nature communications

Article

https://doi.org/10.1038/s41467-024-53191-8

Enhanced stereodivergent evolution of carboxylesterase for efficient kinetic resolution of near-symmetric esters through machine learning

Received: 25 January 2024

Accepted: 7 October 2024

Published online: 20 October 2024

Check for updates

Zhe Dou[1,2], Xuanzao Chen[1], Ledong Zhu[3], Xiangyu Zheng[1], Xiaoyu Chen[1], Jiayu Xue[1], Satomi Niwayama[4], Ye Ni[1] ✉ & Guochao Xu[1,5] ✉

Carboxylesterases serve as potent biocatalysts in the enantioselective synthesis of chiral carboxylic acids and esters. However, naturally occurring carboxylesterases exhibit limited enantioselectivity, particularly toward ethyl 3-cyclohexene-1-carboxylate (CHCE, **S1**), due to its nearly symmetric structure. While machine learning effectively expedites directed evolution, the lack of models for predicting the enantioselectivity for carboxylesterases has hindered progress, primarily due to challenges in obtaining high-quality training datasets. In this study, we devise a high-throughput method by coupling alcohol dehydrogenase to determine the apparent enantioselectivity of the carboxylesterase *Ac*Est1 from *Acinetobacter* sp. JNU9335, generating a high-quality dataset. Leveraging seven features derived from biochemical considerations, we quantitively describe the steric, hydrophobic, hydrophilic, electrostatic, hydrogen bonding, and π-π interaction effects of residues within *Ac*Est1. A robust gradient boosting regression tree model is trained to facilitate stereodivergent evolution, resulting in the enhanced enantioselectivity of *Ac*Est1 toward **S1**. Through this approach, we successfully obtain two stereocomplementary variants, DR3 and DS6, demonstrating significantly increased and reversed enantioselectivity. Notably, DR3 and DS6 exhibit utility in the enantioselective hydrolysis of various symmetric esters. Comprehensive kinetic parameter analysis, molecular dynamics simulations, and QM/MM calculations offer insights into the kinetic and thermodynamic features underlying the manipulated enantioselectivity of DR3 and DS6.

Enzymes have garnered considerable attention in the realms of synthetic biology and biocatalysis, being widely hailed as the preferred choice for the biosynthesis of optically active chemicals[1–4]. The intricate stereochemical structures of enzymes often manifest in distinctive spatial, hydrophobic, and electrostatic properties, particularly within the active center, forming the basis for high stereoselectivity[5–7]. However, enzymes face challenges maintaining high stereoselectivity, especially when confronted with substrates boasting nearly symmetric structures[8–10]. These substrates are commonly deemed "hard-to-be-discriminated," not only by chemical catalysts but also by biocatalysts[11,12].

Chiral cyclohex-3-ene-1-carboxylic acid (CHCA, **P1**) features a nearly symmetric hexatomic ring, serving as a crucial building block

A full list of affiliations appears at the end of the paper. ✉e-mail: yni@jiangnan.edu.cn; guochaoxu@jiangnan.edu.cn; guochaoxu@163.com

for synthesizing a diverse array of pharmaceuticals, agrochemicals, and natural products (Fig. 1A). Compared to cyclohexane, cyclohexadiene, and phenyl groups, **P1** often exhibits distinctive biological activities due to their unique cyclohexene structure[13]. Enantiomeric pairs of **P1** showcase applicability, with both enantiomers serving as versatile building blocks. For example, (*S*)-**P1** is utilized in the synthesis of Edoxaban, an effective oral medication for treating venous thrombosis and surgical bleeding through the inhibition of coagulation factor Xa[14,15]. Other significant applications include the synthesis of immunosuppressant FK-506[16], aglycone of antitumor drug (+)-phyllanthocin[17], toxin pumiliotoxin C[18], and repellent SS220[19], etc. Conversely, (*R*)-**P1** forms the pivotal building block of Oseltamivir, a widely used oral antiviral drug for treating influenza A by inhibiting neurosine glycosidase[20]. It has also been applied to the natural products leustroducsin B[21], phyllanthocin[22], potential antibacterilas[23], and E-selectin antagonists[24]. Because of their structural similarities with only minor differences in the position of C=C that is distant from the chiral center, distinguishing between (*S*)- and (*R*)-**P1** proves challenging for both chemical and biological catalysts.

Current chemical approaches for synthesizing chiral **P1** involve cumbersome resolution steps and necessitate large amounts of chiral reagents and hazardous organic solvents[25,26], resulting in low atomic economy and environmental concerns (Fig. 1B). Biocatalytic synthesis of chiral **P1** presents its own set of challenges. Non-symmetric esters with varying substituent sizes are relatively easily discriminated by the substrate binding pocket, which excludes the incompatible configuration due to steric hindrance or unfavorable repulsive force (Fig. 1C)[27]. However, discriminating nearly symmetric **P1** esters proves challenging for the substrate binding pocket, as the differences between (*R*)- and (*S*)-esters are minor (Fig. 1D)[28]. Commercial enzymes such as Novozym 435 and pig liver esterase (PLE) exhibit no selectivity in discriminating (*S*)- and (*R*)-**P1** esters[28]. While carboxylesterase BioH is reported to possess activity toward **P1** esters, its low *E* value of merely 2.1 necessitated the iterative construction of a triple mutant, resulting in an increased *E* value of 7.1 at a 40 mM substrate concentration[29]. However, further concentration increases led to a significant decrease in stereoselectivity. Consequently, the enantioselective synthesis of chiral **P1** and esters remains a formidable challenge, posing difficulties for both chemical and biocatalysts alike.

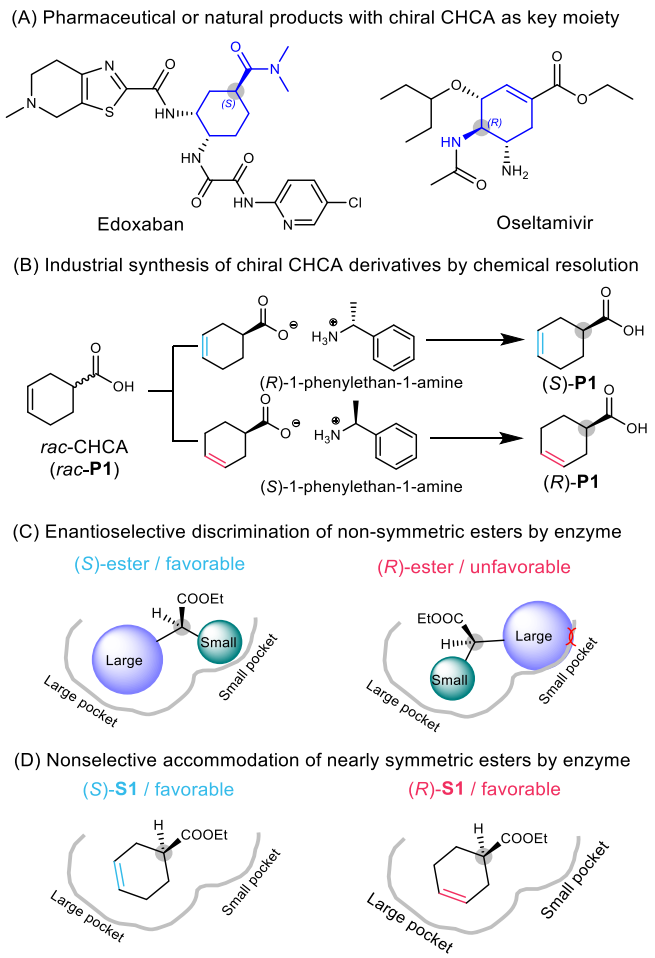


Fig. 1 | Pharmaceuticals containing the moiety of P1 and the asymmetric synthesis of chiral P1 derivatives. Pharmaceuticals with chiral **P1** as the key building block (**A**), industrial synthesis of chiral **P1** by chemical resolution using chiral 1-phenylethane-1-amines as resolution reagents (**B**), enantioselective discrimination of non-symmetric esters (**C**), and nonselective accommodation of nearly symmetric esters (**D**) in the active center of enzymes.

Directed evolution plays a pivotal role in expediting the development of stereoselective enzymes, employing diverse strategies categorized as quantity-intensive and quality-intensive approaches[4,30]. Quantity-intensive strategies involve random mutagenesis across the entire sequence space[31]. This approach necessitates a reliable high-throughput screening (HTS) method with an ideal throughput exceeding 1×10^7 mutants/day[32]. However, the use of chromogenic or electronic substrate analogs in HTS may deviate from research objectives, especially in evolving catalytic activity, enantioselectivity, and substrate specificity. Quality-intensive approaches focus on rational or semi-rational mutagenesis using restricted genetic codons to construct a "smart library"[33,34]. Identifying potential hotspots is crucial for these approaches and is usually determined through empirical, experimental, or computational analysis. While quality-intensive approaches are effective, they may face challenges when dealing with "hard-to-be-discriminated" substrates.

Machine learning (ML) emerges as an alternative, offering a more direct shortcut for boosting directed evolution based on extensive high-quality data and statistical models[35,36]. ML, being data-driven, can identify catalytic patterns, predict promising mutants, and excel in predicting new substitution combinations for directed evolution[37]. ML has found application in manipulating the activity, thermostability, and stereoselectivity of various enzymes, including epoxide hydrolases[38], imine reductase[39], P450 monooxygenase[40], and transaminases[41]. The success of ML predictors depends on the quality of the training dataset[41], with recent efforts by Ran et al. showing promise in predicting enantiomeric excess ratios of hydrolases toward non-symmetric substrates, with positive prediction of 18 out of 28 reactions[42]. The positive prediction ratio was 64%, which might be attributed to the low-quality dataset of hydrolase-catalyzed kinetic resolution reactions, lower in the case of nearly symmetric substrates. Hence, the generation of high-quality dataset and the selection of suitable descriptors are the key to the success of ML predictors.

In this study, we present the establishment of a high-throughput method to generate a high-quality dataset for carboxylesterase *Ac*Est1[43] from *Acinetobacter* sp. JNU9335 using the "real substrate" ethyl cyclohex-3-ene-1-carboxylate (CHCE, **S1**), characterized by its nearly symmetric structure. Subsequently, we constructed an ML predictor utilizing features extracted from biochemical considerations of *Ac*Est1. The trained ML predictor was then applied to design combinatorial mutants. Finally, we successfully achieved stereodivergent evolution of *Ac*Est1, resulting in the generation of two stereocomplementary mutants. These mutants were further employed in the synthesis of both enantiomers of chiral **S1**.

Results

Development of a high-throughput method to obtain a high-quality enantioselectivity dataset

Enantiomeric excess (*e.e.*) is a commonly used parameter to gauge the enantioselectivity of given enzymes. However, when dealing with

hydrolases that exhibit low enantioselectivity, particularly toward nearly symmetric substrates, *e.e.* proves to be an ineffective parameter as its values fluctuate during reaction[44]. To address this limitation and devise a high-throughput method for obtaining a high-quality dataset of enantioselectivity, we adopted the ratio of initial reaction rates between (*R*)- and (*S*)-**S1**, referred to as the apparent enantioselectivity (E_{app}). This approach relies on the "real substrate" and accurately mirrors the actual reaction dynamics (Fig. 2A). The initial reaction rates toward (*R*)- and (*S*)-**S1** were determined by coupling the hydrolytic reaction with an oxidative reaction catalyzed by alcohol dehydrogenase (ADH). An ideal ADH for this method should possess the following characteristics: (1) $NADP^+$-dependent instead of NAD^+-dependent to eliminate background interference; (2) high catalytic efficiency (k_{cat}/K_M), even at lower ethanol concentrations. Consequently, genome mining and rational mutagenesis were employed to identify an $NADP^+$-dependent ADH with high k_{cat}/K_M.

At high ethanol concentration (800 mM), $NADP^+$-dependent ADH6 proved the most efficient, displaying a specific activity as high as 421.2 $U{\cdot}g^{-1}$ (Fig. 2B), while at lower ethanol concentration (2 mM), $NADP^+$-dependent ADH10 from *Kluveromyces polysporus* exhibited the highest specific activity of 9.3 $U{\cdot}g^{-1}$. However, the activity of ADH10 was low and required high loading, which was deemed unsatisfactory for developing a high-throughput method to determine E_{app}. To address this limitation, ethanol was docked into the active center of ADH10 (PDB: 5Z2X), and residues surrounding ethanol were identified (Fig. 2C). In response to the insufficient activity, rational mutagenesis was undertaken to manipulate the hydrophobic interaction and steric hindrance between the methyl group of ethanol and ADH10. Mutations were introduced, turning them into Cys, Phe, Leu, Ile, and Val. Five single mutants, including V84L, V84I, F197L, F197V, and F197I, were obtained, displaying higher activity toward ethanol (Fig. 2D). Among these, F197V proved to be the most efficient, with a specific activity of 17.1 $U{\cdot}g^{-1}$. Consequently, double mutants $ADH10_{V84L/F197V}$ and $ADH10_{V84I/F197V}$ were constructed, with $ADH10_{V84L/F197V}$ displaying a synergistic effect, exhibiting a specific activity of 70.4 $U{\cdot}g^{-1}$, approximately 7.6-fold higher than ADH10. Kinetic parameters analysis revealed that $ADH10_{V84L/F197V}$ not only demonstrated increased ethanol binding affinity with a K_M value of 2.3 mM, much lower than 55.9 mM of WT, but also enhanced catalytic activity, with a k_{cat} of 5.7 min^{-1}. The k_{cat}/K_M of $ADH10_{V84L/F197V}$ was significantly increased to 2.4 $min^{-1}{\cdot}mM^{-1}$ from 0.1 $min^{-1}{\cdot}mM^{-1}$ of WT. The impressive performance of $ADH10_{V84L/F197V}$ encouraged further optimization of conditions for high-throughput determination of E_{app} of *Ac*Est1.

Various factors, including ADH dosage, $NADP^+$ concentration, pH, and **S1** concentrations, were systematically investigated. In Fig. 2E, the changes in A_{340} were similar at ADH dosages higher than 7.5 $U{\cdot}mL^{-1}$. With an increase in $NADP^+$ concentrations from 2.0 to 5.0 mM, the slopes of A_{340} increased accordingly until exceeding 4.0 mM. The effect of **S1** concentrations was also explored, and the slopes of A_{340} increased steadily until reaching 5.0 mM **S1**. Although a further increase in **S1** concentration would result in a higher A_{340} slope, the low solubility of **S1** might interfere with the coupled reaction. Consequently, an ADH dosage of 7.5 $U{\cdot}mL^{-1}$, $NADP^+$ concentration of 4 mM, and **S1** concentration of 5.0 mM were adopted as the optimum conditions for the high-throughput method. Subsequently, the feasibility of this method was evaluated using the entire plate of *Ac*Est1. The average E_{app} of WT was 4.1 ± 0.3, with a CV of less than 8%. This high-throughput method, based on a "real substrate," was developed to obtain a high-quality dataset on the enantioselectivity of *Ac*Est1. Importantly, this method can be extended to characterize other hydrolases with ethanol as the byproduct.

Development of ML predictors for enantioselectivity of *Ac*Est1 toward S1

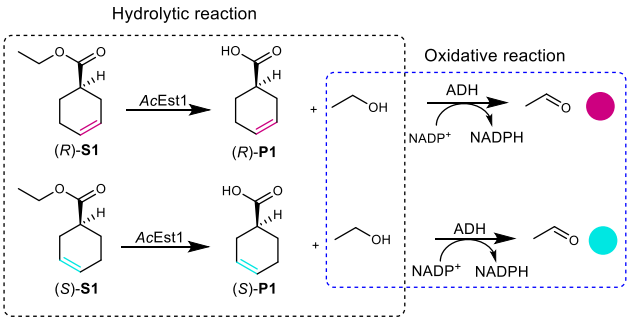
To gather a diverse and high-quality dataset on the enantioselectivity of *Ac*Est1 mutants for constructing an ML predictor, 20 unconservative residues located in the first and second layers surrounding the catalytic S201 were identified for saturation mutagenesis (Fig. 3A). Degenerative codons of 22c-trick[45] including NDT (A/T/C/G, A/T/G, T), VHG (A/C/G, A/T/C, G), and TGG were utilized at molar ratio of 12:9:1 to develop the unbiased library, and a 3-fold excess of mutants was evaluated at each position to ensure the >95% coverage and the quality of library. A total of 1920 mutants were subjected to determining E_{app} values using the high-throughput method. Mutants from I82, V133, Y228, V230, D253, V254, V257, L297, and V328 exhibited significantly enhanced E_{app} values, while mutants from F66, I82, T248, L247, L249, V254, T258, L297, V328, and N329 displayed decreased E_{app} values (Fig. 3B). Among all the single mutants, V257M and L297F had E_{app} values of 9.8 and 0.8, respectively, ranking as the highest and lowest records. After manually removing the data of deactivated mutants, a high-quality dataset containing 760 out of 1920 mutants was obtained for training ML models.

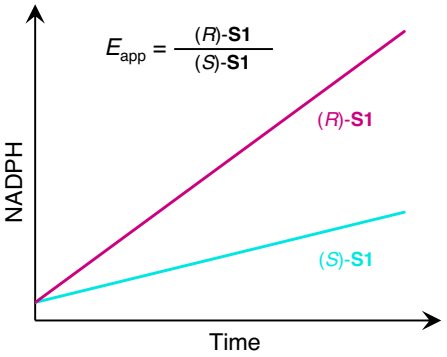
While ML is recognized for its potential in advancing directed evolution, its effectiveness is often hindered by incompatible descriptors or features[46]. In this study, biochemical features, including volume (feature #1), hydrophobicity (feature #2), hydrophilicity (feature #3), electrostatic (feature #4), hydrogen bond (feature #5), π-π interaction (feature #6), and distance to catalytic residue (feature #7), based on 1-D sequence from biochemical considerations, were introduced in this study (Fig. 3C). The use of atom numbers or SMILES format of amino acids as one feature is common. However, this feature is inaccurate for describing the properties of residues, as there is no linear relationship between atom number and spatial volume. For instance, Phe and Lys have the same spatial volume but different molecular weights. Therefore, spatial volume was selected as a feature. The hydrophobicity of the active center is crucial for enzyme performance. Both the hydrophobicity index[47] denoting hydrophobicity (feature #2) and the Kyte-Doolittle hydropathy scale[48] denoting hydrophilicity (feature #3) were introduced as features to describe the hydrophobic interactions. The isoelectric point of an amino acid (feature #4) was selected to describe electrostatic properties. Hydrogen bonding was described by the number of electronegative atoms, such as O and N (feature #5). Stacking interactions, including π-π or alkyl-π interactions, were represented by the number of double bonds (feature #6). Moreover, the distance to nucleophilic S201 was also introduced as feature #7. These features from experimental and biochemical considerations were first introduced in the ML model, providing a robust foundation for training the ML predictor.

Various ML algorithms have been employed to address biocatalytic-related challenges, including Kernel Ridge Regression (KRR), Gaussian Process Regression (GPR), Gradient Boosting Regression Tree (GBRT), Random Forest Regression (RFR), Support Vector Regression (SVR), and Bayesian Ridge Regression (BRR)[35,49]. However, no single algorithm is universally optimal for all tasks. Consequently, we assessed the performance of six regression models in correlating the enantioselectivity of *Ac*Est1 with seven features, namely GBRT, GPR, KRR, RFR, SVR, and BRR. The dataset was randomly divided into a training set (80%) and a testing set (20%), and hyperparameters were tuned for each model. Subsequently, the learning process was executed using the high-quality dataset and the aforementioned models. Based on the regression results, GBRT outperformed GPR, KRR, RFR, SVR, and BRR. The coefficient of determination (R^2) between predicted and experimental E_{app} values of the GBRT model reached a high value of 0.93, with a Mean Square Error (MSE) of 0.12 (Fig. 3D). Landscape analysis revealed a smooth distribution of data, indicating the excellent performance of GBRT. KRR, SVR, and BRR were ineffective in predicting the E_{app} of *Ac*Est1, exhibiting lower R^2 values (below 0.55), and almost all the E_{app} values were underestimated. RFR showed better performance than GPR, with higher R^2 and lower MSE. However, RFR's ability to predict mutants with elevated E_{app} was less robust than GBRT and GPR. The performance of the GBRT model aligned with data-driven protein engineering for the activity and

https://doi.org/10.1038/s41467-024-53191-8

(A)





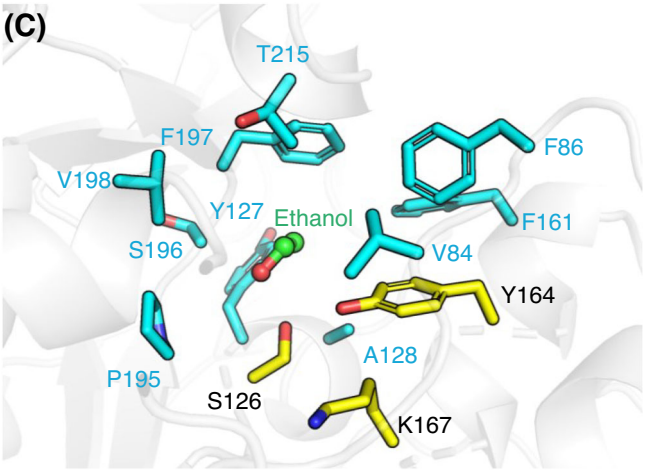
(B)

Genome mining of alcohol dehydrogenases

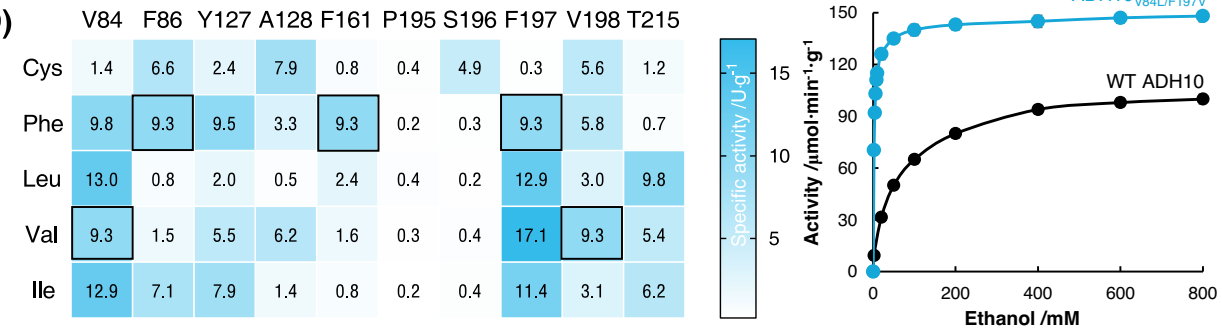
Enzyme	Cofactor	Activity toward 800 mM ethanol (U·g^{-1})	Activity toward 2 mM ethanol (U·g^{-1})
ADH1	NAD$^+$	0.6±0.1	*n. d.*
ADH2	NAD$^+$	2.6±0.1	*n. d.*
ADH3	NAD$^+$	8.5±0.1	*n. d.*
ADH4	NAD$^+$	1.4±0.1	*n. d.*
ADH5	NAD$^+$	45.3±0.1	*n. d.*
ADH6	NADP$^+$	421.2±0.2	4.4±0.1
ADH7	NADP$^+$	10.3±0.1	*n. d.*
ADH8	NADP$^+$	4.2±0.1	*n. d.*
ADH9	NADP$^+$	3.0±0.1	*n. d.*
ADH10	NADP$^+$	100.3±0.5	9.3±0.1
ADH11	NADP$^+$	3.7±0.1	*n. d.*
ADH12	NADP$^+$	76.3±0.2	8.2±0.1

Note: *n. d.* denotes no activity was determined.

(C)



(D)



(E)

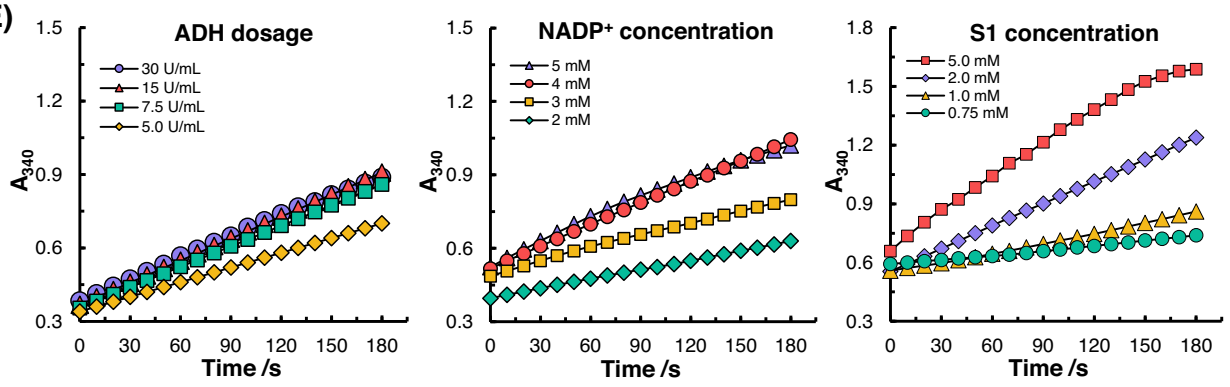


Fig. 2 | Development of a high-throughput method for determining the enantioselectivity value of *Ac*Est1 toward 'real substrate'. A Scheme of the coupled hydrolytic and oxidative reactions for spectrophotometrically determining the E_{app} value. **B** Genome mining of alcohol dehydrogenases. **C** Residues in ADH10 surrounding ethanol for mutagenesis, yellow stick: catalytic triad, green stick: substrate ethanol, cyan stick: residues for mutagenesis. **D** Hydrophobic mutagenesis result and kinetic parameter analysis of WT and ADH10$_{V84L/F197A}$, WT was highlighted with a black border. **E** Effect of ADH dosage, NADP$^+$ and **S1** concentrations on the high-throughput method. $n = 3$ independent biological experiments. Data are presented as mean values ± SD. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-53191-8

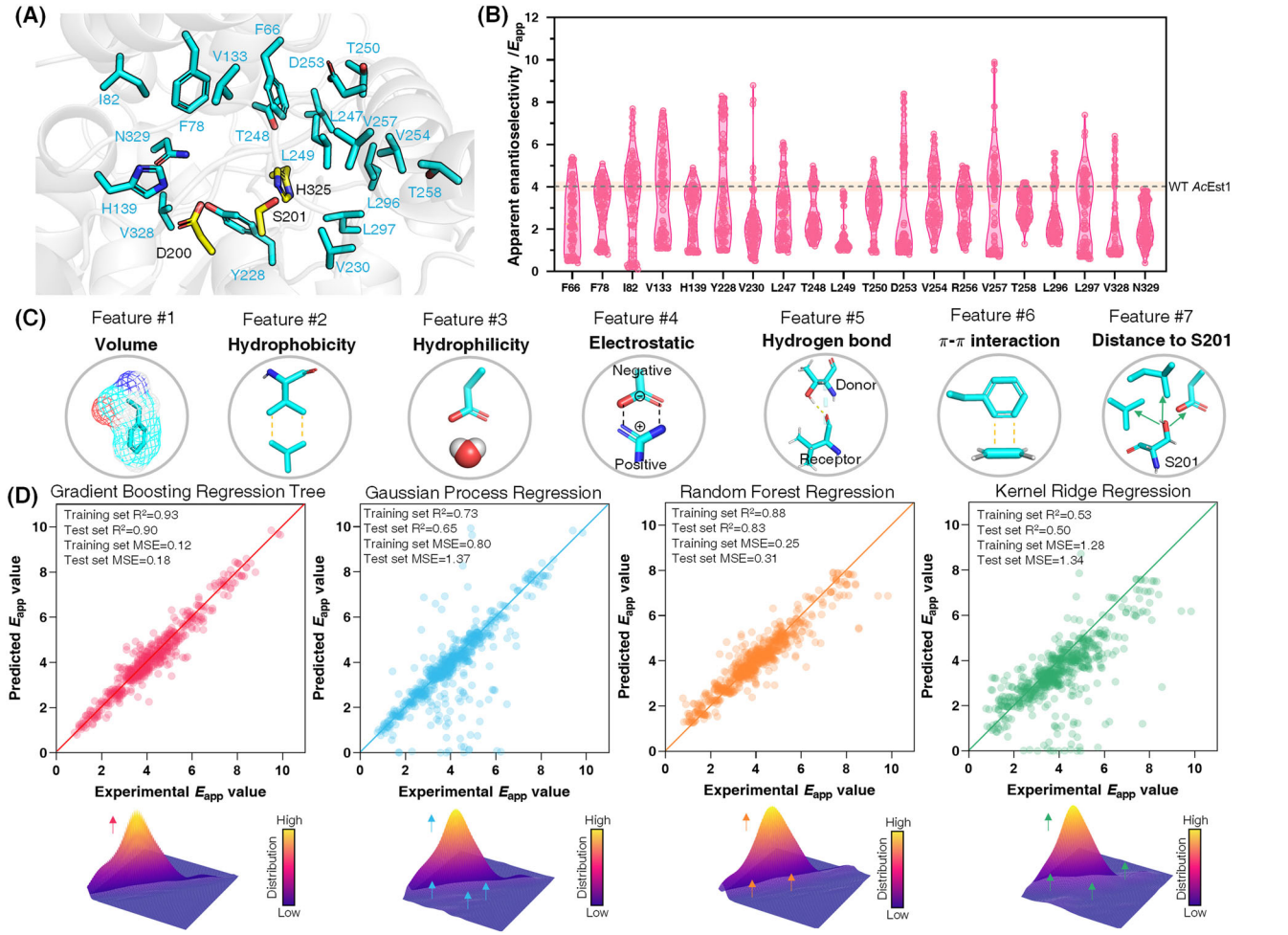


Fig. 3 | High-quality site-specific saturation mutagenesis result and development of machine learning model for predicting the E_{app} value of *Ac*Est1. A First and second layer of residues surrounding catalytic S201. **B** HTS result of site-specific saturation mutagenesis library. **C** Features defined for machine learning of enantioselectivity of *Ac*Est1. **D** Regression performance of different regression models on the training and testing sets. $n = 3$ independent biological experiments. Data are presented as mean values ± SD. Source data are provided as a Source Data file.

stereoselectivity of transaminase, achieving an R^2 of 0.803[41]. The higher R^2 in our model can be attributed to the high-quality dataset and incorporation of biochemical features. Attempts to reduce the number of features and retrain the GBRT model resulted in decreased correlation, proving the synergistic effect of these seven features (Supplementary Figs. S34, S35).

To enhance the accuracy of the trained GBRT predictor, we systematically combined advantageous single mutants to generate double mutants. All single mutants with increased or decreased E_{app} values were paired, resulting in various double mutants. Notably, V257M/Y228M stood out with the highest E_{app} value of 13.8, while L297F/L249A displayed the lowest E_{app} value of 0.36. The incorporation of these double mutants further enriched the dataset used for retraining the GBRT model. The GBRT predictor was then retrained using both the single and double mutants (Fig. 4A). The retrained GBRT predictor demonstrated exceptional performance, achieving an R^2 of 0.97 with an MSE of 0.11, indicative of a robust regression. The remarkable performance of the trained GBRT predictor instilled confidence in its application to guide combinatorial mutagenesis for the stereodivergent evolution of *Ac*Est1.

ML guided stereodivergent evolution of *Ac*Est1

Considering that V257M and L297F were the single mutants with the highest and lowest E_{app} values, they were selected as starting points for the stereodivergent evolution of (*R*)-selective and (*S*)-selective *Ac*Est1 mutants. Consequently, V257M and L297F were designated as DR1 and DS1, respectively. For the (*R*)-selective evolution, we predicted the E_{app} values of all double mutants starting with DR1 and incorporating other mutations (I82M, V133S, Y228M, V230W, D253H, V254Q, V257M, and L297I) with increased E_{app} (Fig. 4B). DR1/Y228M (DR2) emerged as the most (*R*)-selective, consistent with experimental results. Subsequently, we predicted the E_{app} values of triple mutants starting from DR2, leading to DR2/I82M (DR3) with an impressive E_{app} of 18.2. We further forecasted the E_{app} values of quadruple mutants based on DR3, including DR3/V133S, DR3/D253H, DR3/V254Q, and DR3/L297I. V230W was excluded due to the lack of a synergistic effect. As depicted in Fig. 4B, the predicted E_{app} values of quadruple mutants were all higher than their corresponding triple, double, and single mutants, suggesting a synergistic effect among them. However, none of the quadruple mutants exhibited predicted E_{app} values surpassing DR3, which can be attributed to the limited contributions of V133S, D253H, V254Q, and L297I. To validate the accuracy of the GBRT-predicted results, we experimentally constructed DR1, DR2, and DR3, determining their enantioselectivity values (*E* value) through resolution reactions. The *E* values progressively increased from 7.3 for WT to 40.1 for DR1, 59.1 for DR2, and 103 for DR3 (Fig. 4D). Quadruple mutants, such as DR3/V254Q and DR3/L297I, were also experimentally constructed; however, their *E* values were lower than DR3, aligning with the prediction

https://doi.org/10.1038/s41467-024-53191-8

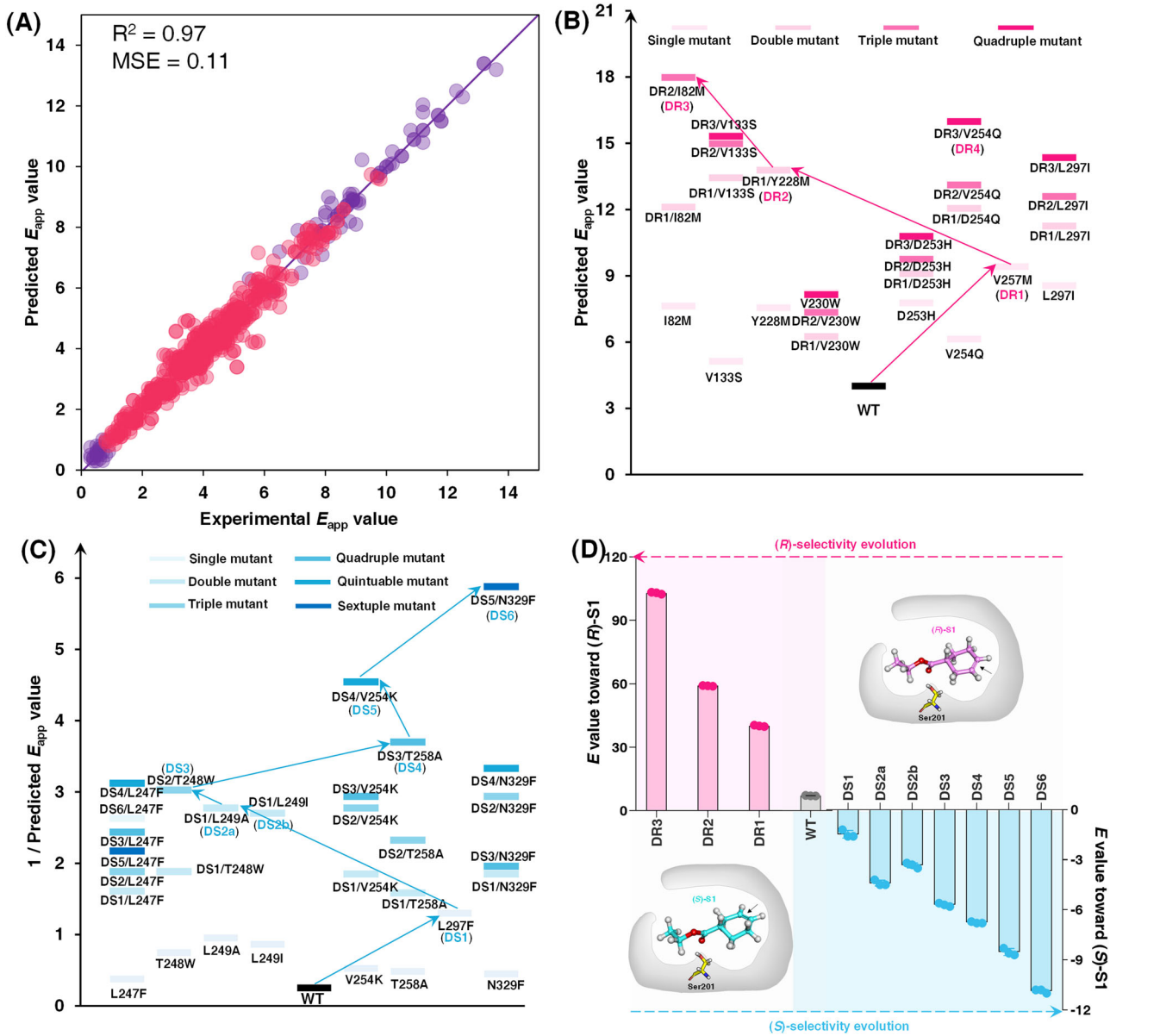


Fig. 4 | ML guided stereodivergent evolution of (*R*)- and (*S*)-selective *Ac*Est1 mutants. A Refined GBRT using double mutants, red dot: single mutant, purple dot: double mutant. **B** GBRT guided (*R*)-selective evolution. **C** GBRT guided (*S*)-selective evolution. **D** Experimental validation of (*R*)- and (*S*)-selective *Ac*Est1 mutants. $n = 3$ independent biological experiments. Data are presented as mean values ± SD. Source data are provided as a Source Data file.

result. The developed DR3 thus represents a promising biocatalyst with excellent enantioselectivity in the resolution of near-symmetric **S1**.

The successful guidance of (*R*)-selective evolution further motivated us to assess the applicability of the trained GBRT predictor in (*S*)-selective evolution. To depict the evolution pathway more clearly, the reciprocal of the predicted E_{app} ((*S*)-**S1**/(*R*)-**S1**) was adopted for (*S*)-selective evolution (Fig. 4C). Among all the single mutants, only L297F exhibited an E_{app} value lower than 1.0, indicating reversed (*S*)-selectivity. Therefore, L297 was designated as DS1 and served as the starting point for (*S*)-selective evolution. All the single mutants with E_{app} values lower than WT, including L247F, T248W, L249A/I, V254K, T258A, and N329F, were added to DS1 to form double mutants. Among all the double mutants, the reciprocal of the predicted E_{app} value for DS1/L249A was 2.8, ranking the highest. Consequently, DS1/L249A was designated as DS2a and subjected to triple mutation. Other mutations were iteratively added to DS2a to predict their E_{app} values. DS2a/T248W displayed a reciprocal predicted E_{app} of 3.0, slightly higher than DS2a, and was regarded as DS3 for further combination. The quadruple mutant DS3/T258A (DS4) exhibited increased enantioselectivity and was combined with other mutations to form quintuple mutants. DS4/V254K showed the highest (*S*)-selectivity among all the quintuable mutants and was regarded as DS5. Furthermore, L247F and N329F were introduced in DS5, resulting in DS5/N329F (DS6) with a reciprocal of predicted E_{app} value as high as 5.9 toward (*S*)-**S1**, even higher than the 4.1 of WT toward (*R*)-**S1**, implying the concrete reversal of the enantioselectivity of *Ac*Est1 from (*R*)-selective to (*S*)-selective.

Although the single mutant L247F exhibited decreased enantioselectivity compared to WT, its contribution to (*S*)-selectivity was limited. The reciprocal predicted E_{app} values of different combinatorial mutants containing L247F were overall increased while locally decreased (Fig. 4C), suggesting the existence of an antagonistic effect. The introduction of L247F in DS6 resulted in decreased enantioselectivity. DS1-DS6 predicted by the GBRT model was experimentally

https://doi.org/10.1038/s41467-024-53191-8

Table 1 | Kinetic parameters of WT *Ac*Est1 and mutants toward (*R*)- and (*S*)-S1

	(*R*)-**S1**			(*S*)-**S1**			
Mutant	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1}•mM^{-1})	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1}•mM^{-1})	Ratio of k_{cat}/K_M
WT	4.3 ± 0.5	675.3 ± 5.2	156.2 ± 4.2	8.4 ± 0.6	109.3 ± 3.2	13.0 ± 0.3	12.0[a]/0.08[b]
DR3	6.2 ± 0.6	60.6 ± 2.5	9.8 ± 1.4	8.4 ± 0.7	0.7 ± 0.1	0.1 ± 0.01	123.3/0.01
V257M	5.8 ± 0.4	324.5 ± 4.4	55.5 ± 2.4	8.9 ± 0.6	8.3 ± 0.6	0.9 ± 0.1	59.0/0.02
Y228M	3.4 ± 0.5	106.4 ± 2.3	31.3 ± 1.8	7.8 ± 0.5	7.4 ± 0.6	0.9 ± 0.1	32.9/0.03
I82M	4.3 ± 0.6	560.5 ± 4.2	131.3 ± 5.1	7.6 ± 0.3	76.7 ± 1.7	10.1 ± 0.3	13.0/0.08
DS6	5.5 ± 0.3	4.1 ± 0.4	0.7 ± 0.1	5.0 ± 0.6	19.9 ± 0.4	4.0 ± 0.2	0.2/5.4
L297F	10.1 ± 0.3	77.4 ± 1.9	7.6 ± 0.5	8.0 ± 0.4	107.2 ± 2.0	13.3 ± 0.3	0.6/1.7
L249A	8.4 ± 0.3	8.6 ± 0.8	1.0 ± 0.1	8.1 ± 0.3	8.9 ± 0.8	1.1 ± 0.1	0.9/1.1
T248W	3.6 ± 0.6	134.4 ± 3.3	36.8 ± 2.5	4.8 ± 0.2	92.1 ± 1.6	19.3 ± 0.4	1.9/0.5
T258A	5.3 ± 0.4	838.8 ± 6.5	158.4 ± 3.5	8.4 ± 0.5	229.4 ± 3.2	27.4 ± 0.4	5.8/0.2
V254K	6.6 ± 0.4	1071.6 ± 5.2	163.3 ± 2.2	9.8 ± 0.6	444.2 ± 4.1	45.4 ± 0.5	3.6/0.3
N329F	2.7 ± 0.2	73.4 ± 1.0	27.6 ± 0.8	5.9 ± 0.2	38.4 ± 1.1	6.5 ± 0.3	4.2/0.2

n = 3 independent biological experiments. Data are presented as mean values ± SD. Source data are provided as a Source Data file.
[a]k_{cat}/K_M of (*R*)-**S1** / k_{cat}/K_M of (*S*)-**S1**.
[b]k_{cat}/K_M of (*S*)-**S1** / k_{cat}/K_M of (*R*)-**S1**.

constructed and evaluated in the resolution of *rac*-**S1**. The *E* value of DS1 was −1.6, reversed from the 7.3 of WT. Since DS2a (DS1/L249A) and DS2b (DS1/L249I) had similar predicted E_{app} values, both were investigated, and DS2a displayed a lower *E*-value of −4.2. From DS3 to DS6, the *E* value further decreased from −5.7 to −11, significantly lower than WT *Ac*Est1. Although the enantioselectivity of DS6 is not as high as DR3, it is quite a significant change for *Ac*Est1, considering the nearly symmetric structure of (*R*)- and (*S*)-**S1**. Stereodivergent evolution of *Ac*Est1 toward nearly symmetric esters has been achieved using this trained GBRT predictor.

Characterization of stereocomplementary *Ac*Est1 mutants

To gain a deeper understanding of enantioselectivity manipulation, kinetic parameters of *Ac*Est1 mutants toward (*R*)- and (*S*)-**S1** were meticulously characterized (Table 1). The WT *Ac*Est1 exhibited k_{cat}/K_M values toward (*R*)- and (*S*)-**S1** of 156.2 and 13.0 s^{-1}·mM^{-1}, respectively, with a resulting calculated ratio of 12.0. Notably, the (*R*)-selective DR3 mutant displayed k_{cat}/K_M values of 9.8 and 0.1 s^{-1}·mM^{-1} toward (*R*)- and (*S*)-**S1**, respectively, leading to a drastically increased ratio of 123.3, approximately 10.3-fold higher than WT, aligning with the observed *E* value. DR3 exhibited k_{cat} values toward (*R*)- and (*S*)-**S1** of 60.6 and 0.7 s^{-1}, respectively, with a ratio of 92, significantly surpassing the 6.2 of WT. In terms of kinetic dynamics, the substantial decrease in the k_{cat} value of DR3 toward (*S*)-**S1** is pivotal for its heightened enantioselectivity. Deconvolution analysis indicated elevated k_{cat}/K_M ratios for V257M and Y228M (59.0 and 32.9, respectively) compared to WT, suggesting their significant contributions to enhanced enantioselectivity. I82M exhibited a modest increase in the k_{cat}/K_M ratio (13.0).

Kinetic parameter and deconvolution analysis of the (*S*)-selective DS6 were also conducted. The k_{cat}/K_M value of DS6 toward (*R*)-**S1** significantly decreased to 0.7 s^{-1}·mM^{-1} from the 156.2 s^{-1}·mM^{-1} of WT, indicating a substantial reduction in catalytic efficiency toward (*R*)-**S1**. Meanwhile, the k_{cat}/K_M value of DS6 toward (*S*)-**S1** was 4.0 s^{-1}·mM^{-1}, with a k_{cat}/K_M ratio of (*S*)-**S1** to (*R*)-**S1** of 5.4, around 67.6-fold higher than that of WT, in agreement with the reversed *E* value of DS6. Deconvolution analysis revealed that L297F exhibited the highest k_{cat}/K_M ratio of 1.7. While L297F maintained the same k_{cat}/K_M value of 13.3 s^{-1}·mM^{-1} toward (*S*)-**S1** as WT, it drastically decreased the k_{cat}/K_M value to 7.6 s^{-1}·mM^{-1} toward (*R*)-**S1**. Additionally, L249A also displayed a higher k_{cat}/K_M ratio of 1.1, about 13.5-fold higher than WT. In contrast to L297F and L249A, single mutants T248W, T258A, and V254K exhibited increased catalytic efficiency toward (*S*)-**S1**, surpassing the increment toward (*R*)-**S1**. T248W, T258A, and V254K contributed not only to the increased (*S*)-selectivity but also to the catalytic efficiency of DS6. T248W and N329F were beneficial for the enhanced binding affinity of DS6 toward (*S*)-**S1**. From the perspective of kinetic dynamics, the significantly decreased catalytic efficiency toward (*R*)-**S1** and enhanced binding affinity toward (*S*)-**S1** are both crucial for the (*S*)-selectivity of DS6.

The k_{cat}/K_M values of DR3 and DS6 were significantly lower than that of WT, hinting at a "trade-off" effect between activity and enantioselectivity toward the near-symmetric ester. This could be attributed to the physical and chemical effects since the catalytic efficiency of WT *Ac*Est1 (k_{cat} of 675.3 s^{-1}) is so high that hard to accurately discriminate (*R*)- and (*S*)-**S1** with nearly symmetric structure. However, it should be mentioned that the catalytic efficiency of DR3 (k_{cat} of 60.6 s^{-1}) and DS6 (k_{cat} of 19.9 s^{-1}) is still high enough for scale-up biocatalytic reactions.

Application potential of stereocomplementary DR3 and DS6 in resolving near-symmetric esters

The potential application of stereocomplementary DR3 and DS6 in the synthesis of chiral **P1** was thoroughly investigated. To qualify as promising biocatalysts with industrial relevance, substrate loading should exceed 100 g·L^{-1}, and the *e.e.* value should surpass 99% within 24 h. After optimization, a substantial 1.0 M *rac*-**S1** (154 g·L^{-1}) could be enantioselectively hydrolyzed into (*S*)-**S1** by DR3 (Fig. 5A). By the 8.0-h mark, the $e.e._s$ reached >99% (*S*) at a conversion ratio of 53% and $e.e._p$ of 88%. (*S*)-**S1** was isolated from the reaction system with a yield of 45.7%. Similarly, DS6 was applied in the resolution of 1.0 M *rac*-**S1** (154 g·L^{-1}) for the synthesis of (*R*)-**S1** (Fig. 5B). At 18 h, $e.e._s$ value reached >−99% (*R*), with a conversion ratio of 74% and $e.e._p$ of −41%. (*R*)-**S1** was isolated with a yield of 23.3%. Notably, the enantioselective resolution of nearly symmetric **S1** for the synthesis of both enantiomers of chiral **S1** was achieved at 1.0 M for the first time, employing stereocomplementary DR3 and DS6.

https://doi.org/10.1038/s41467-024-53191-8

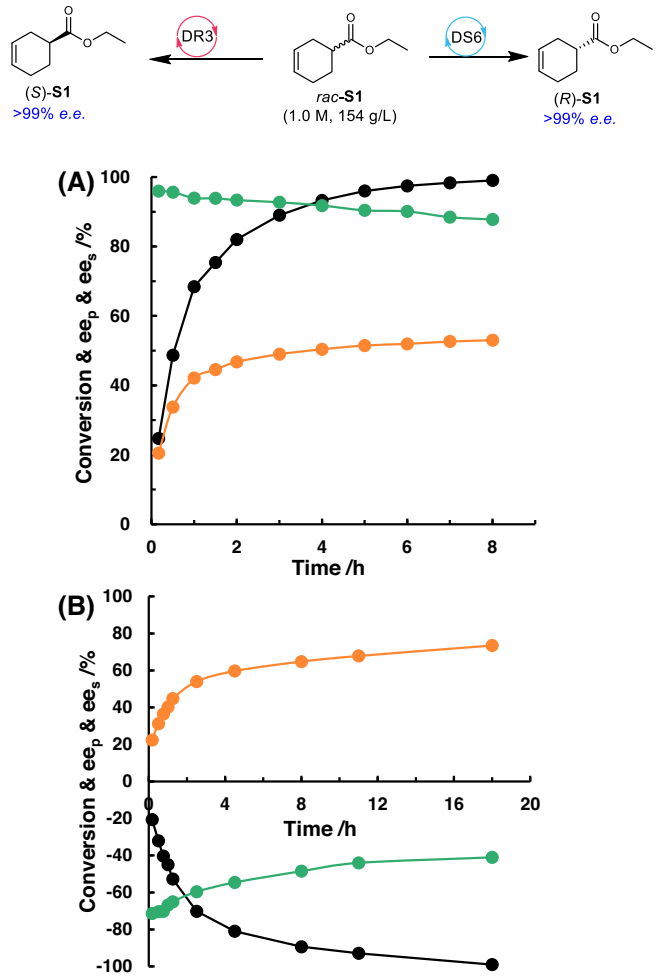


Fig. 5 | Enantioselective resolution of near-symmetric S1 by DR3 and DS6 for the synthesis of chiral (*S*)- and (*R*)-S1. A Time course of DR3 catalyzed enantioselective resolution of *rac*-**S1** for the synthesis of (*S*)-**S1**. **B** Time course of DS6 catalyzed enantioselective resolution of *rac*-**S1** for the synthesis of (*R*)-**S1**. Black dot: $e.e._s$, green dot: $e.e._p$, orange dot: conversion ratio.

The application potential of stereocomplementary DR3 and DS6 was further examined in the enantioselective resolution of esters with a nearly symmetric structure (Table 2). Initially, methyl and isopropyl cyclohex-3-ene-1-carboxylate (**S2** and **S3**) could be enantioselectively hydrolyzed by DR3 and DS6. The *E* values of DR3 toward **S2** and **S3** were 92.3 and 110.4, respectively, dramatically higher than 5.6 and 8.3, consistent with **S1**. Furthermore, with the increase of alcohol groups from methyl to ethyl and isopropyl, the *E* values of DR3 increased accordingly, implying that a larger volume of alcohol groups is favorable for high enantioselectivity. DS6 exhibited reversed (*S*)-selectivity toward **S2** and **S3** with *E* values of −7.0 and −12.3, respectively. DR3 and DS6 demonstrated opposite priorities in the enantioselective accommodation of **P1** esters.

Moreover, other esters with a nearly symmetric structure, specifically oxyheterocyclic esters, were also evaluated by DR3 and DS6 (Table 2). WT *Ac*Est1 displayed a relatively higher activity toward **S4** and **S6** with odd-member rings than **S5** and **S7** with even-member rings. The highest *E* value was 12.1 toward **S7**. DR3 could enantioselectively hydrolyze (*R*)-esters with higher *E* values than WT. The highest *E* value of DR3 was also observed with **S7**. Conversely, DS6 could enantioselective hydrolyze (*S*)-esters with reversed *E* values. The lowest *E* value was observed with **S5** by DS6. All of the above results demonstrate that stereocomplementary DR3 and DS6 can be applied in the enantioselective hydrolysis of other esters containing a nearly symmetric structure.

Molecular mechanism of stereoselectivity control according to MD and QM/MM calculation

In-depth insights into the molecular mechanism of carboxylesterase in the enantioselectivity resolution of nearly symmetric esters were sought by attempting to resolve the crystal structure of stereocomplementary DR3 and DS6. Despite optimization of crystallization conditions and different truncation attempts, crystals with the desired quality for X-ray crystallography were not obtained. Consequently, structural models of DR3 and DS6 were constructed using AlphaFold2, a widely accepted software for building reliable protein structures based on artificial intelligence. The structural models of WT, DR3, and DS6 were minimized and subjected to multiple 100-ns MD simulations. The RMSD achieved stability at about 5 ns, ranging from 1.5 to 2.5 Å.

Near-attack conformation (NAC) analyses were conducted to gain insights into the transition state of nucleophilic attack[50]. According to the transition state, the angle among O of S201, carbonyl C, and carbonyl O of the substrate (θ1: ∠OG-C7-O2) should be within 90° ± 15°, and the nucleophilic attack distance between O of S201 and carbonyl C of the substrate (d1: $dist_{OG\text{-}C7}$) should be less than 3.4 Å. As illustrated in Fig. 6, the percentages of conformations satisfying NAC parameters of WT toward (*R*)- and (*S*)-**S1** were 38.7% and 14.9%, respectively (Fig. 6A, B), consistent with the (*R*)-preference of WT *Ac*Est1. However, the NAC percentage of 14.9% in WT and (*S*)-**S1** hinted that (*S*)-**S1** could also be accommodated by WT.

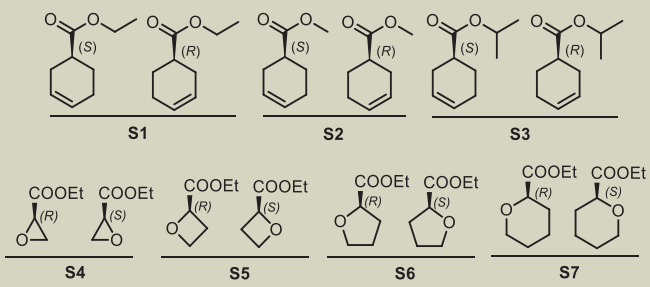
For DR3, the NAC percentage of (*R*)-**S1** was 39.9%, significantly higher than the 5.6% of (*S*)-**S1** (Fig. 6C, D), indicating that hydrolysis of (*R*)-**S1** is more favorable than (*S*)-**S1** in DR3. Regarding DS6, the NAC percentage of (*S*)-**S1** was 29.4%, much higher than the 12.9% of (*R*)-**S1**, proving that (*S*)-**S1** is preferable to (*R*)-**S1** in DS6 (Fig. 6E, F). Binding free energy was also analyzed employing the MMPB/GBSA method[51]. The free energy difference between (*R*)- and (*S*)-**S1** of DR3 was −3.7 kcal·mol^{-1}, higher than the −3.0 kcal·mol^{-1} of WT. For DS6, the free energy difference between (*R*)-**S1** and (*S*)-**S1** was 3.3 kcal·mol^{-1}, suggesting that (*S*)-**S1** was preferable.

Representative conformations with the highest distribution ratio were extracted from MD simulations and are presented in Fig. 7. In the case of WT *Ac*Est1, the distance between the O atom of catalytic S201 and the carbonyl C (d_1) of (*R*)-**S1** measured 3.1 Å, while d_1 of (*S*)-**S1** was 3.9 Å (Fig. 7A, B). Turning to DR3, the d_1 of (*R*)-**S1** decreased to 2.5 Å, and the d_1 of (*S*)-**S1** increased to 4.0 Å (Fig. 7C, D). The mutation of V257 into M257 introduced an alkyl-π interaction with the double bond of (*R*)-**S1**. For DS6, the d_1 of (*S*)-**S1** decreased to 2.6 Å, while the d_1 of (*R*)-**S1** remained the same as in WT (Fig. 7E, F). A favorable π-π interaction between F297 and (*S*)-**S1** was also observed, providing evidence for the increased catalytic efficiency of DS6 toward (*S*)-**S1**.

Five trajectories were randomly selected from MD simulations for calculating the free energy barriers ($\Delta G^{\ddagger}$) for (*R*)- and (*S*)-**S1** of WT, DR3, and DS6 using QM/MM[52]. The energy barriers ($\Delta G^{\ddagger}$) of the rate-determining step for WT&(*R*)-**S1**, WT&(*S*)-**S1**, DR3&(*R*)-**S1**, DR3&(*S*)-**S1**, DS6&(*R*)-**S1**, DS6&(*S*)-**S1** were 13.5, 14.6, 13.7, 16.3, 16.5 and 15.4 kcal·mol^{-1}, respectively (Fig. 8). The free energy difference between (*R*)- and (*S*)-**S1** ($\Delta\Delta G^{\ddagger}$) of WT was calculated to be −1.1 kcal·mol^{-1}, favoring the hydrolysis of (*R*)-**S1**. In contrast, the $\Delta\Delta G^{\ddagger}$ values of DR3 and DS6 were −2.6 and 1.1 kcal·mol^{-1}, respectively. The significant decrease in $\Delta\Delta G^{\ddagger}$ of DR3 and the increase for DS6 were consistent with their performance in the enantioselectivity resolution of **S1**. These findings provide molecular insights into the manipulation of enantioselectivity of stereocomplementary DR3 and DS6 in the accommodation and resolution of **S1** with a nearly symmetric structure.

https://doi.org/10.1038/s41467-024-53191-8

Table 2 | Substrate specificity of WT *Ac*Est1 and its mutants toward other near-symmetric esters



Substrate	WT		DR3		DS6	
	Spec. act.[a] /U·mg^{-1}	*E* value	Spec. act. /U·mg^{-1}	*E* value	Spec. act. /U·mg^{-1}	*E* value[b]
S1	119.2 ± 3.1	7.3 ± 0.1	24.8 ± 0.5	103.2 ± 2.2	2.6 ± 0.2	−11.2 ± 0.5
S2	162.3 ± 4.3	5.6 ± 0.2	29.0 ± 0.7	92.3 ± 2.1	5.2 ± 0.2	−7.0 ± 0.3
S3	91.5 ± 2.2	8.3 ± 0.2	8.1 ± 0.3	110.4 ± 2.3	2.5 ± 0.2	−12.3 ± 0.7
S4	50.5 ± 1.6	3.5 ± 0.1	43.4 ± 1.1	16.3 ± 1.2	8.9 ± 0.4	−1.3 ± 0.1
S5	8.5 ± 0.4	6.3 ± 0.2	6.2 ± 0.2	18.2 ± 1.3	2.5 ± 0.2	−2.7 ± 0.1
S6	46.0 ± 1.5	7.7 ± 0.2	31.4 ± 0.8	22.1 ± 1.3	4.6 ± 0.4	−2.0 ± 0.1
S7	8.0 ± 0.5	12.1 ± 0.2	16.6 ± 0.5	26.4 ± 1.2	2.3 ± 0.2	−1.0 ± 0.1

n = 3 independent biological experiments. Data are presented as mean values ± SD. Source data are provided as a Source Data file.
[a]Spec. act. denotes specific activity.
[b]*E* value refers to enantioselectivity value, and negative *E* value denotes (*S*)-preference.

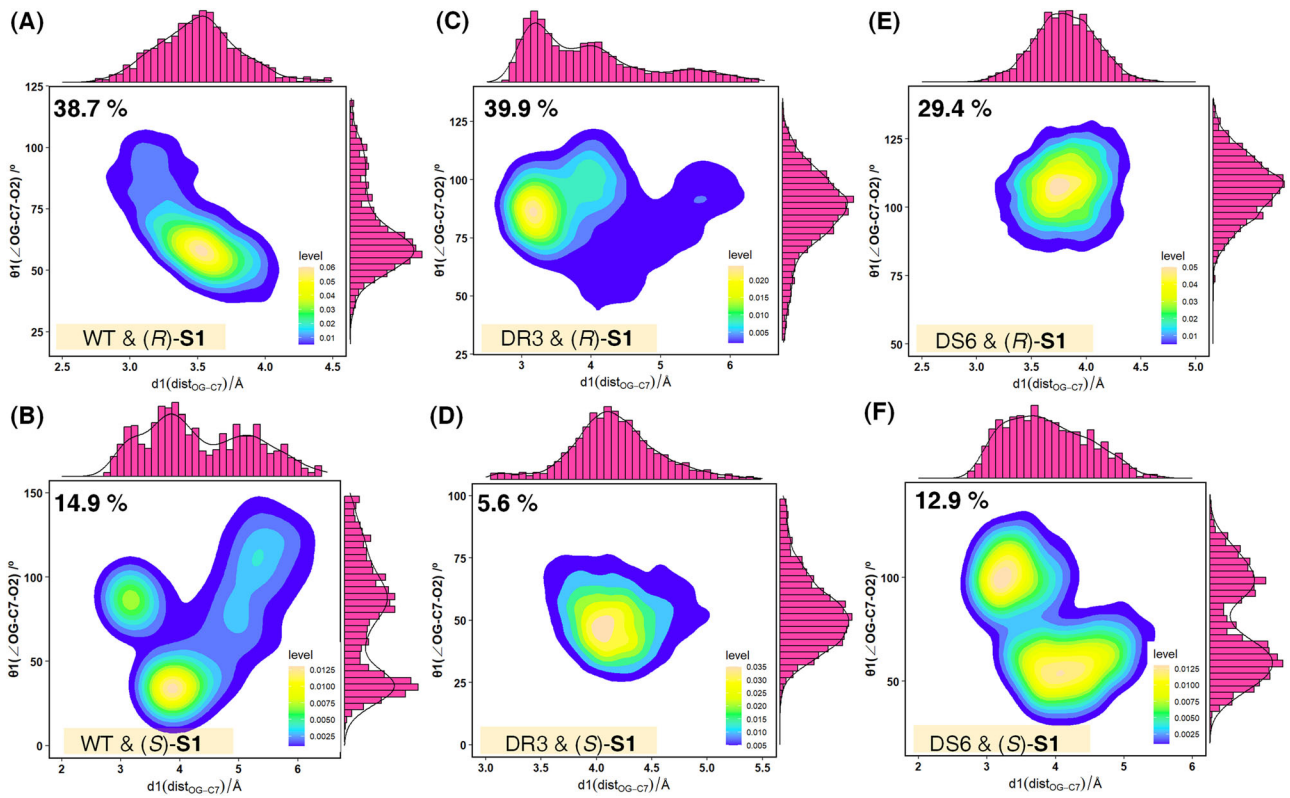


Fig. 6 | Near-attack conformation (NAC) analysis of WT *Ac*Est1 and its mutants from multiple MD simulations. A WT & (*R*)-**S1**. **B** WT & (*S*)-**S1**. **C** DR3 & (*R*)-**S1**. **D** DR3 & (*S*)-**S1**. **E** DS6 & (*R*)-**S1**. **F** DS6 & (*S*)-**S1**. NAC percentage (%) refers to the percentage of conformations satisfying nucleophilic attack criteria with θ1 of 90° ± 15° and d1 of less than 3.4 Å. *n* = 5 independent simulations.

Discussion

Carboxylesterases represent highly promising biocatalysts for synthesizing optically active carboxylic acids and esters[53]. Their appeal lies in their independence from cofactors, high catalytic efficiency, and ease of operation. Recent surveys indicate that approximately 25% of commercialized pharmaceuticals and 40% of crop protection compounds contain at least one carboxylic group[54,55]. This characteristic not only enhances solubility but also improves pharmaceutical

https://doi.org/10.1038/s41467-024-53191-8

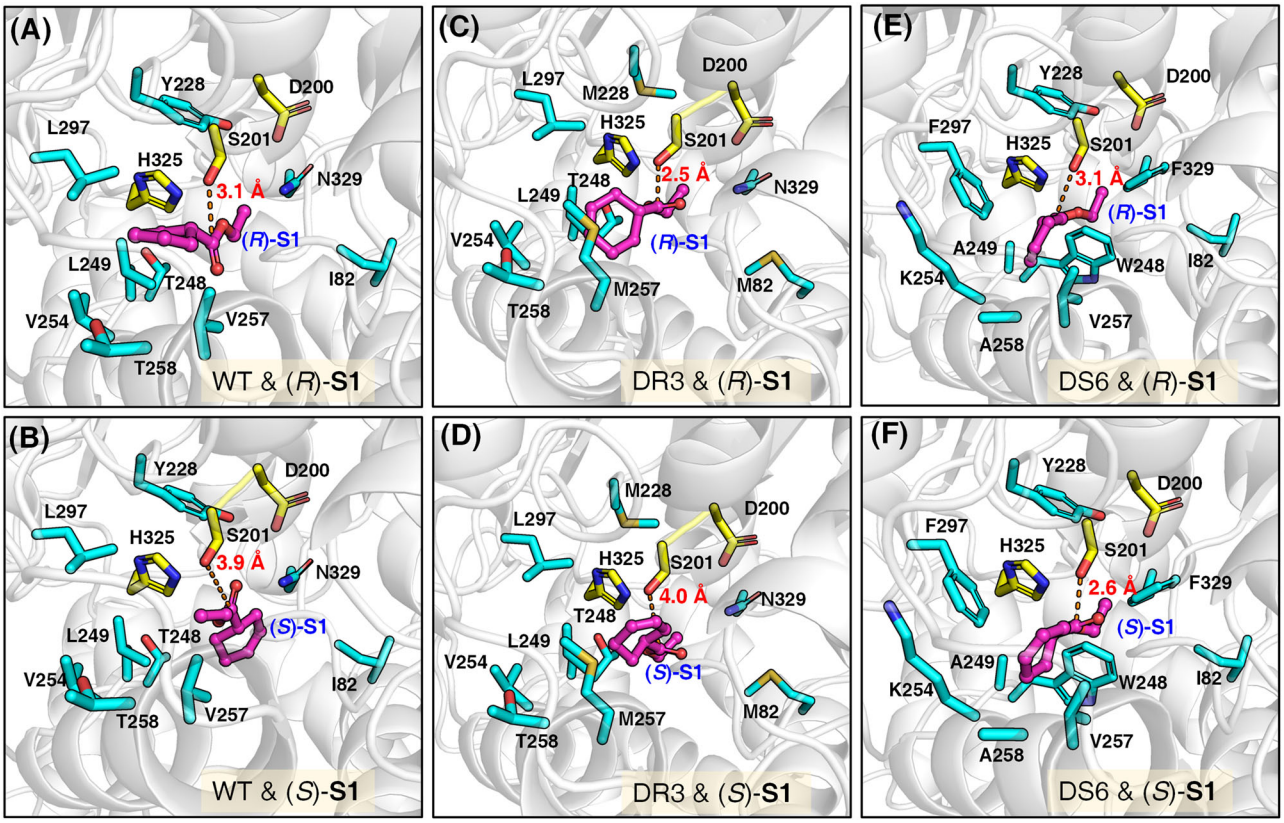


Fig. 7 | Interaction analysis of WT *Ac*Est1 and its mutants toward (*R*)- and (*S*)-S1. A WT & (*R*)-**S1**. **B** WT & (*S*)-**S1**. **C** DR3 & (*R*)-**S1**. **D** DR3 & (*S*)-**S1**. **E** DS6 & (*R*)-**S1**. **F** DS6 & (*S*)-**S1**. Yellow stick: catalytic triad, cyan stick: different residues, purple ball and stick: (*R*)- and (*S*)-**S1**. $n = 5$ independent simulations.

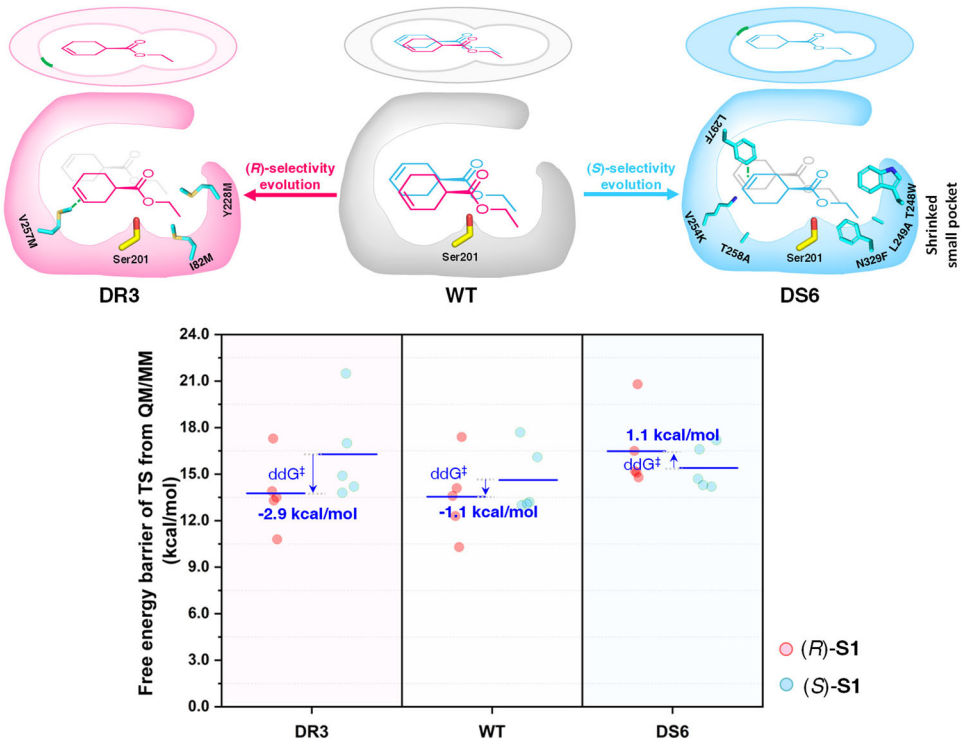


Fig. 8 | Scheme of the substrate binding pockets of WT, DR3 and DS6 with complementary enantioselectivity and free energy barriers analysis using QM/MM calculations. $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{(R)\text{-}S1} - \Delta G^{\ddagger}_{(S)\text{-}S1}$. pink dot: $\Delta G^{\ddagger}$ value toward (*R*)-**S1**, blue dot: $\Delta G^{\ddagger}$ value toward (*S*)-**S1**. $n = 5$ independent simulations.

efficacy, allowing easy derivatization into versatile functional groups like amines or hydroxyl groups[56]. While biocatalysts are generally known for their high enantioselectivity, they often fall short of demonstrating the desired enantioselectivity, particularly for industrially relevant chiral compounds. This challenge becomes more pronounced when dealing with substrates possessing nearly symmetric structures, considered "hard-to-discriminate." Chiral cyclohex-3-ene-1-carboxylic acids (CHCAs, **P1**s), characterized by a nearly symmetric hexatomic ring, are essential building blocks. Examples include (*S*)-**P1** for Edoxaban[15] and (*R*)-**P1** for Oseltamivir[20]. However, commercial and naturally evolved enzymes typically exhibit low enantioselectivity toward **P1** esters[28]. As a result, manipulating carboxylesterase enantioselectivity for the stereocomplementary synthesis of chiral **P1**s has become an area of particular interest. The enzyme enantioselectivity can be engineered through either quantity-intensive directed evolution or quality-intensive (semi-)rational design. To enhance screening throughput, substrate analogs such as *p*-nitrophenol ester, featuring an absorption peak at 405 nm, or 2,3-dibromopropanol ester (toxic to bacterial growth), which replaces glycerol (beneficial to bacterial growth), are commonly employed[57]. However, these substrate analogs often differ, either wholly or partially, from the "real substrate." Consequently, screening outcomes may deviate from the intended objectives. Machine learning (ML) has emerged as a powerful data-driven strategy for expediting directed evolution. The accuracy of a trained ML predictor relies heavily on a high-quality dataset and suitable features. Obtaining accurate enantioselectivity data for carboxylesterases at a high volume can be challenging. A high-throughput method involving an alcohol dehydrogenase coupling was developed to generate a high-quality dataset using the ratio of initial activity between (*R*)-**S1** and (*S*)-**S1**. Additionally, using atom numbers or a binary model of 0 and 1 as features to describe the volume and hydrophobicity of residues and substrates may yield implausible results. Considering parameters from biochemical perspectives, such as spatial volume for residue or substrate size, hydrophobic index for hydrophobicity, Kyte-Doolittle hydropathy scale for hydrophilicity, and isoelectric point for electrostatic properties, is of interest. To address these challenges, we envisioned to introduce biochemical features from experimentalists' considerations in training ML predictors to expedite the stereodivergent evolution of carboxylesterase toward nearly symmetric esters. Seven features—spatial volume, hydrophobic, hydrophilicity, electrostatic, hydrogen bonding, π-π interaction, and distance to catalytic residue—were introduced as descriptors for training ML algorithms. Six ML models were trained, and a GBRT model was derived, exhibiting a high R^2 of 0.93 and a low MSE of 0.12. With the introduction of a dataset containing double mutants, the correlation was further enhanced to 0.97. Contributions of these seven features were also investigated as shown in Supplementary Figs. S34 and S35, proving the synergistic effect of all these features. The GBRT model demonstrated effectiveness in predicting the enantioselectivity of carboxylesterase *Ac*Est1 toward **S1**. Moreover, the potential application of the GBRT model using the aforementioned features was evaluated in learning the enantioselectivity of alcohol dehydrogenase from *Kluyveromyces polysporus*[58] and carbonyl reductase from *Candida glabrata*[59]. As shown in Supplementary Fig. S40, a high coefficient of determination (R^2 values) of 0.92 and 0.89 were obtained. All these prove the application potential of the GBRT model with the aforementioned features. Compared with traditional directed evolution strategies such as iterative saturation mutagenesis and tripe code saturation mutagenesis, this ML predictor is advantageous in saving time and labor force, such as about 95% and 99.3% screening work could be saved for mutagenesis at six positions (Supplementary Fig. S41).

The application of GBRT accelerated the stereodivergent evolution of *Ac*Est1, creating stereocomplementary mutants DR3 and DS6. Kinetic parameter analysis revealed that DR3 displayed a significantly increased k_{cat}/K_M ratio of (*R*)-**S1** to (*S*)-**S1**, while DS6 exhibited an increased k_{cat}/K_M ratio of (*S*)-**S1** to (*R*)-**S1**. DR3 and DS6 were successfully employed in the enantioselective resolution of other nearly symmetric esters. Molecular dynamic simulations and QM/MM calculations provided kinetic and thermodynamic evidence for manipulating the enantioselectivity of *Ac*Est1 toward **S1** with a nearly symmetric structure. This study presents an effective ML predictor for expediting the directed evolution of carboxylesterase enantioselectivity. It introduces two stereocomplementary carboxylesterase mutants for the preparation of enantiomerically enriched 3-cyclohexene-1-carboxylic acids for the synthesis Edoxaban and Oseltamivir, and many other significant compounds. Further evolving (*S*)-selectivity and breaking the trade-off between activity and enantioselectivity are in progress by developing dual-target ML method on 320 residues outside substrate binding pocket.

Methods

Screening and protein engineering of alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) stored in our lab were submitted for determination of the oxidative activity toward ethanol. A 200 μL reaction system containing 0.5 mM NAD^+ or $NADP^+$ (Bontac Bio-Engineering (Shenzhen) co., Ltd), 800 mM or 2 mM ethanol, 10 μL ADH solutions and 170 μL Tris-HCl (pH 8.0, 100 mM) was performed at 30 °C. Absorbance changes at 340 nm referring to the changes of NAD(P)H were monitored. One unit (U) of activity was defined as the amount of ADH required for the production of 1 μmol NAD(P)H per minute. All assays were performed in triplicate.

Substrate ethanol was docked into the crystal structure of ADH10 in complex with $NADP^+$ (PDB ID. 5Z2X). Residues at the methyl terminal of ethanol, including V84, F86M Y127, A128, F161, P195, S196, F197, V198, and T215, were selected for mutagenesis employing High-fidelity KOD Plus Neo polymerase (TOYOBO CO., LTD.) with primers listed in Supplementary Table S1[60]. General protocol for whole-plasmid PCR contained steps of pre-denaturation at 96 °C for 5 min, twenty-cycles of amplification including denaturation at 98 °C for 15 s, annealing at 55 °C for 15 s and elongation at 68 °C for 3 min, and further elongation at 68 °C for 10 min. The resultant PCR products were digested with *Dpn*I to remove the template plasmids and were chemically transformed into *E. coli* BL21(DE3). ADH10 mutants were induced by 0.2 mM IPTG at 25 °C for 12 h and lysed by ultrasonication at 300 W for 15 min in an ice-water bath. The cell debris was removed by centrifugation to obtain the crude enzyme extract. Then, the specific activity toward 2 mM ethanol of ADH10 mutants was determined, as mentioned above. Mutations of V84L and F197V were combined to obtain $ADH10_{V84L/F197V}$. WT and double mutant with His_6-tag were purified employing nickel-affinity chromatography. The kinetic parameters of purified enzymes were determined by varying the concentrations of ethanol (1–800 mM). The K_M, V_{max}, k_{cat} values were calculated by non-linear fitting to the Michaelis-Menten equation. All determinations were conducted in triplicate. Supernatant of $ADH10_{V84L/F197V}$ was obtained by centrifuge and lyophilized under vacuum to obtain crude enzyme power, which was stored at 4 °C for further use.

Establishment of a high-throughput method for determination of E_{app}

A high-throughput method was established for the determination of apparent enantioselectivity (E_{app}) of *Ac*Est1 by coupling $ADH10_{V84L/F197V}$ based on released ethanol from (*R*)- and (*S*)-**S1**. A 200 μL reaction system consisting of (*R*)- or (*S*)-**S1**, $NADP^+$, $ADH10_{V84L/F197V}$ and *Ac*Est1 was conducted at 30 °C to spectrophotometrically monitor the increase of NADPH at 340 nm. First, background interference was excluded using the ADH-free, *Ac*Est1-free, and **S1**-free systems. Then, different dosages of $ADH10_{V84L/F197V}$, including 5.0, 7.5, 15, and 30 $U{\cdot}mL^{-1}$, $NADP^+$ concentrations of 2.0, 3.0, 4.0, and 5.0 mM,

S1 concentrations of 0.75, 1.0, 2.0, and 5.0 mM, were systematically optimized. The final reaction system included 10 μL $ADH10_{V84L/F197V}$ (7.5 U·mL^{-1}), 10 μL $NADP^+$ (4 mM), 10 μL (*R*)- or (*S*)-**S1** (5.0 mM), 10 μL *Ac*Est1 solution in 160 μL Tris-HCl (pH 8.0, 100 mM). E_{app} value was spectrophotometrically obtained by calculating the ratio of (*R*)-**S1** to (*S*)-**S1** or (*S*)-**S1** to (*R*)-**S1** (1). A total of 96 replicates of WT *Ac*Est1 was cultivated for determining their E_{app} and calculating the average E_{app} value and standard deviation of this method.

$$E_{app} = \frac{(R) - \mathbf{S1}}{(S) - \mathbf{S1}} \tag{1}$$

Site-directed saturation mutagenesis and combinatorial mutagenesis of *Ac*Est1

Site-directed saturation mutagenesis of *Ac*Est1 was conducted with plasmid pET28-*AcEst1* as template. Degenerative codons of 22c-trick, including NDT, VHG and TGG, were employed to introduce saturation mutagenesis (Supplementary Table S2). Primers were mixed at a ratio of 12:9:1 to encode all the twenty amino acids with only two redundant. Whole-plasmid PCR was performed using KOD polymerase, and the resultant product was digested by *Dpn*I as mentioned above and transformed into *E. coli* BL21(DE3). After evaluation by colony PCR and sequencing, positive colonies were inoculated into a 96-well plate supplemented with LB medium and kanamycin. Then, the plate was cultivated at 37 °C and sub-transferred to another 96-well plate for induction expression at 25 °C for 12 h. The cells were harvested by centrifugation and lysed by lysozyme at 37 °C for 2 h. After centrifugation at 8000×*g* and 4 °C for 20 min, the supernatants were subjected to E_{app} analysis using the above-described high-throughput method. All measurements were performed in triplicate. The mutants with distinctly different E_{app} (higher or lower) compared to WT were recultivated for further evaluation. A high-quality dataset was established for subsequent machine learning.

Beneficial single mutants were randomly combined to construct double mutants employing the primers listed in Supplementary Table S3. PCR reaction and transformation were conducted as mentioned above. Double mutants were verified by sequencing, and positive clones were inoculated in LB medium and cultivated overnight. After inductive expression by IPTG, double mutants were obtained and subjected for determination of E_{app}. Triple, quadruple, quintuable and sextuple mutants guided by machine learning were also constructed using primers listed in Supplementary Table S3. After verification, their E_{app} and enantioselectivity value (*E* value) were determined.

Machine learning

The high-quality dataset of E_{app} was built based on features including volume, hydrophobicity, hydropathy scale, isoelectric point, hydrogen bond, π-interaction, and distance to nucleophilic S201 listed in Supplementary Table S4. Features of single mutants were obtained by subtraction with those of WT (Supplementary Fig. S25), features of combinatorial mutants were obtained by addition of single mutants. Machine learning was performed on Jupyter Notebook installed with the Scikit-Learn package. Six regression algorithms, including GBRT, KRR, GPR, RF, SVR, and BRR, were trained using the high-quality dataset. These hyperparameters were tuned with a ten-fold cross-validated grid-search on the training set (Supplementary Table S5). Subsequently, these models were retrained on the training set with the above-optimized hyperparameters and evaluated on the test set. The quality of ML predictors was evaluated by the coefficient of determination (*R*-square, R^2) and mean squared error (MSE) (2–4). Based on the initial dataset with 760 mutants, the GBRT was proved to be the best algorithm to build the ML model, using hyperparameters of n_estimators=250, learning_rate=0.04, max_depth=10, min_samples_split=2, min_samples_leaf=4.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3}$$

$$MSE(y, \hat{y}) = \frac{i}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

Where y_i and are $\hat{y}_i$ the true and predicted values of the i-th mutant, respectively.

E_{app} data of (*R*)- and (*S*)-selective double mutants was added into the dataset, which was subsequently retrained by the GBRT predictor. The refined GBRT predictor was evaluated by R^2 and MSE. Furthermore, combinatorial mutants were virtually constructed, and their E_{app} values were predicted by GBRT algorithm. Mutants with the highest E_{app} values at each round of combination were further experimentally verified using the above-mentioned method. The activity and enantioselectivity of combinatorial mutants were measured to achieve the stereodivergent evolution of (*R*)- and (*S*)-selective mutants.

Characterization of *Ac*Est1 mutants

General protocol for determination of the activity and enantioselectivity of *Ac*Est1 and mutants was performed in a 10-mL reaction mixture containing 50 mM *rac*-**S1** and appropriate amount of enzyme solution in PBS buffer (pH 8.0, 100 mM) at 30 °C and 180 rpm. At different time intervals, samples were withdrawn from the reaction mixture, and the reaction was terminated by the addition of 1.0 M HCl, followed by extraction with equal volume of ethyl acetate supplemented with 1 mM dodecane as the internal standard. The upper organic phase was isolated and dried over anhydrous Na_2SO_4, analyzed by chiral GC or HPLC (Supplementary Table S6). The activity (U) was defined as the amount of enzyme required for the decrease of 1 μmol **S1** at the above conditions. The enantiomeric excess (*e.e.*) and enantioselectivity value (*E* value) were calculated according to Eqs. 5–7.

$$ee_s = \frac{[(S) - \mathbf{S1}] - [(R) - \mathbf{S1}]}{[(S) - \mathbf{S1}] + [(R) - \mathbf{S1}]} \times 100\% \tag{5}$$

$$ee_p = \frac{[(R) - \mathbf{P1}] - [(S) - \mathbf{P1}]}{[(R) - \mathbf{P1}] + [(S) - \mathbf{P1}]} \times 100\% \tag{6}$$

$$E\ \text{value} = \frac{\mathrm{Ln}[(1 - \text{conversion ratio}) \times (1 - ee_s)]}{\mathrm{Ln}[(1 - \text{conversion ratio}) \times (1 + ee_s)]} \tag{7}$$

Kinetic parameters of *Ac*Est1 and mutants were determined using standard protocol except for different substrate concentrations ranging from 1 mM to 20 mM. The initial activities were obtained. Subsequently, the K_M, V_{max} and k_{cat} values were calculated by non-linear fitting to the Michaelis-Menten equation employing Origin package. All assays were performed at least three times.

Application potential of *Ac*Est1, DR3, and DS6 in the resolution of other nearly symmetric esters was performed employing the standard activity assay conditions. Different substrates were added in the reaction system and the specific activity was calculated according to the initial reaction rates. *E* values toward different substrates were determined at conversion ratios of 50% based on Eq. 7. All determinations were conducted in triplicate.

Enantioselective resolution of *rac*-S1 by DR3 and DS6 at gram scale

Recombinant *E. coli* whole cells expressing (*R*)-selective DR3 and (*S*)-selective DS6 were prepared as mentioned above. The cells were lyophilized under vacuum to form the dry cells, which were stored at 4 °C for further use. To achieve stereoselective synthesis of chiral **P1** at gram scale, 100 mmol (1.0 M) *rac*-**S1**, appropriate of dry cells of DR3 and DS6 were dissolved in 100 mL PBS buffer (pH 8.0, 100 mM) in a three-neck flask. Preparative reaction was magnetically stirred at 180 rpm and 30 °C. Then, 1.0 M Na_2CO_3 was spontaneously titrated to maintain the pH of the reaction mixture at 8.0. The reaction was terminated until the ee_s achieving >99% by adjusting the pH to basic condition and adding equal volume of ethyl acetate three times. All the organic phases were combined and dried over anhydrous Na_2SO_4 overnight. Products were obtained by evaporation under vacuum and verified by ^{1}H-NMR and ^{13}C-NMR.

MD simulations and QM/MM calculation

Structure models of WT *Ac*Est1, DR3 and DS6 were built employing AlphaFold2 and verified by SAVES[61]. (*R*)- and (*S*)-**S1** were docked into the active center of *Ac*Est1, DR3 and DS6 using AutoDock Vina. Docking poses with relative higher scores were regarded as the initial structure from molecular dynamic (MD) simulations. Subsequently, *Ac*Est1, DR3 and DS6 were protonated using H^{++}. All the MD simulations were performed using the Amber20 packages with GPU acceleration on the hypercomputer at the School of Biotechnology. The force fields for protein, **S1** and water molecules were ff14SB, gaff and TIP3P, respectively. Sodium and chloride were added to neutralize simulation system, and a TIP3P water box with a clearance distance of 15 Å around the protein was added. There were 27,208 atoms in the final simulation systems First, 10,000-step of energy minimization with the steepest descent method was employed, followed by 1 ns heating at NVP from 0 K to 300 K, and 2 ns equilibrating at NPT and 300 K. Production was performed in NPT ensembles for 100 ns with dt of 0.002 ps. All MD simulations were conducted in five replicates. Binding free energy was analyzed using MMPB/GBSA, distance and angle were statistically analyzed employing Chimera 1.6, interactions between enzymes and substrates were analyzed with Discover studio 4.5 and visualized with Pymol.

Quantum mechanics/molecular mechanics (QM/MM) calculations were performed employing Turbomole 7.21[62] and DL-POLY2 modules within the Chemshell 3.7.1 software package3[52]. The QM region includes the catalytic triad Asp200-Ser201-Hid325 and the substrate molecule, totaling 71 atoms. The atoms in the QM region are treated with DFT methods, while the atoms in the MM region are subjected to energy calculations using the CHARMM36 force field[63]. For the polarization effects in the QM region, the static electric field embedding method is employed, and the QM/MM boundary is treated using the hydrogen-linking method of the charge transfer model. Considering the medium- and long-range dispersion interactions of the system, the dispersion correction method Gimme-D3 is applied in all QM/MM calculations. The structural optimizations are completed using B3-LYP-D3/def-SVP/CHARMM36[64], and single-point energy calculations are performed using M06-2X-D3/def2-TZVP[65]. The energy barrier of the reaction was calculated using the Boltzmann-weighted average method, with the formula as follows:

$$\Delta G^{\ddagger} = -RT\ln\left(\frac{1}{N}\sum_{j=1}^{N}\exp\left(-\frac{\Delta G_j}{RT}\right)\right) \quad (8)$$

Where $\Delta G^{\ddagger}$ represents the average barrier, R is the gas constant, T is the temperature at which the reaction occurs, *N* is the number of the selected conformations, and ΔG_j represents the barrier of conformation j.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data needed to evaluate the conclusions in the paper are available in the main text or the supplementary information. There is no restriction on data availability. Source data are provided with this paper.

Code availability

All codes are provided separately with this paper and deposited in GitHub (https://github.com/guochaoxu2019/NCOMMS-24-05359.git). There is no restriction on code availability.

References

1. Bornscheuer, G. W. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
2. Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).
3. Nazor, J., Liu, J. & Huisman, G. Enzyme evolution for industrial biocatalytic cascades. *Curr. Opin. Biotechnol.* **69**, 182–190 (2021).
4. Reetz, M. T. Witnessing the birth of directed evolution of stereoselective enzymes as catalysts in organic chemistry. *Adv. Synth. Catal.* **364**, 3326–3335 (2022).
5. Falivene, L. et al. Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* **11**, 872–879 (2019).
6. Cheng, F. & Chen, Y. et al. Controlling stereopreferences of carbonyl reductases for enantioselective synthesis of Atorvastatin precursor. *ACS Catal.* **11**, 2572–2582 (2021).
7. Wu, L. J. et al. Computer-aided understanding and engineering of enzymatic selectivity. *Biotechnol. Adv.* **54**, 107793 (2022).
8. Sousa, H. A., Afonso, C. A. M. & Crespo, J. G. Kinetic study of the enantioselective hydrolysis of a meso-diester using pig liver esterase. *J. Chem. Technol. Biotechnol.* **75**, 707–710 (2000).
9. Niwayama, S. & Cho, H. Practical large scale synthesis of half-esters of malonic acid. *Chem. Pharm. Bull.* **57**, 508–510 (2009).
10. Dou, Z. et al. Kinetic resolution of nearly symmetric 3-cyclohexene-1-carboxylate esters using a bacterial carboxylesterase identified by genome mining. *Org. Lett.* **23**, 3043–3047 (2021).
11. Garcia-Urdiales, E., Alfonso, I. & Gotor, V. Enantioselective enzymatic desymmetrizations in organic synthesis. *Chem. Rev.* **105**, 313–354 (2005).
12. Borissov, A. et al. Organocatalytic enantioselective desymmetrization. *Chem. Soc. Rev.* **45**, 5474–5540 (2016).
13. Herdewijn, P. & de Clercq, E. The cyclohexene ring as bioisostere of a furanose ring: synthesis and antiviral activity of cyclohexenyl nucleosides. *Bioorg. Med. Chem. Lett.* **11**, 1591–1597 (2001).
14. Nagata, T. et al. Stereoselective synthesis and biological evaluation of 3,4-diaminocyclohexanecarboxylic derivatives as factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **18**, 4587–4592 (2008).
15. Michida, M. et al. Development of an efficient manufacturing process for a key intermediate in the synthesis of Edoxaban. *Org. Process Res. Dev.* **23**, 524–534 (2019).
16. Marco, C. et al. Practical, asymmetric synthesis of the cyclohexyl C_{28}-C_{34} fragment of the immunosuppressant FK-506 via (*S*)-(−)-3-cyclohexenecarboxylic acid. *Tetrahedron* **48**, 539–544 (1992).
17. Trost, B. M. & Kondo, Y. An asymmetric synthesis of (+)-phyllanthocin. *Tetrahedron Lett.* **32**, 1613–1616 (1991).
18. Toyota, M., Asoh, T., Matsuura, M. & Fukumoto, K. Stereoselective transformation of enantiopure cyclohexanol into cis-hydrindan. An enantioselective formal total synthesis route to (+)-pumiliotoxin C. *J. Org. Chem.* **61**, 8687–8691 (1996).

19. Klun, J. A. & Gupta, R. Optically active arthropod repellents for use against disease vectors. *J. Med. Entomol.* **37**, 182–186 (2000).
20. Raghavan, S. & Babu, V. S. Enantioselective synthesis of oseltamivir phosphate. *Tetrahedron* **67**, 2044–2050 (2011).
21. Miyashita, K. et al. Total synthesis of leustroducsin B via a convergent route. *Tetrahedron Lett.* **48**, 3829–3833 (2007).
22. Martin, S. F. et al. Application of nitrile oxide cycloadditions to a convergent, asymmetric synthesis of (+)-phyllanthocin. *J. Org. Chem.* **54**, 2209–2216 (1989).
23. Miles, T. J. et al. Novel cyclohexyl-amides as potent antibacterials targeting bacterial type IIA topoisomerases. *Bioorg. Med. Chem.* **21**, 7483–7488 (2011).
24. Schwizer, D. & Patton, J. T. et al. Pre-organization of the core structure of E-selectin antagonists. *Chem. Eur. J.* **18**, 1342–1351 (2012).
25. Wang, X., Ma, M. L., Reddy, A. G. K. & Hu, W. H. An efficient stereoselective synthesis six stereoisomers of 3,4-diaminocyclohexane carboxamide as key intermediates for the synthesis of factor Xa inhibitors. *Tetrahedron* **73**, 1381–1388 (2017).
26. Riedl, R., Heilmayer, W. & Spence, L. Enantiomerically Pure Amines. PCT Patant WO 2011/146953 A1 https://worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&II=0&ND=3&adjacent=true&locale=en_EP&FT=D&date=20111201&CC=WO&NR=2011146953A1&KC=A1 (2011).
27. Zhao, D. T. et al. Enzymatic resolution of ibuprofen in an organic solvent under ultrasound irradiation. *Biotechnol. Appl. Biochem.* **61**, 655–659 (2014).
28. Dou, Z., Xu, G. C. & Ni, Y. Efficient microbial resolution of racemic methyl 3-cyclohexene-1-carboxylate as chiral precursor of Edoxaban by newly identified *Acinetobacter* sp. JNU9335. *Enzym. Microb. Technol.* **139**, 109580 (2020).
29. Wu, X. F. et al. Improved enantioselectivity of *E. coli* BioH in kinetic resolution of methyl (*S*)-3-cyclohexene-1-carboxylate by combinatorial modulation of steric and aromatic interactions. *Biosci. Biotechnol. Biochem.* **83**, 1263–1269 (2019).
30. Arnold, F. H. Innovation by evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed. Engl.* **58**, 14420–14426 (2019).
31. Shivange, A. V. & Marienhagen, J. Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* **13**, 19–25 (2009).
32. Ma, F. Q. et al. Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nat. Commun.* **9**, 1030 (2018).
33. Qu, G. et al. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed. Engl.* **59**, 13204–13231 (2020).
34. Cheng, F., Zhu, L. L. & Schwaneberg, U. Directed evolution 2.0: improving and deciphering enzyme properties. *Chem. Commun.* **51**, 9760–9772 (2015).
35. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
36. Li, G. Y., Dong, Y. J. & Reetz, M. T. Can machine learning revolutionize directed evolution of selective enzymes? *Adv. Synth. Catal.* **361**, 2377–2386 (2019).
37. Siedhoff, N. E., Schwaneberg, U. & Davari, M. D. Machine learning-assisted enzyme engineering. *Methods Enzymol.* **643**, 281–315 (2020).
38. Cadet, F. et al. A machine learning approach for reliable prediction of amino acid interactions and its application in the directed evolution of enantioselective enzymes. *Sci. Rep.* **8**, 16757 (2018).
39. Ma, E. J. et al. Machine-directed evolution of an imine reductase for activity and stereoselectivity. *ACS Catal.* **11**, 12433–12445 (2021).
40. Romero, P. A., Krause, A. & Arnold, F. H. Navigating the protein fitness landscape with Gaussian processes. *Proc. Natl Acad. Sci. USA* **110**, E193–E201 (2013).
41. Ao, Y. F. et al. Structure- and data-driven protein engineering of transaminases for improving activity and stereoselectivity. *Angew. Chem. Int. Ed. Engl.* **62**, e202301660 (2023).
42. Ran, X. C., Jiang, Y. Y. K., Shao, Q. Z. & Yang, Z. Y. J. EnzyKR: a chirality-aware deep learning model for predicting the outcomes of the hydrolase catalyzed kinetic resolution. *Chem. Sci.* **14**, 12073 (2023).
43. Dou, Z., Xu, G. C. & Ni, Y. A novel carboxylesterase from *Acinetobacter* sp. JNU9335 for efficient biosynthesis of Edoxaban precursor with high substrate to catalyst ratio. *Bioresour. Technol.* **317**, 123984 (2020).
44. Chen, C. S., Fujimoto, Y., Girdaukas, G. & Sih, C. J. Quantitative analyses of biochemical kinetic resolutions of enantiomers. *J. Am. Chem. Soc.* **104**, 7294–7299 (1982).
45. Kille, S. et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
46. Mazurenko, S., Prokop, Z. & Damborsky, J. Machine learning in enzyme engineering. *ACS Catal.* **10**, 1210–1223 (2020).
47. Monera, O. D. et al. Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix. *J. Pept. Sci.* **1**, 319–329 (1995).
48. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
49. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
50. Hur, S. & Bruice, T. C. The near attack conformation approach to the study of the chorismite to prephenate reaction. *Proc. Natl Acad. Sci. USA* **100**, 12015–12020 (2003).
51. 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.* **51**, 69–82 (2011).
52. Metz, S. et al. ChemShell—a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* **4**, 101–110 (2014).
53. Romano, D. et al. Esterases as stereoselective biocatalysts. *Biotechnol. Adv.* **33**, 547–565 (2015).
54. Halama, A. & Zapadlo, M. Synthesis, isolation, and analysis of stereoisomers of sacubitril. *Org. Process Res. Dev.* **23**, 102–107 (2019).
55. Lanigan, R. M. & Sheppard, T. D. Recent developments in amide synthesis: direct amidation of carboxylic acids and transamidation reactions. *Eur. J. Org. Chem.* **33**, 7453–7465 (2013).
56. David, S. E., Timmins, P. & Conway, B. R. Impact of the counterion on the solubility and physicochemical properties of salts of carboxylic acid drugs. *Drug Dev. Ind. Pharm.* **38**, 93–103 (2012).
57. Fernández-Álvaro, E. et al. A combination of in vivo selection and cell sorting for the identification of enantioselective biocatalysts. *Angew. Chem. Int. Ed. Engl.* **50**, 8584–8587 (2011).
58. Xu, G. C. et al. Hydroclassified combinatorial saturation mutagenesis: reshaping substrate binding pockets of *Kp*ADH for enantioselective reduction of bulky-bulky ketones. *ACS Catal.* **8**, 8336–8345 (2018).
59. Sun, Z. W. et al. Novel alcohol dehydrogenase CgADH from *Candida glabrata* for stereocomplementary reduction of bulky-bulky ketones featuring self-sufficient NADPH regeneration. *ACS Sustain. Chem. Eng.* **10**, 13722–13732 (2022).
60. Xu, G. C. et al. Engineering an alcohol dehydrogenase for balancing kinetics in NADPH regeneration with 1,4-butanediol as a cosubstrate. *ACS Sustain. Chem. Eng.* **7**, 15706–15714 (2019).
61. Laskowski, R. A. et al. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
62. Furche, F. et al. Turbomole. *WIREs Comput. Mol. Sci.* **4**, 91–100 (2014).

63. Lee, J. et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
64. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
65. Walker, M. et al. Performance of M06, M06-2X, and M06-HF density functionals for conformationally flexible anionic clusters: M06 functionals perform better than B3LYP for a model system with dispersion and ionic hydrogen-bonding interactions. *J. Phys. Chem. A* **117**, 12590–12600 (2013).

Acknowledgements

We are grateful to the National Key R&D Program (2019YFA0906401), the National Natural Science Foundation of China (22078127, 22378169), and the Fundamental Research Funds for the Central Universities (111-2-06) for the financial support of this research. We are thankful for the support from the high-performance computing cluster platform of the School of Biotechnology, Jiangnan University.

Author contributions

Z.D., G.X. and Y.N. designed the study. Z.D. and Xu.C. performed the experiments. X.Z., Xi.C. and J.X. analyzed the data. L.Z. performed QM/MM calculations. Z.D., S.N. and G.X. wrote the manuscript. G.X. and Y.N. provided financial support and supervised the project. All authors contributed to the manuscript.

Competing interests

All authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53191-8.

Correspondence and requests for materials should be addressed to Ye Ni or Guochao Xu.

Peer review information *Nature Communications* thanks Carlos Acevedo-Rocha, Yu-Fei Ao and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

[1]Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, 214122 Wuxi, Jiangsu, P. R. China. [2]Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, College of Pharmacy, Zhejiang University of Technology, 310014 Hangzhou, Zhejiang, P. R. China. [3]Environmental Research Institute, Shandong University, Jimo, 266237 Qingdao, Shandong, P. R. China. [4]Graduate School of Engineering, Muroran Institute of Technology, Muroran, Hokkaido 050-8585, Japan. [5]The Research Center of Chiral Drugs, Innovation Research Institute of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, 201203 Shanghai, China. ✉e-mail: yni@jiangnan.edu.cn; guochaoxu@jiangnan.edu.cn; guochaoxu@163.com