Anuradha, K.; Shyamala, B.N.; Naidu, M.M. Vanilla—Its Science of Cultivation, Curing, Chemistry, and Nutraceutical Properties. *Crit. Rev. Food Sci. Nutr.* **2013**, *53*, 1250–1276. [CrossRef] [PubMed]
2. Zhu, Y.; Liu, J.; Liao, Y.; Lv, W.; Ma, L.; Wang, C. Degradation of Vanillin During Lignin Valorization Under Alkaline Oxidation. *Top. Curr. Chem.* **2018**, *376*, 29. [CrossRef]
3. Martău, G.A.; Călinoiu, L.-F.; Vodnar, D.C. Bio-vanillin: Towards a sustainable industrial production. *Trends Food Sci. Technol.* **2021**, *109*, 579–592. [CrossRef]

4.   García-Bofill, M.; Sutton, P.W.; Straatman, H.; Brummund, J.; Schürmann, M.; Guillén, M.; Álvaro, G. Biocatalytic synthesis of vanillin by an immobilised eugenol oxidase: High biocatalyst yield by enzyme recycling. *Appl. Catal. A Gen.* **2020**, *610*, 117934. [CrossRef]

5.   Paul, V.; Rai, D.C.; Ramyaa Lakshmi, T.S.; Srivastava, S.K.; Tripathi, A.D. A comprehensive review on vanillin: Its microbial synthesis, isolation and recovery. *Food Biotechnol.* **2021**, *35*, 22–49. [CrossRef]

6.   Tadasa, K. Degradation of Eugenol by a Microorganism. *Agric. Biol. Chem.* **1977**, *41*, 925–929. [CrossRef]

7.   Lone, B.A.; Bhushan, A.; Ganjoo, A.; Katoch, M.; Gairola, S.; Gupta, P.; Babu, V. Biotransformation of eugenol by an endophytic fungus *Daldinia* sp. IIIMF4010 isolated from *Rosmarinus officinalis*. *Nat. Prod. Res.* **2023**, *37*, 535–541. [CrossRef]

8.   Ashengroph, M.; Nahvi, I.; Zarkesh-Esfahani, H.; Momenbeik, F. Conversion of Isoeugenol to Vanillin by *Psychrobacter* sp. Strain CSW4. *Appl. Biochem. Biotechnol.* **2012**, *166*, 1–12. [CrossRef]

9.   Ashengroph, M.; Nahvi, I.; Zarkesh-Esfahani, H.; Momenbeik, F. Pseudomonas resinovorans SPR1, a newly isolated strain with potential of transforming eugenol to vanillin and vanillic acid. *New Biotechnol.* **2011**, *28*, 656–664. [CrossRef]

10. Plaggenborg, R.; Overhage, J.; Loos, A.; Archer, J.A.C.; Lessard, P.; Sinskey, A.J.; Steinbüchel, A.; Priefert, H. Potential of *Rhodococcus* strains for biotechnological vanillin production from ferulic acid and eugenol. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 745–755. [CrossRef]

11. Karakaya, H.; Yilmaztekin, M. Natural Vanillin Production from Isoeugenol by Using *Pseudomonas putida* in Biphasic Bioconversion Medium. *Tarım Bilim. Derg.* **2021**, *28*, 423–429. [CrossRef]

12. Ashengroph, M.; Nahvi, I.; Zarkesh-Esfahani, H.; Momenbeik, F. Use of Growing Cells of *Pseudomonas aeruginosa* for Synthesis of the Natural Vanillin via Conversion of Isoeugenol. *Iran. J. Pharm. Res.* **2011**, *10*, 749–757.

13. García-Bofill, M.; Sutton, P.W.; Guillén, M.; Álvaro, G. Enzymatic synthesis of vanillin catalysed by an eugenol oxidase. *Appl. Catal. A Gen.* **2019**, *582*, 117117. [CrossRef]

14. Liu, H.; Zou, Y.; Yao, C.; Yang, Z. Enzymatic synthesis of vanillin and related catalytic mechanism. *Flavour Fragr. J.* **2020**, *35*, 51–58. [CrossRef]

15. Ni, J.; Wu, Y.-T.; Tao, F.; Peng, Y.; Xu, P. A Coenzyme-Free Biocatalyst for the Value-Added Utilization of Lignin-Derived Aromatics. *J. Am. Chem. Soc.* **2018**, *140*, 16001–16005. [CrossRef]

16. Gasson, M.J.; Kitamura, Y.; McLauchlan, W.R.; Narbad, A.; Parr, A.J.; Parsons, E.L.H.; Payne, J.; Rhodes, M.J.C.; Walton, N.J. Metabolism of Ferulic Acid to Vanillin. *J. Biol. Chem.* **1998**, *273*, 4163–4170. [CrossRef] [PubMed]

17. Overhage, J.; Priefert, H.; Steinbüchel, A. Biochemical and Genetic Analyses of Ferulic Acid Catabolism in *Pseudomonas* sp. Strain HR199. *Appl. Environ. Microbiol.* **1999**, *65*, 4837–4847. [CrossRef] [PubMed]

18. Yoon, S.-H.; Li, C.; Lee, Y.-M.; Lee, S.-H.; Kim, S.-H.; Choi, M.-S.; Seo, W.-T.; Yang, J.-K.; Kim, J.-Y.; Kim, S.-W. Production of vanillin from ferulic acid using recombinant strains of *Escherichia coli*. *Biotechnol. Bioprocess Eng.* **2005**, *10*, 378–384. [CrossRef]

19. Graf, N.; Altenbuchner, J. Genetic engineering of Pseudomonas putida KT2440 for rapid and high-yield production of vanillin from ferulic acid. *Appl. Microbiol. Biotechnol.* **2013**, *98*, 137–149. [CrossRef]

20. Di Gioia, D.; Luziatelli, F.; Negroni, A.; Ficca, A.G.; Fava, F.; Ruzzi, M. Metabolic engineering of *Pseudomonas fluorescens* for the production of vanillin from ferulic acid. *J. Biotechnol.* **2011**, *156*, 309–316. [CrossRef]

21. Kaur, B.; Chakraborty, D.; Kumar, B. Metabolic engineering of *Pediococcus acidilactici* BD16 for production of vanillin through ferulic acid catabolic pathway and process optimization using response surface methodology. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 8539–8551. [CrossRef]

22. Liu, X.; Kong, J.-Q. Steroids hydroxylation catalyzed by the monooxygenase mutant 139-3 from *Bacillus megaterium* BM3. *Acta Pharm. Sin. B* **2017**, *7*, 510–516. [CrossRef]

23. Chen, P.; Yan, L.; Wu, Z.; Li, S.; Bai, Z.; Yan, X.; Wang, N.; Liang, N.; Li, H. A microbial transformation using *Bacillus subtilis* B7-S to produce natural vanillin from ferulic acid. *Sci. Rep.* **2016**, *6*, 20400. [CrossRef]

24. Yang, W.; Tang, H.; Ni, J.; Wu, Q.; Hua, D.; Tao, F.; Xu, P. Characterization of Two Streptomyces Enzymes that Convert Ferulic Acid to Vanillin. *PLoS ONE* **2013**, *8*, e67339. [CrossRef] [PubMed]

25. Fleige, C.; Meyer, F.; Steinbüchel, A. Metabolic Engineering of the Actinomycete *Amycolatopsis* sp. Strain ATCC 39116 towards Enhanced Production of Natural Vanillin. *Appl. Environ. Microbiol.* **2016**, *82*, 3410–3419. [CrossRef]

26. Arnold, F.H.; Volkov, A.A. Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **1999**, *3*, 54–59. [CrossRef] [PubMed]

27. Kuchner, O.; Arnold, F.H. Directed evolution of enzyme catalysts. *Trends Biotechnol.* **1997**, *15*, 523–530. [CrossRef] [PubMed]

28. Wang, Y.; Xue, P.; Cao, M.; Yu, T.; Lane, S.T.; Zhao, H. Directed Evolution: Methodologies and Applications. *Chem. Rev.* **2021**, *121*, 12384–12444. [CrossRef]

29. Yadav, S.; Pandey, A.K.; Dubey, S.K. Molecular modeling, docking and simulation dynamics of β-glucosidase reveals high-efficiency, thermo-stable, glucose tolerant enzyme in *Paenibacillus lautus* BHU3 strain. *Int. J. Biol. Macromol.* **2021**, *168*, 371–382. [CrossRef]

30. Huang, P.; Lv, A.; Yan, Q.; Jiang, Z.; Yang, S. The structure and molecular dynamics of prolyl oligopeptidase from *Microbulbifer arenaceous* provide insights into catalytic and regulatory mechanisms. *Acta Crystallogr. D Struct. Biol.* **2022**, *78*, 735–751. [CrossRef] [PubMed]

31. Taidi, L.; Maurady, A.; Britel, M.R. Molecular docking study and molecular dynamic simulation of human cyclooxygenase-2 (COX-2) with selected eutypoids. *J. Biomol. Struct. Dyn.* **2022**, *40*, 1189–1204. [CrossRef] [PubMed]

32. Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr. Protoc. Bioinform.* **2003**, *2*, 2–3. [CrossRef] [PubMed]

33. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef]

34. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef] [PubMed]

35. Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinform.* **2016**, *54*, 5.6.1–5.6.37. [CrossRef] [PubMed]

36. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. [CrossRef]

37. Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An overview of the Amber Biomolecular Simulation Package: Amber Biomolecular Simulation Package. *WIREs Comput. Mol. Sci.* **2013**, *3*, 198–210. [CrossRef]

38. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef]

39. Sprenger, K.G.; Jaeger, V.W.; Pfaendtner, J. The General AMBER Force Field (GAFF) Can Accurately Predict Thermodynamic and Transport Properties of Many Ionic Liquids. *J. Phys. Chem. B* **2015**, *119*, 5882–5895. [CrossRef]

40. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.A.; et al. *Gaussian 09*; Revision E.01; Gaussian, Inc.: Wallingford, CT, USA, 2009.

41. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

42. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]

43. Izaguirre, J.A.; Catarello, D.P.; Wozniak, J.M.; Skeel, R.D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **2001**, *114*, 2090–2098. [CrossRef]

44. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82. [CrossRef] [PubMed]

45. Shao, J.; Tanner, S.W.; Thompson, N.; Cheatham, T.E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334. [CrossRef]