Evolutionary coupling-inspired engineering of alcohol dehydrogenase reveals the influence of distant sites on its catalytic efficiency for stereospecific synthesis of chiral alcohols

Jie Gu, Byu Ri Sim, Jiawen Li, Yangqing Yu, Lei Qin, Lunjie Wu, Yu Shen, Yao Nie, Yi-Lei Zhao, Yan Xu

Lab of Brewing Microbiology and Applied Enzymology, School of Biotechnology and Key Laboratory of Industrial Biotechnology of Ministry of Education, Jiangnan University, Wuxi 214122, China

State Key Laboratory of Food Science and Technology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, MOE-LSB & MOE-LSC, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

Keywords:
Evolutionary coupling
Distant sites
Saturation mutation
Alcohol dehydrogenases
Catalytic efficiency
Chiral alcohols

Abstract
Alcohol dehydrogenase (ADH) has attracted much attention due to its ability to catalyze the synthesis of important chiral alcohol pharmaceutical intermediates with high stereoselectivity. ADH protein engineering efforts have generally focused on reshaping the substrate-binding pocket. However, distant sites outside the pocket may also affect its activity, although the underlying molecular mechanism remains unclear. The current study aimed to apply evolutionary coupling-inspired engineering to the ADH CpRCR and to identify potential mutation sites. Through conservative analysis, phylogenetic analysis and residues distribution analysis, the co-evolution hotspots Leu34 and Leu137 were confirmed to be highly evolved under the pressure of natural selection and to be possibly related to the catalytic function of the protein. Hence, Leu34 and Leu137, far away from the active center, were selected for mutation. The generated CpRCR-L34A and CpRCR-L137V variants showed high stereoselectivity and 1.24–7.81 fold increase in k_cat/K_m value compared with that of the wild type, when reacted with 8 aromatic ketones or α-ketoesters. Corresponding computational study implied that L34 and L137 may extend allosteric fluctuation in the protein structure from the distal mutational site to the active site. Moreover, the L34 and L137 mutations modified the pre-reaction state in multiple ways, in terms of position of the hydride with respect to the target carbonyl. These findings provide insights into the catalytic mechanism of the enzyme and facilitate its regulation from the perspective of the site interaction network.

1. Introduction
Alcohol dehydrogenases (ADHs; EC 1.1.1.1–423) catalyze the reduction of prochiral ketones to chiral alcohol compounds with high stereoselectivity [1–4]. Compounds such as (S)-tert-butyl-3-hydroxybutanoate (1b), (R)-1-phenyl ethanol (5b) and (S)-1-phenyl-1,2-ethanediol (6b), shown in Scheme 1, are widely used in the preparation of high value-added products and in the synthesis of pharmaceutical intermediates with anti-tumor, anesthetic, sedative, and anti-inflammatory effects [5–10]. However, due to the low catalytic efficiency and narrow substrate spectrum of natural ADH [11], their application remains limited. Therefore, protein engineering of the enzyme was considered [12–14].

Currently, directed evolution is a kind of protein engineering that allows proteins to evolve in the desired direction by simulating natural selection. The required protein properties are obtained through the commonly used methods error-prone PCR and DNA
shuffling coupled with a large number of screenings [15–18]. However, the associated workload is quite heavy due to random mutagenesis. The rational design is based on the analysis of protein structure information and catalytic mechanism to determine the key sites or regions that affect the characteristics and modify the protein [19–21]. Reetz [57] proposed combinational active-site saturation test (CAST), in which pairs of residues are selected for saturation mutation in the active center of the enzyme, which can change the characteristics of the enzyme more effectively than single-point saturation mutation. Sun [22–23] et al. improved the enantioselectivity of ADH using the three-code saturation mutagenesis (TCSM) method. Iterative saturation mutagenesis (ISM) was introduced on the basis of preliminary screening to further optimize the screening results [17]. Yu [20] et al. utilized the difference in enzyme/substrate-binding mode between restricted and free molecular dynamics to find mutation sites suitable for reshaping the substrate pocket and, subsequently, improving catalytic efficiency of ADH. However, both rational and semi-rational design remain mostly limited to the catalytic center, whereas there could be innumerable functional elements in the distal sites outside the functional area. Distal effect has been confirmed to influence remote sites; however, method of identifying the latter in the desired enzyme and the mechanism underlying the remote effects are still unclear [24–28]. If the selection of mutation hotspots can be extended to the entire protein molecule while keeping the mutation library limited to a relatively small number of residues, new functional sites may be discovered. Co-evolution is a hot topic in current research; it can predict the residue pairs that are highly likely to undergo co-variation under natural pressure by analyzing the results of multiple sequence alignments (MSA). The co-variating residue pairs are in contact with each other in the 3D structure and are related to the structure and function of the enzyme molecule [29–32]. Co-evolutionary analysis of protein sequence and calculation of evolutionary coupling are widely used in protein 3D structure prediction [33–36]. Moreover, Wang [37] et al. showed that the structural correlation of evolutionary coupled residue pairs (ECs) is incorporated into functional correlation as well.

In order to improve the catalytic efficiency of the stereospecific ADH from Candida parapsilosis (CpRCR) toward various prochiral carbonyl compounds and broaden the enzyme’s substrate spectrum, Nie [4] et al. reshaped large and small catalytic pockets according to the crystal structure and increased conformational flexibility of the active site. However, this study was limited to the catalytic pocket only, and catalytic effect on some substrates was not obvious. In the present study, we aimed to efficiently identify new functional sites throughout the protein, not being confined to the catalytic pocket, in order to improve the catalytic efficiency and maintain high stereoselectivity. We analyzed the full-length sequence of CpRCR, and obtained residue pairs with strong evolutionary coupling as potential mutation sites. Double-site saturation mutagenesis, employing merge codon NNK and single-site saturation mutations, was conducted to obtain mutants with improved catalytic efficiency. By characterizing the specific activity, kinetic parameters, and stereoselectivity toward aromatic ketones and β-ketoesters, CpRCR variants with enhanced catalytic efficiency for stereospecific synthesis of chiral alcohols were identified.

2. Materials and methods

2.1. Materials

DNA Polymerase and other reagent kits for DNA cloning and amplification were purchased from Takara-Bio Co. (Japan). Oligonucleotides were synthesized by Sangon Biotech (Shanghai, China). Substrates 1a–8a, NADH and (R)-[(S)-1-bromo-2-methylbutane] were ordered from Sigma-Aldrich (USA). Chromatography grade hexane and isopropanol used in high performance liquid chromatography (HPLC) were purchased from Aladdin Co. (China).

2.2. Co-evolution analysis

The ADH CpRCR was derived from Candida parapsilosis (GenBank: DQ295067.1) with a total of 336 residues, which is characteristic of the medium-chain dehydrogenase family. The full sequence was entered into the EVCouplings server (http://evfold.org/); the program utilized the jackHMMER algorithm to search the UniProt database for generating MSA. Alternative pseudolikelihood maximization direct coupling analysis (PLM) was subsequently performed to evaluate the above results with default parameters.

2.3. Evolutionary conservation analysis

The ConSurf algorithm was used to estimate the evolutionary conservation of CpRCR. ConSurf performs accurate computation of the evolutionary rate using an empirical Bayesian method [38]. The conservation score, which is continuous from 1 to 9, calculated by ConSurf, is a relative measure of evolutionary conservation at each residue. Evolutionary conservation analysis was performed on the web server: https://consurf.tau.ac.il/.

2.4. Primer design and saturation mutagenesis library construction

The recombinant plasmid pET-32Xa/LIC-CpCR (WT) was constructed previously [4]. The mutagenesis primer shown in Table S1 was used with the WT CpCR gene sequence as template. The PCR product was digested with DpnI eliminate any residual template. The variants obtained in this work were all constructed using ConExpress II One Step Cloning Kit from Vazyme. After homologous recombination in vitro, circular plasmids were directly introduced into E. coli BL21 (DE3) competent cells. Single colonies of E. coli on the plate were randomly selected after 8-h culture, and their plasmids were extracted and sequenced to confirm successful construction of the saturation mutation library.
2.5. Gene expression and enzyme purification

A single transformant was cultured in LB medium including ampicillin (100 μg·mL⁻¹) and 0.2 mM zinc acetate at 37 °C and 200 rpm until optical density of the culture reached 0.6–0.8. The protein was expressed without induction by isopropyl-β-D-thiogalactopyranoside at 30 °C for 8 h. After harvested, the cells were resuspended in Tris- HCl (pH 8.0) buffer with 25 mM Tris- HCl, 150 mM NaCl, and 20 mM imidazole. At 4 °C, the cells were lysed by sonication and centrifuged to collect the supernatant. The target protein was purified by affinity chromatography using His-Tag. The sample was loaded with 2.0 mL/min. The protein with His-tag was bound to the column and eluted with approximately 300 mM imidazole. Factor Xa was used to remove the fusion protein TrxA. CpRCR which was not labeled with His-tag was eluted by the buffer without imidazole in the second purification using HisTrap HP column. The purified protein was verified to be a single band in SDS-PAGE and thereafter considered for subsequent experiments.

2.6. Screening of positive variants

All the strains were inoculated into 96-well plates and cultivated as mentioned above. After the cells were broken by lysozyme, crude enzyme activity was measured by the change of coenzyme NADH absorbance at 340 nm. The standard assay was performed at 30 °C in a reaction mixture of 100 μl, composed of 100 mM sodium phosphate buffer, pH 7.0, 5 mM substrate, and 0.5 mM NADH, together with appropriate amounts of the crude or purified enzymes. One unit of enzyme activity was defined as the amount of enzyme required to oxidize 1 μmol NADH per min under the standard assay conditions [4].

2.7. Determination of kinetic parameters

The standard assay of enzyme activity has been described above. The kinetic parameters were obtained through GraphPad Prism 7 to curve-fit the Michaelis–Menten equation, and the specific activity at different substrate concentrations (0.1–15 mM) was measured. Protein concentrations were determined by measuring UV absorption at 280 nm using the molar extinction coefficients calculated by ExPASy based on the amino acid compositions. All data were averaged over three replicates.

2.8. Phylogenetic analysis

MSA obtained from conservative analysis were used as input data, and MEGA (version 7.0) software was employed to build a phylogenetic tree of dehydrogenase. The neighbor-joining method was used to construct the phylogenetic tree, and the bootstrap method was performed to detect the quality of the tree, which was set to 1000. According to the Uniprot database (http://www.uniprot.org), the source of each gene was listed.

2.9. Stereoselectivity assay

Enzymatic reduction was performed by adding 10 mM substrate, 10 mM NADH, 10% isoamyl alcohol, and proper amount of enzyme to 100 mM phosphate buffer (pH 7.0), and then allowing them to react at 30 °C and 200 rpm for 8 h. After the reaction mixture was extracted with ethyl acetate, the organic layer was filtered through a 0.22-μm PVDF syringe filter (Sangon Biotech, China) for analysis. The specific peak of (R)/(-)-alcohol was detected by HPLC or GC [4]. The details of chiral HPLC or GC analysis are shown in Table S2.

2.10. Molecular dynamic (MD) simulations of pre-reaction state (PRS)

All MD simulations were performed using the Amber Suite [39]. The initial protein structure used in the molecular dynamic simulation was constructed based on the crystal structure (PDB: 3WLE). The enzyme-substrate complexation was referred to for the catalytic mechanisms of the zinc-dependent medium-chain dehydrogenase/reductase [40]. Appropriate substrate conformations were selected for multiple MD simulations. Protonation states of titratable residues were assigned at pH 6.5 using the H+ web server [41], and Zn/substrate-binding residues were visually inspected. Particularly, the zinc interacting cystine residues (C44, C95, C98, C101, and C104) were manually deprotonated, and the zinc interacting histidine residue (H65) was switched to HID. The restrained electrostatic potential (RESP) charge was fitted on the substrate at the level of B3LYP/6-31G* using the antechamber program and Gaussian09 software [42]. The force field of Zn was treated in the python-based Metal Center Parameter Builder [43]. The NADH force field parameters by walker were applied in the MD simulations [58].

Topology and coordinate files for the enzyme-substrate complexes were prepared using the tleap module, and solvation was performed in a cubic box of TIP3P water molecules with a water thickness of 10 Å from the protein surface. The systems were neutralized by adding chloride ions. MD simulations were conducted by the PMEMD.cuda program. Long-range electrostatic interactions were calculated using the particle mesh Ewald method. Lengths of bonds involving hydrogen atoms were fixed by the SHAKE algorithm. Periodic boundary conditions were used to simulate the effects of a larger system. Minimization was carried out in two steps to correct any possibly unrealistic arrangements: the first step only involved the water molecules relaxation; the second step involved minimization of the whole system. Langevin dynamics with a collision frequency of 2 ps⁻¹ was applied to gradually increase the system temperature from 0 to 300 K over 100 ps. All MD-simulations were conducted with the non-bonded cut-off limit set to 10 Å. Before generating simulations, the system was subjected to a 50-ps equilibration to adjust the density under constant pressure and temperature (NPT), followed by a second 5-ns equilibration step. During the 10-ns PRS simulation, distances of the migrating hydride of NADH and proton of Ser46 with the substrate carbonyl were constrained within a range of 1.0–3.0 Å to enhance the PRS sampling. Finally, 100-ns simulations with no restraint were conducted under NPT conditions (Fig. S8). A 2-fs integration time step was utilized, with structural snapshots being extracted every 1000 steps. The simulation trajectory was analyzed by cpptraj in Amber Tools 18.

3. Results and discussion

3.1. Prediction of evolutionary couplings

After entering the full-length 336 amino acids of CpRCR into EVcouplings, JackHMMER was used to search the UniProt database to generate multiple sequence alignment results. Since the CpRCR sequence had fewer than 600 amino acids, accuracy of the derivation results of the maximum entropy model for evaluating MSA with PLM was higher than that with DI (mean-field direct coupling analysis).

EVcouplings found 113,514 homologous alignment sequences in the database and generated pairs of ECs; thirteen pairs of ECs scored higher than 1.2 (Table 1), which was higher than the significance threshold of 0.8 suggested by Hopf [44]. Amino acid residues with high scores were considered to be the hotspots of evolution under natural selection and to have a significant impact.
on protein function. Moreover, Lee [45] found the mutation effect of conserved positions during evolution to be more obvious than that of non-conserved positions. However, the majority of protein engineering efforts to date have focused on non-conserved sites [46–47]. Therefore, conserved sites in these co-variant residue pairs would be significant to the catalytic function of the enzyme and be the “hotspot” for further investigation of their influence on catalytic activity and stereoselectivity.

3.2. Conservation analysis

Conservation analysis of all amino acid residues were performed based on evolutionary conservation scores calculated using ConSurf. Among all the ECs shown in the table above, L34 and L137 had the highest conserved scores of 8 and 7 (the maximum score being 9), respectively. The highest score represents the most conserved position in a protein. It does not necessarily indicate 100% conservation (meaning no mutation at all) in the case of L34 and L137. However, the positions have co-evolved compared to other sites of the enzyme. In MSA, the residue at position 34 was either serine, isoleucine, valine, alanine, or leucine while that at position 137 was either cysteine, valine, isoleucine, alanine, methionine, or leucine (Fig. 1D–E). The substitutions at these two positions mainly focused on uncharged amino acids, mostly nonpolar ones, under natural evolutionary pressure.

After imposing the amino acid conservation degree onto the spatial structure of CpRCR (Fig. 1B), it was more obvious that although L34 and L137 are far apart in sequence, they are very close spatially. The distance of 3.6 Å implied that they were indeed close in contact with each other (Fig. 1C). Therefore, the positions have co-evolved in response to the pressure exerted by nature. In order to explore the influence of these two sites on protein function, saturation mutagenesis at both the residues was further conducted.

3.3. Saturation mutagenesis at both residues

Tert-butyl acetoacetate (1a) was chosen as the model substrate for screening, since its chiral product (S)-tert-butyl-3-hydroxybutyrate ((S)-1b) is a raw material for the synthesis of cryptocarya triacetate [48], and Sam [49] reported that it has physiological activity, not only in treating headaches and morning sickness, but also against cancer and lung diseases. Based on previous studies, the catalytic activity of CpRCR is poor and needs further improvement.

As shown in Fig. 2, approximately 3072 mutants at L34/L137, which are far away from the catalytic triad, were screened, covering 95% probability [50]. Expectedly, the enzyme activities of mutants differed from those of the wild type. Most mutations led to a decrease in crude enzyme activity, and some mutations were even severely inactivated. Only about 5% of the mutations have higher enzyme activity than the wild type. Therefore, these two sites would have the impact on protein function. After sequencing all the mutants (at the position L34/L137) with increased enzyme activity, an interesting phenomenon was observed. All mutants with increased enzyme activity in this library were single-point mutations (Fig. S1). In addition, the positive mutants (CpRCR-L34A, CpRCR-L34I, CpRCR-L137M, CpRCR-L137T, and CpRCR-L137V) were obtained by replacing the original residues with specific ones. Another interesting point is that the substituted residues of positive mutants are mainly concentrated on a small part of the special nonpolar amino acids.

Honda [51] discovered that conserved residues are essential for protein function and that some substitutions exert serious effects on the NADP-malic enzyme activity. However, amino acid substitutions with similar properties can maintain the protein function. By combining evolutionary coupling prediction and conservation analysis, L34 and L137 were presumed to have strong synergistic relationship and were considered to be important functional sites of CpRCR; thus catalytic activity of the enzyme can only be maintained when one of these two sites was mutated to several specific amino acid residues. In order to further confirm the presumption, we performed single-point saturation mutagenesis at each of these two residues.

3.4. Single-point saturation mutagenesis at L34 or L137

Substrate 1a as the model substrate was still applied to scan 38 saturated mutants at two sites. All the protein was purified by affinity chromatography and verified by SDS-PAGE. At 5 mM substrate concentration, the specific activity of all mutants was shown in Fig. S3. The results were similar to the primary screening, with most of the single mutants showing a decrease in specific activity compared to the wild type. Further more, all charged residue mutants except L137D were inactivated. Not all non-polar amino acids were active either, e.g., L34W, L34F, which may also result in the 34 site being more conserved than the 137. None of the amino acid substitutions present in the conservation analysis above showed a significant decrease in specific activity (all less than 2-fold).

Subsequently, two classes of typical carbonyl compounds β-ketoesters (1a, 2a, 3a, and 4a) aryl ketones (5a and 6a) and were selected to scan the substrate spectrum of purified CpRCR and its 10 positive mutants at L34 or L137 (Fig. 3). The results were consistent with screening of positive variants with double mutation and conservation analysis. Almost all single-site mutants formed by replacement with specific amino acid residues, based on conservation analysis, performed better catalytic activity for one or more substrates. Among them, the mutants CpRCR-L34A and CpRCR-L137V exhibited the most significant improvement in catalytic activity. These two amino acids are also highlighted in Fig. 1D–E. The specific activity of CpRCR-L34A toward 1a, 2a, 3a, 4a, 5a, 6a was 3.57, 3.74, 3.64, 1.65, 2.30 and 5.50-fold increased than that of the wild type; the specific activity of CpRCR-L137V for 3a and 6a was 3.60 times and 4.75 times increased over that of the wild type, respectively. All the mutations in Fig. 3 showed the effect in improving the catalytic activity of the enzyme toward the aromatic substrate 6a. Specific activity data of all positive mutants are presented in Table S3.

Honda [51] discovered that conserved residues are essential for protein function and that some substitutions exert serious effects on the NADP-malic enzyme activity. However, amino acid substitutions with similar properties can maintain the protein function. By combining evolutionary coupling prediction and conservation analysis, L34 and L137 were presumed to have strong synergistic relationship and were considered to be important functional sites of CpRCR; thus catalytic activity of the enzyme can only be maintained when one of these two sites was mutated to several specific amino acid residues. In order to further confirm the presumption, we performed single-point saturation mutagenesis at each of these two residues.
tion of residues at any point of these two could lead to conspicuous changes in the catalytic ability of the protein, and a large number of negative mutants formed by two-point saturation mutations and inactivation of the combination of dominant mutants CpRCR-L34A/L137V implied that these two points have a profound influence on each other, further confirming the reliability of evolutionary coupling analysis. Consistently, Aakre [53] discovered that mutations in the second position of co-evolved residues could compensate for the decreased fitness caused by the mutation in the first position. Therefore, we reasonably speculated that when the amino acid residue at position L34 becomes alanine, leucine at position L137 is the most suitable compensatory residue; on the other hand, when the amino acid residue at position L137 becomes valine, leucine at position L34 restored the adaptability.

3.5. Phylogenetic analysis and residues distribution analysis

In order to further reveal why combined mutations lead to the decreased protein activity, we performed phylogenetic analysis and residues distribution analysis on the corresponding position of 34 and 137 in the homologous sequence obtained from the previous conservation analysis.

According to the results of MSA (Fig. 1D–E), there are 5 and 6 possible amino acid residues at position 34 and 137, respectively, and thus there should theoretically be 30 combinations at these two positions. By analyzing these homologous sequences, however, only 15 combinations can be observed for these two sites, and the frequency of V34/L137, L34/L137, and I34/L137 was 66.7%, 10%, and 5.3%, respectively, as shown in Fig. 4. These three combinations account for 82% of all cases. Therefore, during the evolution at these two sites, residues may be not in one-to-one correspondence, but have a fixed collocation pattern. Leucine may be an all-matched residue at these two positions. In addition, for dehydrogenases derived from bacteria, the residue pair at positions 34 and 137 generally prefers to V34/L137; while the dehydrogenases isolated from eukaryotes exhibit a broader distribution of amino acid residues at these sites. Interestingly, almost all dehydrogenases derived from Candida possess this residue pair as V34/L137 or L34/L137. Thus, these two sites exhibit significant particularity in phylogenetic analysis and residues distribution analysis. These observations also help us for further understanding why only single mutations at each site showed better performance when saturation mutating at predicted correlated pairs L34/L137.

According to the results in Fig. 3, the mutants L34V/L137 and L34I/L137 with the highest probability of occurrence displayed a higher activity towards 6 substrates (L34/L137 is the wild type).
substrate 1a were characterized (Fig. S6). The specific activity of these double mutations did not differ from that of the wild type. Only the specific activity of L34I/L137V was slightly increased, while all other mutations were slightly reduced relative to the wild-type. The double mutants maintained their protein catalytic function, which may explain why they were not eliminated during natural evolution. Although the coevolution analysis is based on the natural function, we believe that these evolutions under natural pressure are still instructive for our research on the enzymes catalyzing stereospecific reactions of unnatural carbonyl substrates, because natural evolution can eliminate residues that have a major impact on protein folding and activity [56].

The residues distribution of all sites was shown in Fig. S5. The hydrophobicity and conservation of residues at 34 and 137 were evident compared to the other sites. There are other sites such as 81, 131, 134, etc. that have evolutionary coupling with 34 or 137 and close to each other in the protein spatial structure. According to the results of the co-evolution analysis, the scores of these sites coupled with L34 and L137 were lower than L34/L137 pairs and below 0.8 which is the threshold recommended by Hopf [44]. The residues at these sites were widely distributed and not concentrated in a particular class of amino acids with specific properties. Therefore, the compensating effect of other co-evolution site may not be obvious.

### 3.6. Characterization of CpRCR-L34A and CpRCR-L137V

In addition to the above-mentioned substrates, 7a and 8a, which are difficult-to-reduce substrates, were used to evaluate the effect of these mutations. As shown in Fig. 5, the kinetic parameters of CpRCR-L34A and CpRCR-L137V for each substrate showed the same trend as their characterized specific activities. The $k_{cat}$ and $k_{cat}/K_m$ values of the two mutants were improved to a certain extent compared with those of the wild type, and modification effect of CpRCR-L34A was higher than that of CpRCR-L137V. Detailed $k_{cat}$ and $k_{cat}/K_m$ values are shown in Table S4.

For 2a, 3a, 5a and 6a, the favorable substrates of CpRCR [4], $k_{cat}$ of the two mutants were still more than twice of that of the wild type (Fig. 4A). Toward these substrates, CpRCR-L34A displayed up to 3.76, 3.79, 2.23, and 5.69 times higher $k_{cat}$ compared with that of the wild type. The $k_{cat}$ of CpRCR-L137V also increased from 2.06 times to 5.17 times towards these substrates. This result obtained from mutations at distant sites is highly significant, especially compared with the previous work on structure-based rational design at the enzyme catalytic pocket [4]. Although structure-based site-directed mutagenesis expanded the overall substrate spectrum, for some substrates such as 2a, 3a 5a and 6a which the $k_{cat}$ of wild-type is already higher than the $6 \text{ s}^{-1}$, the positive mutation effect was not obvious. Even most mutants lead to a reduction in catalytic efficiency. Therefore, not only the increased fold is an important factor in judging the effect of mutation, but the value of $k_{cat}$ and $k_{cat}/K_m$ should also be taken into account. It is more difficult to modify an enzyme with high original catalytic efficiency than that with poor catalytic capacity. This study could improve the catalytic efficiency of the enzyme toward these substrates through remote-site mutation, which further indicated the functional contribution of these distal sites.

For 1a, 4a, 7a, and 8a, $k_{cat}$ of the two mutants showed a certain degree of improvement (Fig. 5A). In addition, increased affinity of the mutants toward 1a, 4a and 8a was noticeable. Especially toward 8a, $K_m$ of the mutant was 0.36 while that of the wild type CpRCR was 1.09, and the ratio of $k_{cat}/K_m$ of CpRCR-L34A was 7.81 while that of CpRCR-L137V was 5.72 (Fig. 5B).

Since active-site engineering may well affect the stereoselectivity of ketone reduction and, thereby, the enantiomeric purity of alcohols [54], we suspected that distant sites could also play an important role in stereoselectivity. Therefore, the optical purity of chiral alcohols was measured (see Table 2). Overall, the remote mutations L34A and L137V basically maintained the stereoselectivity of CpRCR. WT and the two positive mutants toward 1a–8a followed the Prelog’s rule in stereoselectivity. In addition, when reducing 1a, 3a, 6a and 7a, WT and the two mutants yielded the products in Prelog configuration with >99% ee. Towards 4a and
Fig. 4. Phylogenic analysis and residues distribution analysis. The legends and probability on the right is the occurrence frequency of amino acids at corresponding position in the homologous sequence. Residues were marked with abbreviations.

Fig. 5. Ratios of $k_{cat}$ (A) and $k_{cat}/K_m$ (B) of mutants (CpRCR-L34A, CpRCR-L137V) to those of WT toward 8 ketones.
Table 2
Optical purity (% ee) of products from asymmetric reduction of 1a–8a, catalyzed by purified CpRcr and mutants.

| Substrate (ee) | CpRcr | CpRcr-L34A | CpRcr-L137V |
|---------------|-------|------------|-------------|
| 1a            | >99 (S) | >99 (S)    | >99 (S)     |
| 2a            | 85 (S)  | 80 (S)     | 84 (S)      |
| 3a            | >99 (R) | >99 (R)    | >99 (R)     |
| 4a            | 82 (R)  | 83 (R)     | 92 (R)      |
| 5a            | 93 (S)  | 95 (S)     | 94 (S)      |
| 6a            | >99 (R) | >99 (R)    | >99 (R)     |
| 7a            | >99 (S) | >99 (S)    | >99 (S)     |
| 8a            | 94 (S)  | 95 (S)     | >99 (S)     |

8a, CpRcr-L137V exhibited slightly higher stereoselectivity than CpRcr, specifically increased from 82% ee to 92% ee and from 94% ee to >99% ee, respectively.

In our previous study, it has been confirmed that CpRcr is an alcohol dehydrogenase with high stereoselectivity towards a series of substrates [4]. Thus, mutants at the sites L34 and L137, which are far from the catalytic pocket, meet the expectations of protein engineering, improving catalytic efficiency while maintaining high stereoselectivity. Evolutionary coupling-inspired protein engineering can efficiently identify remote functional sites with evolutionary potential. It would be an efficient supplement to the conventional catalytic pocket-focused protein engineering.

3.7. Molecular dynamic simulations of pre-reaction state

Structural perturbation at the active site was investigated with the pre-reaction state analysis. To minimize the computational error for hydrogen bonding and metal–ligand coordination in the active pocket, caused by metal center parameter builder (MCPB), we only used the simplest substrate (5a) for comparison, with only carbon and hydrogen atoms in the two substitutions of reactive carbonyl group. Four identical PRS structures, except for the L34 and L137 mutations, were carefully constructed according to the previous QM/MM results in literature [4] and as per crystal structure (PDB: 3WLE). Fig. 6 shows the dynamic structures of pre-reaction states in the wild-type, CpRcr-L34A, CpRcr-L137V, and CpRcr-L34AL137V mutants. All the three mutations increased certain disturbance in the wild-type structure, including either a larger deviation on the hydride-carbon distance or higher distribution on the attacking angle and dihedral angle or both, as per the thermal fluctuation of molecular dynamics simulation. It is in good agreement with the co-evolutionary phenomenon that the two coupling amino acids were optimized in evolution to stabilize the enzyme structure.

The distal mutations changed the relative geometry between substrate and NADH. In L34A, the average bending angle of hydride-carbonyl remained at 81.2–81.5°, although the hydride aligned with the carbonyl perpendicular π* better than that in the wild-type, with a larger dihedral angle (−82.7°) from the original −80.0° in wild type. As a result, the reacting population increased from 12.5% to 14.2%. In CpRcr-L137V, the H---C distance decreased by 0.01 Å on an average, and the highest-probable point of the attacking angle and dihedral angle slightly shifted to that in the pre-reaction state. However, the mean values of attacking angle decreased to 79.6° and the calculated active populations decreased by 1.7%. In the CpRcr-L34A/L137V double mutation, the two angles were calculated to be 73.8° and −75.1°, respectively, far from the pre-reaction state and the other three cases. The CpRcr-L34A/L137V mutation possessed the smallest amount of active population (approximately 5.3%) and the longest hydride-carbon distance of 2.30 Å. This was consistent with the lowest enzyme activity observed for the double mutation, the specific activity was only 33% of the wild type. In addition, we determined the kinetic parameters of the double mutation on substrate 5a. The $k_{cat}$ of the combined mutant was 2.72 s$^{-1}$, while $K_m$ was not significantly different from the wild type at 0.79 mM. Thus, L34 and L137 mutations modified the pre-reaction state in multiple ways in terms of the hydride positions with respect to the target carbonyl, which could either enhance or damage the enzyme activity. However, how the distal mutations could transfer their structural perturbation to the active site still remains unclear.

The Normal mode analysis (NMA) indicates that motions of the Michaelis’ complex could be considered as two parts, corresponding to the coenzyme binding domain and the catalytic domain in the previous reported X-ray structure (PDB: 3WLE), (see Fig. S7) in the low-energy modes, the two individual parts wiggle as an independent community. The substrate-bound catalytic zinc
center lies in the interface of the two parts. The 34 and 137 sites are located at the core part of the catalytic domain during the dynamic movement. Among of them, L34 was more dynamic, with the largest motion compared with other three parts, i.e. L137, NADH, and substrate. The dynamic effect was promoted significantly in L34A for the attenuated vdW interaction between residues 34 and 137. In the low-energy NMA modes, substrate motion was observed to be more essential than that of the cofactor NADH. Because the substrate molecule mainly binds the catalytic zinc via carbonyl oxygen lone pair, the distal mutation of L34/L137 pair inside of a beta-sheet barrel may extend the geometric fluctuation on the substrate via the rigid zinc-coordination of Cys44-His65-Asp154 triad. In addition, previously stated hydrophobic residues of the catalytic domain which contributes to the small and the large substrate binding pocket may constrain the substrate motion via weak interactions as referred by the residues H49, S46, W116, L119, and F285 [54]. Such a quasi-“Newton’s cradle” manner from the distal mutational site to the active site usually reflects allosteric fluctuation in the protein structure [55].

4. Conclusion

In our study, evolutionary coupling-inspired protein engineering was applied to stereospecific ADH CpRKR, which catalyzes the asymmetric synthesis of chiral hydroxyl compounds with high stereoselectivity. In order to identify new functional sites to improve the catalytic efficiency, the high EC score pairs L34 and L137, which were highly conserved compared to other sites, were discovered to significantly affect the catalytic function of the enzyme, although these residues are distant from the active center of the enzyme. In this study, using 8 aromatic ketones and β-ketoesters as model substrates, the mutants obtained from screening the saturation mutagenesis library exhibited significantly enhanced catalytic efficiency compared with the wild type while maintaining the stereoselectivity preference. Mutations at the distal sites showed increased $k_{cat}$ and $k_{cat}/K_m$ values, by 5.69 and 7.81 times, respectively, compared with those of the wild type, and this result was comparable to the results of a structure-based rational design at the substrate-binding pocket [44]. Using computational analysis, the mutations at L34 and L137 could be transferred to the rigid zinc-coordination of Cys44-His65-Asp154 triad through geometric fluctuation. Changes in distribution of the C–H distance, attack angle, and dihedral angle revealed changes in pre-reaction state after the mutation.

The large number of negative results also implied the importance of highly conserved sites for the catalytic function of the protein. Whether in the primary screen for double site saturation mutations, in single site saturation mutations at each site or in combinatorial mutations of dominant mutations, a large number of apparently inactivating mutants were present. The substituted residues of positive mutants were mainly concentrated on uncharged amino acids. Phylogenetic and residue distribution analyses further revealed the co-evolutionary patterns and compensatory effects at L34 and L137.

Overall, evolutionary coupling-inspired protein engineering could help identify new functional sites to systematically study the catalytic mechanism of enzymes and facilitate their regulation from the perspective of site interaction networks.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31872891), the Program of Introducing Talents of Discipline to Universities (111-2-06), the High-End Foreign Experts Recruitment Program (G20190010083), the National Program for Support of Top-notch Young Professionals, the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, Top-notch Academic Programs Project of Jiangsu Higher Education Institutions, the Program for the Key Laboratory of Enzymes of Suqian (M201803), and the National First-Class Discipline Program of Light Industry Technology and Engineering (LITE2018-09).

Author contributions

J.G., J.L., Y.Y., L.Q., L.W., and Y.S. constructed the mutants and performed activity assay and biocatalysis. B.R.S. and Y.-L.Z. performed computational study. Y.N., Y.-L.Z., and Y.X. conceived and designed the experiments. J.G., B.R.S., Y.N., and Y.-L.Z. analyzed the data. J.G., B.R.S., Y.N., and Y.-L.Z. wrote the manuscript with input from coauthors. All authors have given approval to the final version of the manuscript.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.10.031.

References

[1] Wohlgemuth R. Asymmetric biocatalysis with microbial enzymes and cells. Curr Opin Microbiol 2010;13(3):283–92.
[2] Matsuda T, Yamanka R, Nakamura K. Recent progress in biocatalysis for asymmetric oxidation and reduction. Tetrahedron Asymmetry 2009;20(5):513–57.
[3] Musa MM, Philips RS. Recent advances in alcohol dehydrogenase-catalyzed asymmetric production of hydrophobic alcohols. Catal Sci Technol 2011;1(8):1311–23.
[4] Nie Y, Wang S, Xu Y, Luo S, Zhao Y-L, Xiao R, et al. Enzyme engineering based on X-ray structures and kinetic profiling of substrate libraries: alcohol dehydrogenases for stereospecific synthesis of a broad range of chiral alcohols. ACS Catal 2018;8(6):5145–52.
[5] Yamada-Onodera K, Kojima K, Takase Y, Tani Y, Cloning, sequence analysis, and expression in Escherichia Coli of gene encoding n-benzyl-3-pyrrolidinol dehydrogenase from Geotrichum Capitatum. J Biosci Bioeng 2007;104(5):379–84.
[6] Yadav GD, Saigude AD, Dhoott SB. Insight into microwave irradiation and enzyme catalysis in enantioselective resolution of RS(±)-methyl mandelate. J Chem Technol Biotechnol 2008;83(8):1145–53.
[7] Guo JL, Mu QG, Xu Y. Integration of newly isolated biocatalyst and resin-based in situ product removal technique for the asymmetric synthesis of R(±)-methyl mandelate. Bioprocess Biosyst Eng 2010;33(7):797–804.
[8] Yamada-Onodera K, Fukuji M, Tani Y. Purification and characterization of alcohol dehydrogenase reducing n-benzyl-3-pyrrolidinol from Geotrichum Capitatum, J Biosci Bioeng 2007;103(2):174–88.
[9] Chen X, Liu ZQ, Lin CF, Zheng YC. Chemoenzymatic synthesis of (±)-duloxetine using carbonyl reductase from Rhodospirillum Toruloides. Biocor Eng 2016;65:82–9.
[10] Oda S, Kikuchi Y, Nanishi Y. Synthesis of optically active mandelic acid via microbial oxidation of racemic 1-phenyl-1,2-ethanediol. Biosci Biotech Biochem 2014;56(8):1216–20.
[11] Kulig J, Simon RC, Rose CA, Husain SM, Häckl M, Ludeke S, et al. Stereoselective synthesis of bulky 1,2-diols with alcohol dehydrogenases. Catal Sci Technol 2012;2(9):1350–5.
[12] Nealon CM, Musa MM, Patel JM, Phillips RS. Controlling substrate specificity and stereospecificity of alcohol dehydrogenases. ACS Catal 2015;5(4):2190–14.
[13] An J, Nie Y, Xu Y. Structural insights into alcohol dehydrogenases catalyzing asymmetric reductions. Crit Rev Biotechnol 2019;39(3):366–79.

[14] Noey EL, Tiherval N, Jimenez-Oses G, Osuna S, Park J, Bond CM, et al. Origins of stereoselectivity in evolved ketoreductases. Proc Natl Acad Sci USA 2015;112(51):7065–72.

[15] Turner NJ. Directed evolution drives the next generation of biocatalysts. Nat Chem Biol 2009;5(8):567–73.

[16] Denard CA, Ren H, Zhao H. Improving and repurposing biocatalysts via directed evolution. Curr Opin Chem Biol 2015;25:55–64.

[17] Reetz MT. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. Angew Chem Int Edit 2011;50(1):138–74.

[18] Ren C, Wen X, Mencius J, Quan S. Selection and screening strategies in directed evolution to improve protein stability. Bioreor Bioproc Biotechnol 2019;6(1):53–67.

[19] Wang Y, Ying BB, Shun W, Zheng RC, Zheng YG. Rational design of Kluyermonces Manusinos ZBJ14056 Aldo-Keto reductase KmAKR to enhance diastereoselectivity and activity. Enzyme Microb Tech 2017;107:32–40.

[20] Su B-M, Shao Z-H, Li A-P, Naeem M, Lin J, Ye L-D, et al. Rational design of dehydrogenase/reductases based on comparative structural analysis of prevostration and state-free/grame simulations for efficient asymmetric reduction of bulky alky ketones. ACS Catal 2019;10(1):864–76.

[21] Zhou J, Wang Y, Chen X, Xu M, Yang T, Zheng J, et al. Rational engineering of Bacillus cereus leucine dehydrogenase towards alpha-keto acid reduction for improving unnatural amino acid production. Biotechnol J 2019;14(3):e1800253.

[22] Sun Z, Lonsdale R, Ilie A, Li G, Zhou J, Reetz MT. Catalytic asymmetric reduction of difficult-to-reduce ketones: triple-code saturation mutagenesis of an alcohol dehydrogenase. ACS Catal 2016;6(3):1598–605.

[23] Sun Z, Lonsdale R, Wu L, Li G, Li A, Wang J, et al. Structure-guided triple-code saturation mutagenesis: efficient tuning of the stereoselectivity of an epoxide hydrolase. ACS Catal 2016;6(3):1590–7.

[24] Keszef AE, Scherf FM. Mechanism of catalysis, E2 recognition, and autoinhibition for the ipha family of bacterial E3 ubiquitin ligases. Proc Natl Acad Sci USA 2017;114(6):1311–6.

[25] Schuetz, A. K.; Kay, L. E., A Dynamic molecular basis for malfunction in disease mutants of p97/VCP. Elife. 2016, 5, e20143.

[26] Motlaff HN, Wribel JO, Li J, Hilser VJ. The ensemble nature of allostericity. Nature 2014;508(7496):331–9.

[27] Sumbalova L, Stourac J, Martinek T, Bednar D, Damborsky J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. Nucleic Acids Res 2018;46(1):356–62.

[28] Suplatov D, Sharapova Y, Timonina D, Kopylov K, Svedas V. The visualcmat: a web-tool to efficiently guide enzyme engineering. ChemBioChem 2015;16(2):e180023.

[29] Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 2011;108(49):1293–301.

[30] Sulkowska JL, Morcos F, Weigt M, Hwa T, Onuchic JN. Genomics-aided structure prediction. Proc Natl Acad Sci USA 2012;109(26):10340–5.

[31] Hopf TA, Collwell LJ, Sheridan R, Rost B, Sanders C, Marks DS. Three-dimensional structures of membrane proteins from genomic sequence. Cell 2012;149(7):1607–21.

[32] Simkovic F, Ochovnikov S, Baker D, Rigden DJ. Applications of contact predictions to structural biology. IUCrJ 2014;5(10):e3689.

[33] Michel M, Hayat S, Skwark MJ, Sander C, Marks DS, Elphsson A. PconsFold: improved contact predictions improve protein models. Bioinformatics 2014;30(17):482–8.

[34] Ochovnikov S, Kamssetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. Elife 2014;3:e02030.

[35] Ochovnikov S, Kim DE, Wang RY, Liu Y, DiMaio F, Baker D. Improved De Novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. Proteins 2016;84(1):67–75.

[36] Ochovnikov S, Park H, Varghese N, Huang PS, Pavlopoulos GA, Kim DE, et al. Protein structure determination using metagenome sequence data. Science 2017;355(6322):294–8.

[37] Wang X, Jing X, Deng Y, Nie Y, Xu F, Xu Y, et al. Evolutionary coupling saturation mutagenesis: coevolution-guided identification of distant sites influencing Bacillus Noguensis Pulillulase activity. FEBS Lett 2020;594:799–812.

[38] Celikner G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: using evolutionary data to raise testable hypotheses about protein function. J Struct Chem 2013;53(4):199–206.

[39] Berryman JT, Cai Q, Cheatham T, Darden T. AMBER 12. San Francisco: University of California; 2014.

[40] Dhoke GV, Davari MD, Schwanenberg U, Boccola M. QM/MM calculations revealing the resting and catalytic states in zinc-dependent medium-chain dehydrogenases/reductases. ACS Catal 2015;5(6):3207–15.

[41] Anandakrishnan R, Aguilar B, Onufriev AV. Boccola M. QM/MM calculations revealing the resting and catalytic states in zinc-dependent medium-chain dehydrogenases/reductases. ACS Catal 2015;5(6):3207–15.

[42] Wang J, Kollman PA. How well does a restrained electrostatic potential (Resp) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 2015;21(12):1049–74.

[43] Li P, Merz KM. MCPBP.py: a python based metal center parameter builder. J Chem Inf Model 2016;56:599–604.

[44] Hopf TA, Schärfe CP, Rodrigues JPC, Green AG, Marks DS. Sequence co-evolution gives 3D contacts and structures of protein complexes. eLife 2014;3(6):380–9.

[45] Lee J, Goodey NM. Catalytic contributions from remote regions of enzyme structure. Chem Rev 2011;111(12):7595–624.

[46] Long M, Xu M, Ma Z, Pan X, Kao Z. Significantly enhancing production of trans-4-hydroxy-l-proline by integrated system engineering in Escherichia Coli. ACS Adv 2020;6(21):eaba2383.

[47] Liu Y, Xu G, Zhou J, Ni J, Zhang L, Hou X, et al. Structure-guided engineering of D-carnosylamone reveals a key loop at substrate entrance tunnel. ACS Catal 2020;10(1):12393–402.

[48] Dreesen SE, Selhapelo BM, Horn MM, Scott-Shaw R, Sander P. 5,6-Dihydro-pyrone derivatives from Cryptocarya ilifulis. Phytochemistry 1995;38(6):1427–30.

[49] Jodynis-Liebert Jädwiga, Murias Marek. Effect of sesquiterpene lactones on antioxidant enzymes and some drug-metabolizing enzymes in rat liver and kidney. Planta Med 2000;66(3):199–205.

[50] Li A, Qiu G, Sun Z, Reetz MT. Statistical analysis of the benefits of focused saturation mutagenesis in directed evolution based on reduced amino acid alphabets. ACS Catal 2019;9(9):7769–78.

[51] Honda H, Akagi H, Shimada H. An isoyme of the NADP-malic enzyme of a Bacillus subtilis isozyme. J Bioinform Comput Biol 2009;5(8):567–73.

[52] Kong XD, Ma Q, Zhou J, Zeng BB, Xu JH. A smart library of epoxide hydrolase variants and the top hits for synthesis of (5)-beta-blocker precursors. Angew Chem Int Ed Engl 2014;53(26):6641–4.

[53] Aakre CD, Herrou J, Phung TN, Perchuk BS, Crosson S, Laub MT. Evolving new protein-protein interaction specificity through promiscuous intermediates. Cell 2015;163(3):594–606.

[54] Wang S, Nie Y, Xu Y, Zhang R, Ko T-P, Huang C-H, et al. Unconserved substrate-binding sites direct the stereoselectivity of medium-chain alcohol dehydrogenase. Chem Commun 2014;50(58):7770–2.

[55] Liu L, Shi T, Hoku KN, Zhao Y-L. Understanding the R882H mutation effects of DNA methyltransferase Dnmt1: a combination of molecular dynamics simulations and QM/MM calculations. RSC Adv 2019;9(34):31425–34.

[56] Nobili A, Tao Y, Pavlidis IV, van den Bergh T, Joosten HJ, Tan T, et al. Simultaneous use of in silico design and a correlated mutation network as a tool to efficiently guide enzyme engineering. ChemBioChem 2015;16(5):805–10.

[57] Reetz MT, Boccola M, Carballera JD, Zha D, Vogel A. Expanding the range of ketoreductase for asymmetric reductions. Crit Rev Biotechnol 2019;39(5):3664–5873.