JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

pubs.acs.org/JAFC Review

Advances in the Modification of Enzymatic Properties Based on Protein Engineering Strategies

Yaoguang Huang,* Lingrong Wen, and Bao Yang*

Cite This: J. Agric. Food Chem. 2026, 74, 185−210 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Enzymes are environmentally friendly biocatalysts that play a crucial role in catalyzing biochemical reactions. However, the catalytic performance of natural enzymes often cannot fully meet the demands of industrial production. Rapidly developing AI tool-assisted protein engineering modification strategies can be used to significantly enhance the catalytic properties of enzymes. The review highlights the functional features of different protein language models and their representative tools. In addition, the system reviews various software tools used for analyzing protein catalytic performance evaluation indicators and provides a detailed comparison of the success rates and application permissions of different tools. Finally, the challenges and future directions of AI techniques in protein engineering are emphasized to strengthen the role of rational computations in protein design and personalization. This review provides a comprehensive perspective for researchers to promote further development in the field of protein engineering.

KEYWORDS: *protein engineering, thermostability engineering, enzymatic activity, enzymatic selectivity*

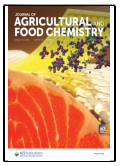
1. INTRODUCTION

Enzymes serve as biological catalysts, exhibiting high specificity, exceptional efficiency, and the capacity to operate under mild conditions. These properties grant them a paramount advantage over conventional chemical catalysts in manufacturing. As a result, enzymatic catalysis is now extensively employed in a wide array of fields, including biomedicine, environmental protection, and synthetic biology.[1] Early industrial biocatalysis predominantly utilized readily available natural wild-type (WT) enzymes, which were often repurposed from the food or other established industries for manufacturing detergents, semisynthetic antibiotics, and simple chiral precursors for pharmaceuticals. In contrast, modern engineered biocatalysts must be tailored to catalyze desired reactions, accept noncanonical substrates, and maintain stability and activity under specified process conditions.[2] Nevertheless, natural enzymes exhibit significant drawbacks under industrial conditions. Their functional stability is often compromised, as the elevated temperatures typical of many industrial processes can induce denaturation and irreversible activity loss. Moreover, their native catalytic activity is frequently inadequate to meet the demands of industrial-scale kinetics. Although high substrate specificity can be advantageous in some contexts, it also narrows their applicability by hindering the processing of non-natural or structurally diverse substrates.

Protein engineering is an established discipline that is critical to biotechnology, as it enables the rational redesign of enzyme function through structure-guided mutagenesis to overcome performance barriers in industrial applications. The catalytic mechanism of the protein and its related modification process are shown in Figure 1. Current dominant protein engineering methodologies comprise three principal strategies: directed evolution, semirational design, and rational design.[3] The comparison among the different characteristics of three strategies is exhibited in Table 1. Breakthroughs in artificial intelligence (AI) are propelling protein design beyond the constraints of natural evolution, enabling the creation of novel protein structures and functions. By leveraging the iterative refinement of machine learning algorithms, AI-driven de novo strategies circumvent the traditional dependence on natural templates. These methods autonomously optimize catalytic sites and global stability, thereby significantly enhancing enzyme performance and applicability. Furthermore, emerging protein language models (PLMs) and AI tools can generate innovative protein architectures with tailored functions, offering new technological pathways for biomolecular engineering. We compared AI tool-driven protein engineering with traditional bioinformatics-based protein engineering approaches, and the detailed information is presented in Table 2.

The future of industrial biotechnology promises to systematically overcome fundamental enzymatic constraints. This review bridges two critical, synergistic fronts: the ongoing revolution in AI-powered enzyme design tools and the persistent refinement of classical protein engineering strategies. We provide a synthesized analysis of these state-of-the-art

Received: October 8, 2025
Revised: December 29, 2025
Accepted: December 29, 2025
Published: January 5, 2026



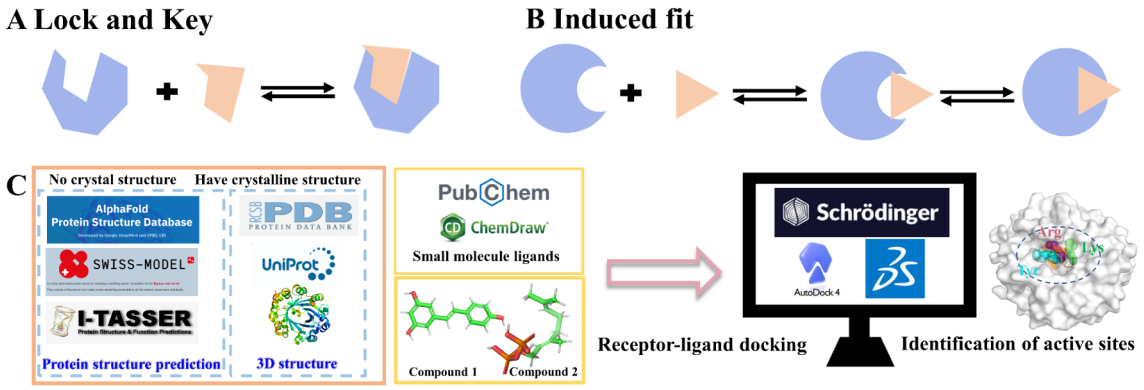


Figure 1. Enzyme catalysis theory. (A) Lock and key model, (B) Induced fit model, (C) Common procedures for enzyme modification. AlphaFold, Swiss-MODEL, and I-TASSER are protein structure prediction tools. PDB and UniProt are related databases of proteins. Schrodinger, AutoDock, and Discovery Studio are three different molecular docking tools. PubChem and ChemDraw are compound databases and compound structure drawing software, respectively.

approaches, highlighting how their integration is paving the way for a new generation of biocatalysts. This study first systematically consolidates and summarizes the landscape of PLMs and the computational tools derived from them. It then methodically delineates established strategies for enhancing key enzymatic properties-including stability, catalytic activity, and substrate selectivity. For each strategic domain, the work evaluates the associated computational and experimental tools, assessing their respective strengths and limitations, and concludes by identifying promising future trajectories for AI development in protein engineering. Finally, this review provides a critical assessment of the HTS technologies, the core concepts discussed are graphically summarized in Figure 2. In summary, this review offers a critical analysis of the prospects and challenges within contemporary enzyme engineering. It subsequently delivers an actionable framework for implementing AI-driven biocatalyst design, thereby bridging the gap between computational innovation and practical application in food processing, environmental remediation, and biomedical fields.

2. PLMS

PLMs integrate the predictive and generative capacities of AI into biological research, enabling the rational design of proteins with tailored functions.[4] PLMs leverage deep learning architectures, such as Transformers, and are trained unsupervised on extensive data sets of natural protein sequences.[5] Through this process, PLMs internalize the evolutionary patterns, structural constraints, and functional principles embedded in amino acid sequences, thus producing context-aware representations for individual residues or entire proteins. The different functions of PLMs in protein engineering are shown in Figure 3. In the discussion of this section, we divide the PLMs into 3 categories, which not only helps to understand the advantages and applicable scenarios of different models but also provides a basis for researchers to choose suitable models for specific applications. The workflow of the PLMs is shown in Figure 4. Based on this classification discussion, the characteristics of different PLMs and their applications in enzyme engineering are summarized in Table 3.

2.1. Sequence Models

These models utilize exclusively the amino acid sequences of proteins as input during both training and inference. By analyzing these sequences, the models are capable of identifying inherent patterns and extracting evolutionary signals, thereby enabling accurate prediction of how mutations affect protein function. Sequence-based models can be broadly categorized into two types: those that depend on multiple sequence alignment (MSA) and those that operate without it. MSA-dependent models derive evolutionary insights by comparing multiple related sequences, which allows them to capture coevolutionary relationships across residues. In contrast, MSA-independent models process individual sequences directly, leveraging deep learning methods to detect complex sequence motifs and long-range interdependencies within the protein. Examples of such models include AlphaMissense,[6] FusionProt,[7] and S-PLM.[8]

While MSA-dependent protein language models offer significant advantages in capturing evolutionary information, and have accumulated substantial theoretical and practical experience in predicting the effects of gene mutations, their performance is limited when handling sequences incompatible with the MSA coordinate system—such as those involving insertions or deletions. This constraint restricts the broader applicability of such models. Furthermore, a considerable number of disordered regions exist in the proteome that cannot be effectively aligned through MSA. Studies indicate that approximately 50% of human proteins contain at least one intrinsically disordered region spanning 40 amino acids or longer.[9] By contrast, MSA-independent PLMs overcome the limitations of their traditional counterparts by training directly on raw protein sequences, demonstrating considerable proficiency in predicting mutation effects. Notable examples of such models include ESM-1v,[10] ESM-1b,[11] etc.

2.2. Structure Models

Structural models elucidate the relationship between protein structure and function by capturing detailed three-dimensional conformational information. Representative examples of such models include Rosetta,[12] FoldX,[13] and GVP-GNN.[14] These models are capable not only of inferring sequences from protein structures (inverse folding) but also of establishing a novel paradigm for predicting mutation adaptability. Nevertheless, they face persistent challenges, including the need to account for protein flexibility and dynamics, and to more effectively integrate sequence and structural information. Looking ahead, incorporating molecular dynamics (MD) simulations offers a promising path to enhance prediction

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 1. Comparison of Protein Engineering Strategies

Strategies	Throughput specifications	Mutagenesis strategies	Experimental cost	Advantages	Disadvantages	Future directions	Ref.
Directed evolution	Ultrahigh throughput, >10^6 mutants to be screened	Error-prone PCR, DNA shuffling	High	Less dependent on structural information, diverse methods of mutation and screening, and the ability to combine with artificial intelligence	The high cost of experiments, the large mutant library, and the tendency to cause antagonistic effects when optimizing multiple characteristics	Developing HTS methods to enhance screening efficiency	71
Rational design	Low throughput, only 100 target mutants to be verified	Implement site-specific residue substitutions guided by the target protein's tertiary structure	Low	High-throughput efficiency and resource economy, predictive accuracy, and transcendence of natural evolutionary constraints	The necessity of relying on precise structural information of proteins, and the tendency to produce antagonistic effects when combining mutations	Streamlining experimental strategies to accelerate the "design-build-test-learn" cycle and further expand the mutant library	72
Semirational design	Moderate- to high-throughput, screening 10^3–10^5 mutants	Generate a defined set of mutants targeting computationally predicted structural hotspots	Medium	Only partial structural information is required, combined with computational tools to narrow down the mutant library	Data dependency and local optima issues, technical barriers and resource intensity, multiobjective optimization complexity	Combining with artificial intelligence to further accurately predict the effects of mutant modifications	73

accuracy, thereby jointly advancing the field of protein mutation effect prediction.

2.3. Sequence–Structure Models

Sequence-structure integrative models combine information from both sequence and structural domains, creating a unified framework that leverages multimodal data for more comprehensive and accurate predictions. By preserving the strengths of PLMs in sequence analysis while simultaneously incorporating spatial conformational information, these models achieve enhanced accuracy in predicting mutation effects. ESM-3 is an important breakthrough in sequence-structure models, and ESM-3 can generate entirely new protein sequences based on functional requirements and successfully design functional proteins with significant differences from known sequences. Common sequence-structure models include ProSST,[15] SaProt,[16] etc.

PLMs are playing an increasingly significant role in protein engineering. A current research hotspot involves modeling enzyme kinetic parameters to better understand and predict catalytic activity. Despite this prominence, practical applications of these models face several limitations. A primary challenge is the scarcity of high-quality, large-scale mutation data sets essential for robust model training and validation. Future efforts could address the issue of noisy, large-scale data by developing transfer learning strategies that integrate multiomics information. Moreover, as many model predictions rely solely on computational outputs without experimental backing, researchers must interpret these results with caution, grounding them in biological knowledge and empirical validation. Promising future directions include introducing multimodal training with structural and functional supervision, and advancing the customized design of proteins from functional or structural prompts.

3. APPROACHES FOR PROTEIN STABILITY ENGINEERING

The inherent structural instability of enzymes is a fundamental limitation for their industrial application, making stability engineering a paramount objective in protein engineering. Semirational design addresses this by combining the targeted precision of rational design with the extensive explorative capacity of directed evolution, thereby generating focused, high-quality mutant libraries. This integrated approach enables targeted structural modifications while significantly reducing screening burdens. Within this strategy, AI extracts and integrates key features from the three-dimensional protein structure, including B-factors, salt bridges, and hydrogen bond networks, to predict stabilizing mutations. This section systematically reviews key protein engineering strategies, offers a comparative analysis of their advantages in biocatalyst optimization, and critically examines the current challenges and future directions for each methodology. The established strategies for enhancing protein thermostability are outlined in Figure 5.

Currently, directed evolution requires a typical cycle of several months to years, with a screening capacity on the order of 10^3 -10^4 variants. In contrast, its ideal timeline would be reduced to several weeks to months, with a throughput exceeding 10^6 variants. Computer-aided rational design typically spans several days to weeks and handles roughly 10^1 -10^2 variants; its ideal cycle remains in the range of days to weeks, but with an improved throughput of 10^2 -10^3

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 2. Comparison of Different Protein Engineering Strategies

Comparative metrics	Traditional computational protein engineering	AI-Driven protein engineering
Representative methods/tools	Site-Directed mutagenesis	Deep learning models
	Error-Prone PCR	Convolutional neural networks (CNNs)
	Sequence alignment	
	Phylogenetic analysis	Generative adversarial networks (GANs)
	Conservation analysis	
	Statistical coupling analysis	Variational autoencoders (VAEs)
Core algorithm	Sequence/Structure alignment	Unsupervised learning of high-dimensional vector representations (embeddings) from large-scale sequence data.
	Physicochemical models	Directly predicts function from sequence or structure without the need for manually engineered features.
	Statistical analysis	Learns the underlying "grammar" of proteins to generate novel, physiologically plausible sequences.
Training data set	A limited yet high-quality experimental data set.	Massive-scale sequence databases, containing hundreds of millions to billions of sequences.
	Heavy reliance on manually curated and annotated protein family databases	Protein structure databases.
		Unlabeled sequence data, such as metagenomic data sets.
Accuracy and benchmark metrics	Rational design has a low success rate (typically <10%)	Significantly enhanced predictive accuracy.
	Directed evolution involves lengthy iterative cycles	Capable of making accurate predictions for unseen sequences or functions.
	Benchmarks: Experimentally validated fold improvement in activity/stability, change in melting temperature (T_m)	Benchmarks: root-mean-square deviation (RMSD), recall/precision, experimentally validated success rate.
Scope of application	Local optimization within known protein families	Direct mining of novel enzymes from metagenomic data.
	Heavy reliance on prior knowledge and expert input	De novo generation of proteins with specified folds or functions not found in nature.
	Limited capacity for cross-family or de novo design	Simultaneous optimization of multiple, unrelated protein properties.
		Prediction and design of protein−protein interactions.

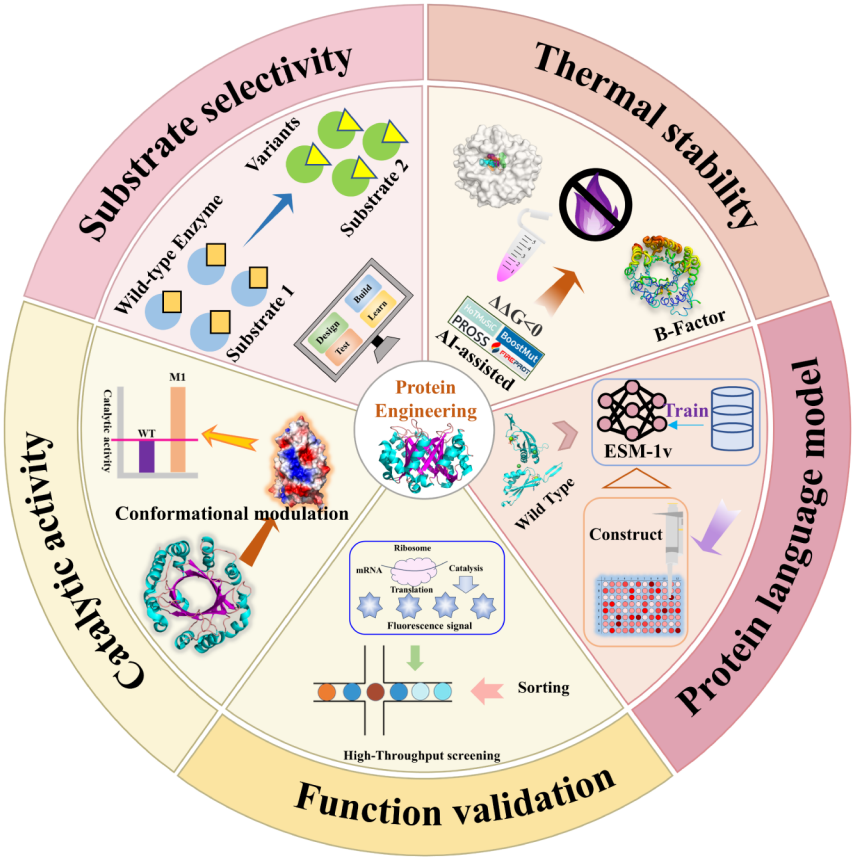


Figure 2. Key components of the review.

variants. AI-driven protein design currently operates on a time scale of days to weeks, with a typical throughput of 10^2 −10^4 variants, while the ideal target for such approaches is real-time or hourly turnaround, accompanied by a virtually infinite

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

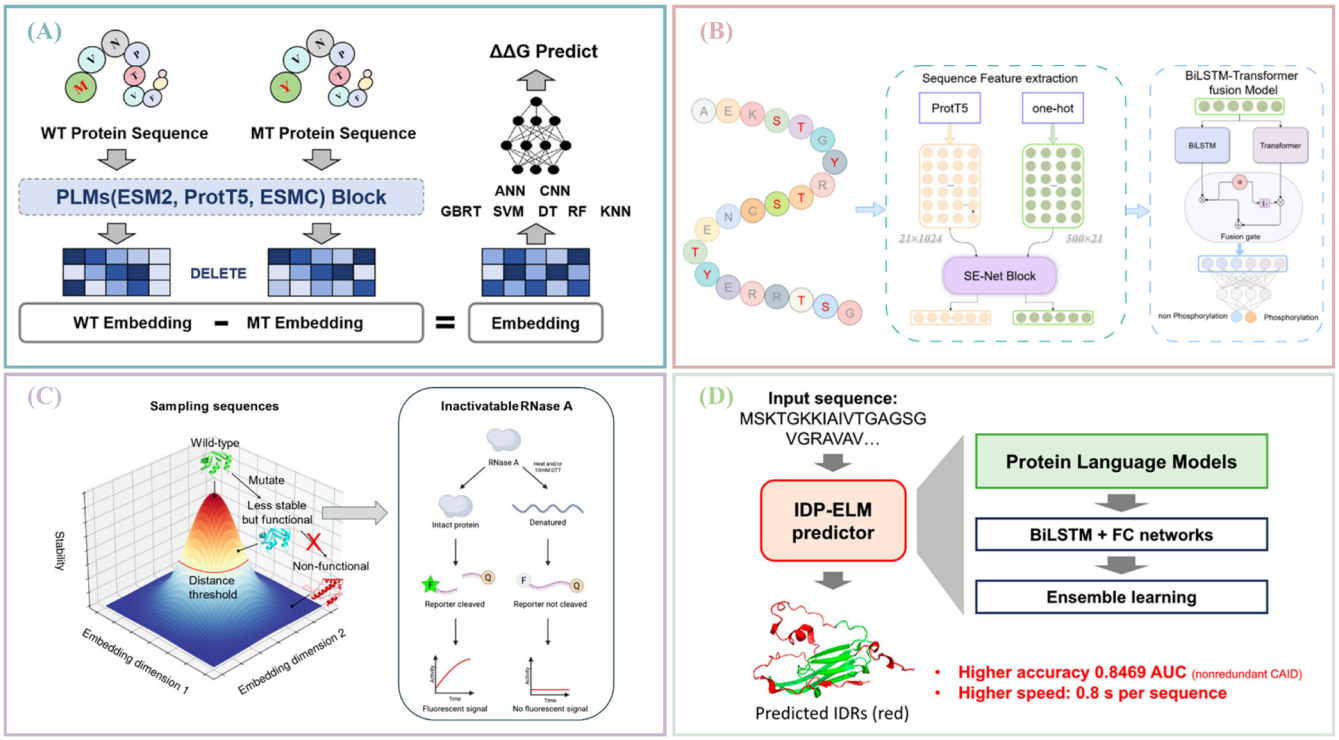


Figure 3. An overview of the different functions of the PLMs. (A) Prediction of protein−protein binding affinity changes.[142] Reprinted with permission from ref 142. Copyright 2025 American Chemical Society. (B) Determination of protein phosphorylation sites.[143] Reprinted with permission from ref 143. Copyright 2025 American Chemical Society. (C) Design of a Labile RNase A.[144] Reprinted with permission from ref 144. Copyright 2025 American Chemical Society. (D) Accurate prediction of intrinsically disordered proteins.[145] Reprinted with permission from ref 145. Copyright 2024 American Chemical Society.

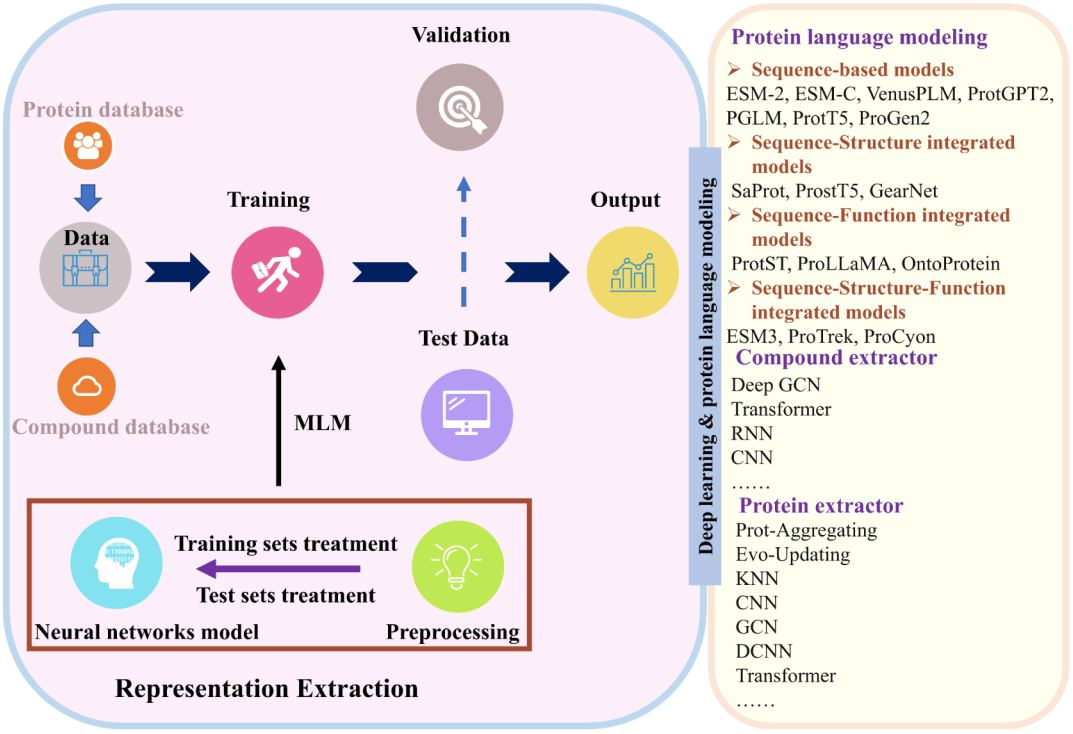


Figure 4. General workflow of PLMs. PLMs, compound feature extraction, and protein feature extraction-related content.[146] Adapted with permission from ref 146. Copyright 2025 Elsevier.

capacity for in silico design. The current cycle of enzyme thermostability engineering is long and the throughput is low, which severely limits the application of proteins in industrial production, and the future development of high-throughput methods for strengthening protein thermostability detection remains a key direction for research.[17]

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 3. Comparing the Characteristics of Different PLMs

Type	Model name	Key features	Training objectives	Primary applications	Advantages	Limitations	Ref.
Sequence models	ESM-2	Big in scale (up to 15 billion parameters), powerful in performance	Masked language modeling (MLM)	Structure prediction, function annotation, mutation effect analysis	The model has a parameter size of up to 15 billion, does not require the generation of MSA, and is open-source.	An inability to use the coevolutionary signal provided by homologous sequences	74
	ProtBERT/BFD	Training on large databases was a widely applied benchmark in early stages	MLM	Protein representation learning, downstream tasks fine-tuning	Training on large databases, providing rich protein knowledge	The architecture is conventional, without much special optimization for the protein field	75
	ProT5	Text-to-text framework, which is very flexible	Denoising self-coding	Sequence generation, translation,	Text-to-text framework based on T5, which is very flexible	High computational cost	76
	ProGen	Focusing on controlled protein generation	Autoregressive language model	De novo protein design	Designed to generate protein sequences, generating new sequences by controlling the labels	The generated sequences may reflect biases in the training data	77
	AlphaMissense	Optimized for missense mutation prediction, high practicality	MLM based fine-tuning	Clinical mutation pathogenicity assessment	Optimization for the prediction of pathogenicity of missense mutations	The predicted partial structure information has biases	78
	MSA Transformer	Utilizing homologous sequence information, strong inferential ability	Masking the language model	High-precision structure prediction, evolutionary analysis	Utilizing a relatively small amount of parameters, it demonstrates strong performance in unsupervised structure prediction tasks	High computational cost	79
	GVP-GNN	Processing vector features through a geometric vector perceptron	Protein structure quality assessment and inverse folding	Protein design	Native support for 3D data	Reliance on high quality structures	14
Structure models	S-PLM	Comparative learning is used	Aligning sequence and structure representations in the latent space	Protein clustering, function prediction	When reasoning, only the sequence	The generated structure representation is nonexplicit 3D coordinates	80
	FusionProt	Learnable fusion tokens enable iterative, bidirectional information exchange between sequence and structure encoders.	Multiview contrastive learning	Enzyme function prediction, mutation stability prediction	Fine-grained interactions between sequence and structure modes	The complexity of the model is high	7
	StructTrans/GearNet	Graph neural network	Geometric structure pretraining	Functional site prediction, protein engineering	Enriching structural information by multitask prelearning	High computational cost	81
	LM-GVP	Frameworks that integrate sequence (language model) and structure (GVP-GNN) earlier.	Prediction of protein properties	Prediction of mutation effect	Performs well on specific downstream tasks such as fluorescence intensity, stability prediction, etc.	Fusion model is too simple	82
	ProSST	Structured token sequences		Prediction of mutation effect	The structure is converted into discrete tokens, which, along with sequence tokens, are input into the standard transformer.	Fusion model is too simple	15
Sequence-structure models	ESM3	Powerful generative model	Generative mask language modeling	Mutation effect prediction, protein design	With up to 98 billion parameters	High consumption of computing resource	83
	SaProt	"Lexicalizing" structural information to achieve a deep integration of sequences and structures: masking language models	Masking the language model	Mutation effect prediction, function annotation	Incorporating 3D structures seamlessly into language model training	Reliance on external computing tools	16
	ProTrek	Exploring the evolutionary trajectories of proteins	Learning the evolutionary patterns and functional constraints of protein sequences	Functional prediction, evolutionary analysis	Insights from a dynamic evolutionary perspective	The computation is complex and the application threshold is high	84
	SST-ResNet	Effectively integrating sequence, evolutionary information, static structure, and dynamic conformation.	Protein dynamic property analysis	Mutation effect prediction, function annotation	The information is comprehensive	There is a lack of high-quality dynamic data, which affects pretraining	85

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

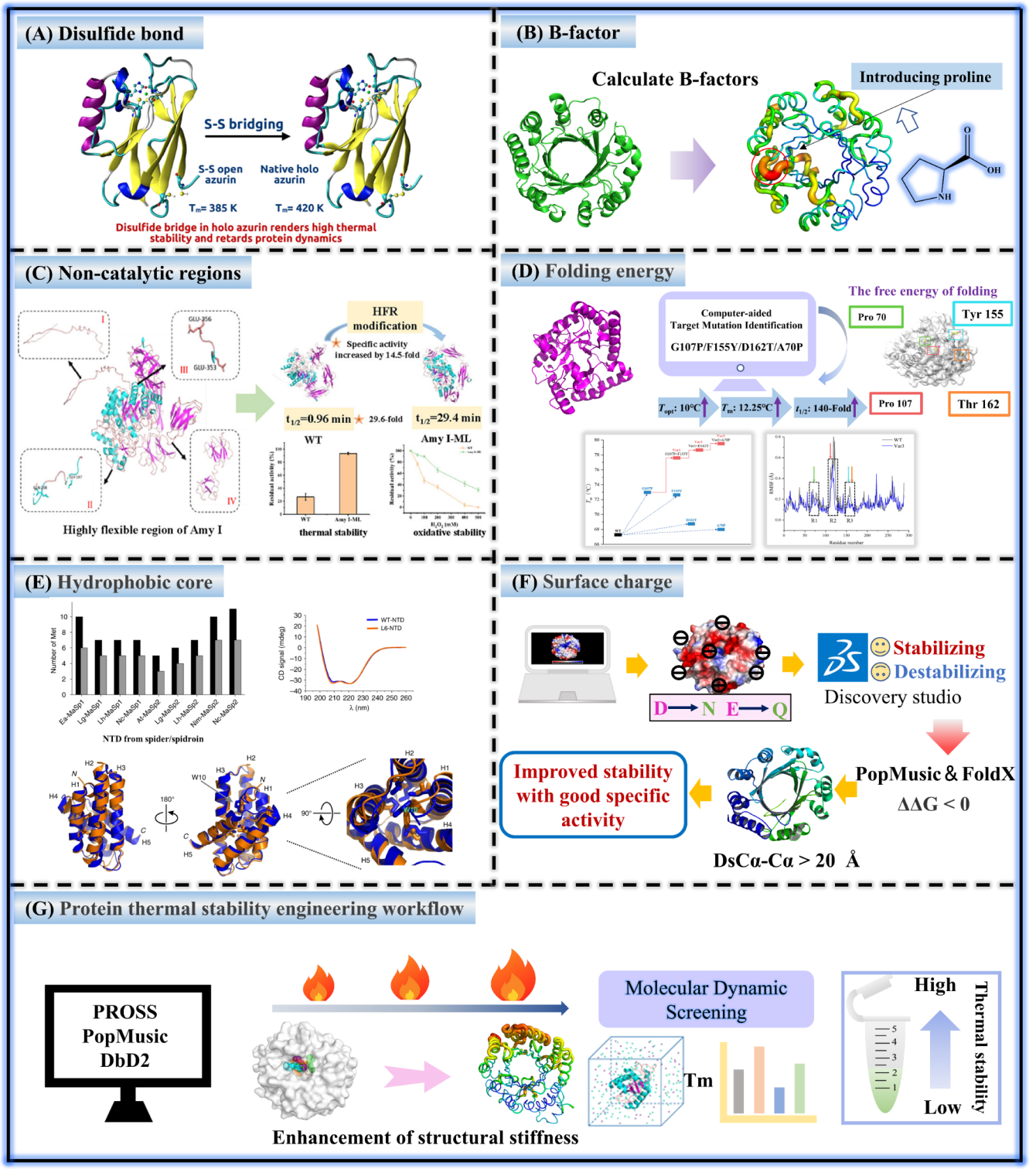


Figure 5. Tactical engineering strategies for enzyme thermostability. (A) Disulfide engineering.[147] Reprinted with permission from ref 147. Copyright 2024 American Chemical Society. (B) Proline scanning based on B-factors. (C) Increase structural rigidity of noncatalytic regions.[148] Reprinted with permission from ref 148. Copyright 2025 American Chemical Society. (D) Folding energy.[149] Reprinted with permission from ref 149. Copyright 2022 American Chemical Society. (E) Hydrophobic core.[150] Reprinted with permission from ref 150. Copyright 2019 Springer Nature. (F) Surface charge. (G) General engineering pipeline for protein stabilization.

3.1. B-Factor

The B-factor is a key parameter in protein crystal structure analysis, reflecting the amplitude of atomic positional fluctuations due to dynamic or static disorder. Notably, structurally flexible regions, such as enzyme active sites, substrate-binding pockets, and allosteric regulatory sites, typically display elevated B-factor values.[18] In enzyme engineering, such localized flexibility can be strategically constrained through the introduction of stabilizing elements

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

such as proline residues, disulfide bridges, or salt bridges, offering a rational means to enhance overall protein stability.[19] A series of thermostability engineering tools based on the PLMs have been developed using the B-factor as a measurement indicator, such as RONN, Maranas, SCHEMA, FireProt, CAN, RosettaDesign, FoldX, FRESCO, and OPUS-BFactor. Nevertheless, the B-factor provides an incomplete representation of overall protein stability. The structural stability of a protein is governed by the collective contribution of numerous weak interactions throughout its three-dimensional fold. Consequently, mutation strategies that target only individual or a limited number of high B-factor sites often yield limited improvements and can even prove counterproductive to stability.

Ban et al.[20] have improved the thermostability of a branching enzyme by engineering salt bridges. The K137E mutation increased the enzyme's half-life by 36.6%, while K137D enhanced stability by 26%. When combined with active-site mutations, the double mutants L25A/K137E and L25R/K137E increased catalytic activity by 17.12% and 31.62%, respectively. This structure-guided approach successfully enhanced thermostability without compromising enzyme function.

3.2. Disulfide Bond

A disulfide bond is a covalent linkage formed through the oxidation of thiol groups from two cysteine residues. In protein architecture, these bonds play a critical role in stabilizing the folded three-dimensional structure, improving thermal stability and resistance to denaturation, regulating functional activity, and protecting key active sites.[21] Disulfide bonds can be rationally introduced into protein structures by mutating key functional sites to cysteine residues. Computational tools, such as Discovery Studio, Disulfide by Design 2.0, and Rosetta, are often employed to guide the design and placement of these bonds. Successful incorporation is typically verified using mass spectrometry to identify linkage positions, along with biophysical and functional assays to evaluate improvements in thermal stability, enzymatic activity, or binding capacity. It should be noted, however, that newly engineered disulfide bonds may perturb the native protein conformation. Therefore, MD simulations are recommended to assess the structural impact of such mutations prior to experimental validation.

Polyethylene terephthalate (PET) is widely utilized in the textile industry, household appliances, and food packaging due to its advantageous properties, including low cost, portability, stability, excellent barrier performance, and high transparency.[22] However, its resistance to natural degradation leads to persistent environmental pollution. Enzymatic depolymerization offers a sustainable and eco-friendly strategy for the recycling of waste PET plastics. Among known enzymes, IsPETase exhibits the highest hydrolytic activity against PET under ambient conditions, though its practical application is limited by poor stability. Li et al.[23] have enhanced the performance of the high-efficiency PET hydrolase variant FAST-PETase-N212A by integrating site-directed mutagenesis with machine learning methodologies, yielding seven variants with significantly improved activity (e.g., ACC-T140D and ACC-T140E). These variants demonstrated enhanced PET degradation capability under elevated temperatures. Notably, the disulfide bond-introduced variant Pp-FAST-ACC exhibited a 10-fold increase in activity compared to the wild-type enzyme at 70 °C.

Disulfide bonds are critically important for enhancing the thermal stability of proteins. Future work could employ high-throughput experimental platforms, including microfluidic devices and fluorescence-activated cell sorting (FACS), to efficiently identify stabilized protein variants. Paradoxically, although disulfide bonds are generally considered stabilizing, their introduction may in some cases promote partially folded or unfolded intermediate states, potentially leading to structural disruption. As an alternative strategy, the rational placement of oppositely charged residue pairs at key structural positions can be employed to strengthen electrostatic interactions and improve thermostability without introducing covalent constraints.

3.3. Noncatalytic Region

The noncatalytic regions of proteins are structural segments that do not directly engage in catalytic reactions but play essential roles in maintaining structural integrity, modulating activity, facilitating protein−protein interactions or subunit assembly, and performing other auxiliary functions. Stabilizing these regions can indirectly preserve the proper conformation of the catalytic site, thereby preventing activity loss. Such an approach holds promise for extending the shelf life and enhancing the robustness of proteins in industrial or medical settings, improving stability without compromising high catalytic performance. However, this strategy is not without limitations. Excessive rigidification may impede essential conformational dynamics, thereby reducing functional efficiency, and the introduction of proline substitutions may inadvertently disrupt key molecular interactions. Future efforts could employ directed evolution or computational design strategies to precisely enhance rigidity at targeted positions, achieving an optimal balance between stability and functional flexibility.

Li et al.[24] have enhanced the thermal stability and activity of β-galactosidase Aga 0917 by introducing a proline substitution at position V140. This mutation promoted new hydrogen bonds and salt bridges, stabilized key loops, and narrowed the substrate-binding cleft. The resulting structural rigidity restricted substrate fluctuation, increased binding free energy, and shortened the catalytic distance, collectively improving the enzyme's performance at high temperatures. This study enhanced the enzyme's thermal stability by engineering its noncritical, noncatalytic regions, while fully retaining its catalytic activity.

3.4. Folding Energy

Protein stability is intrinsically linked to the folding process, wherein a well-defined folding pathway and a stable conformational state are essential for functional integrity. Folding energy-guided strategies for protein engineering typically involve identifying key folding intermediates and critical residues, improving folding kinetic efficiency, and optimizing hydrophobic core packing and hydrogen-bond networks.[25] Implementing thermostability engineering based on folding mechanisms requires integrated kinetic analysis, in which MD simulations play a central role by revealing conformational fluctuations, mapping free energy landscapes, and delineating folding pathways. Commonly employed MD packages such as AMBER, GROMACS, and NAMD are widely used for these purposes. In these simulations, the root-mean-square fluctuation (RMSF) serves as a key metric, quantifying the deviation of each residue's α-carbon (Cα) from its average position over the simulation trajectory, thus providing insight into local

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 4. Prediction Tools for Mutation-Based Protein Stability Engineering by Rational Design[a]

Tool name	Workflow	Application examples	Success rate	Ease of use and input	Advantage	Limitations	License	Computational resource requirements	Ref.
DEPROTECT-Consensus Analysis	At given positions, the corresponding consensus amino acids contribute more than the average to protein stability compared with nonconserved amino acids	The thermal stability of UDP-dependent glycosyltransferase has been improved[86]	-	User friendly, sequence	Protein structure information is not required, only the amino acid sequence is needed	Over-reliance on evolutionary conservation and ignoring functional innovation	Open source	Web server	87
Mutcompute	Identification of HydG hydrolase stabilizing mutations using a 3D self-supervised convolutional neural network	Enhanced thermostability of PET hydrolase	-	User friendly, 3D protein structure	Only PDB ID is required for prediction and visualization	Unable to provide contextualized predictions in microenvironments with atoms from cofactors ligands or nucleic acids	Open source	Web server	34
PROSS	Predicting enzyme stability mutation sites	The stability of haloalkane dehalogenase DhaA115 is enhanced[88]	The success rate of thermal stability engineering for L-rhamnose kinase (RhaB) is 80%[89]	User friendly, sequence	Thermal stability design for multiple-point mutations can be applied	Over-reliance on the accuracy of protein structure information leads to inaccurate prediction of thermal stability for proteins with large dynamic conformational changes	Open source	Web server	90
DeepDDG	Prediction of thermal stability changes caused by amino acid mutations based on neural networks	The thermal stability of carbonyl reductase LsCRM4 is enhanced[91]	The success rate of thermal stability engineering for α-carbonic anhydrases (CAs) is 84.61%[92]	User friendly, 3D protein structure	Prediction based on a large amount of experimental data and neural networks	Considering only the local environment has limitations	Open source	Web server	93
FireProt	Designing multifunctional sites while retaining the active site to enhance stability	The catalytic activity and stability of phospholipase D (PLD) were significantly improved[94]	The thermal stability engineering success rate of Inulosucrase is 11.11%[95]	User friendly, sequence	Allowing users to directly analyze and selectively modify thermostable mutants	The prediction accuracy is affected on structures obtained from low-resolution structures or homology modeling	Open source	Web server	96
Disulfide by Design 2	Designing disulfide bonds	The stability of k-carrageenase from the genus *Pseudoalteromonas* in seaweed has been improved[97]	The success rate of thermal stability engineering for the serine protease PB92 is 66.67%[98]	User friendly, 3D protein structure	It can accurately identify potential disulfide bond formation sites	There is a size limitation (5000 residues) for proteins that can be analyzed	Open source	Web server	99
Yosshi	Designing disulfide bonds	-	-	User friendly, the format of multiple sequence alignments (FASTA) and protein structures (PDB)	Introducing disulfide bonds at suitable positions based on homologous sequence alignment	The designed positions of the disulfide bonds did not consider the active sites, which may lead to loss of protein activity	Open source	Web server	100
Popmusic	Designing mutation sites based on the change in folding free energy before and after mutation	Enhanced catalytic activity and stability of Chitin-binding Lyase[101]	-	User friendly, PDB ID	The change in thermostability due to all the single point mutations can be calculated	The predictive results are greatly influenced by the information on crystal structure	Open source	Web server	102
Hotspot Wizard 3.0	Improving the stability, catalytic activity, substrate specificity, and enantioselectivity of proteins	Increased stability of superheated β-Mannanase[103]	The success rate of thermal stability engineering for the WT lactate dehydrogenase LrLDH from *Lactobacillus rhamnosus* is 10.52[104]	User friendly, sequence	Improving the stability, catalytic activity, substrate specificity, and enantioselectivity of proteins	The protein structures obtained through homology modeling have a significant impact on mutant design	Open source	Web server	64
GRAPE-WEB	Screening for the optimal mutation combination to enhance protein stability	-	The success rate of thermal stability engineering for 2648 mutants of 131 proteins is 76.7%[105]	User friendly, sequence or structure	No extensive bioinformatics knowledge required to design, check and combine stabilizing mutations	Unable to perform optimally for highly dynamic proteins	Open source	Web server	105

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 4. continued

Tool name	Workflow	Application examples	Success rate	Ease of use and input	Advantage	Limitations	License	Computational resource requirements	Ref.
FoldX	The stability changes before and after mutation based on folding free energy are presented	Improving the thermal stability of 3-Ketosteroid Δ1-Dehydrogenase 5[106]	The thermal stability engineering success rate of GH10 xylanase is 45%[107]	User friendly, 3D protein structure	Predicting destabilizing mutations more accurately	The prediction accuracy depends on the quality of the protein crystal structure	Open source	Commonly integrated with YASARA, it requires local deployment	108
FoldX	The stability changes before and after mutation based on folding free energy are presented	Improving the thermal stability of 3-Ketosteroid Δ1-Dehydrogenase 5[107]	The thermal stability engineering success rate of GH10 xylanase is 45%[107]	User friendly, 3D protein structure	Predicting destabilizing mutations more accurately	The prediction accuracy depends on the quality of the protein crystal structure	Open source	Commonly integrated with YASARA, it requires local deployment	108
I Mutant	Automatic prediction of stability changes in proteins upon single-point mutations	Improving the stability of *Rhizomucor miehei* lipase[109]	-	User friendly, sequence or structure	The prediction of the stability of mutants can be achieved solely based on the amino acid sequence	Ignoring the influence of protein stereoconformation on protein stability	Open source	Web server	91
B-Fitter	Calculate the B-factor values for each amino acid residue	Thermostability enhancement of alkaline amylase Amy I[20]	The success rate of thermal stability engineering for (+)-γ-lactamase is 54.55%[110]	User friendly, 3D protein structure	Assisting in identifying flexible regions of proteins, selecting potential mutation sites, and enhancing the efficiency of modification	Unable to predict proteins with unknown structures	Open source	Local deployment mandates high-performance computing resources	111
FRESCO	A 20−35 °C increase in enzyme stability is achieved with relatively few experimental screens	Alcohol Dehydrogenase (ADH) Stability Enhancement[112]	The success rate of thermal stability engineering (interface mutations) for ω-transaminase is 56%[113]	User friendly, 3D protein structure	High degree of automation, well suited for large-scale mutants	Unable to be applied to proteins with unknown structures	Open source	Local deployment mandates high-performance computing resources	114
ProteinMPNN	Given the backbone structure of a protein, find a corresponding amino acid sequence	Improved stability of TEV protease[115]	Thermal stability natural human myoglobin (nMb) with a 40% improvement[115]	User friendly, 3D protein structure	The modification of protein expression, stability, and function can be achieved simultaneously	The number of amino acid residues to be screened needs to be determined empirically	Open source	Local deployment is Linux-dependent	116
Rosetta	Designing mutants with significantly improved stability through two mutation strategies	The obtained combined mutant has an increased thermal stability of more than 20 °C and an expression level that is more than 10 times higher	The success rate of thermal stability engineering for glycosyltransferase UGT76G1 is 50%[117]	User friendly, 3D protein structure	Engineering protein stability under equilibrium activity conditions	Recommendations for evolutionarily nonconservative mutations affecting protein folding	Open source	Computationally intensive	118
BoostMut	Design of protein mutation Stability	-	The success rate of thermal stability engineering for limonene-1,2-epoxide hydrolase (LEH) was 42.11%[119]	Performed through MD simulations, relatively complex	It can be used for customization of protein local structures	MD simulations increase the computational cost	Open source	Computationally intensive	119
HoTMuSiC	Prediction of the effects of mutations on protein melting temperature	The thermal stability of arginine deiminase in *Enterococcus faecalis* SK23.001 has been improved[120]	The success rate of thermal stability engineering in *Bacillus thuringiensis* is 25%[121]	User friendly, 3D protein structure	Screening all possible amino acid substitutions in the target protein	The mutation data is very unbalanced toward unstable mutations	Open source	Web server	122

[a]"-" indicates no relevant success rate reported.

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

flexibility and structural stability. PROSS is a computational design server based on evolutionary analysis and structural stability, whose core logic involves introducing stabilizing mutations. It inherently incorporates AI-related concepts by integrating the Rosetta energy function for comprehensive scoring. Through training of deep learning models, this platform takes protein sequences (or mutant sequences) as input to directly predict their folding free energy or changes in thermal stability.

Tryptophan hydroxylase (TPH) enables the enzymatic production of 5-hydroxytryptophan from L-tryptophan. Wang et al.[26] have employed ΔΔG as a design criterion to obtain two TPH mutants, M1 (S422V) and M30 (V275L/I412K), which exhibited significantly enhanced thermal stability. The half-lives of these two mutants at 37 °C were 5.66-fold and 6.32-fold longer than that of the WT, respectively, while their T_m increased by 4.2 and 6.0 °C compared to WT. Tian et al.[27] have engineered *Bacillus subtilis* lipase A using an SVM model based on site-directed mutagenesis. The model analyzed 181 mutation sites and successfully identified four known stabilizing residues (G80V, G111D, M134D, N161Y), along with other predicted high-stability variants. Nevertheless, MD simulations approaches face several inherent limitations: (1) Force field inaccuracies introduce systematic biases in modeling hydrophobic interactions, hydrogen bonds, and ionic contacts, potentially compromising predictive reliability, (2) Strong dependency on experimental validation creates bottlenecks, as high-throughput experimentation remains cost-prohibitive for large-scale applications. Future advancements should focus on integrated computational-experimental closed-loop systems, leveraging automated platforms to establish iterative "simulation-experimentation-model refinement" cycles for accelerated design validation and optimization.

3.5. Surface Charge Distribution

Optimizing surface charge is a core strategy in protein stability engineering. Introducing salt bridges is an important method to improve protein stability based on the strategy of protein surface charge distribution. A salt bridge is the electrostatic attraction between positively charged cationic groups and negatively charged anionic groups in proteins, which stabilizes the natural conformation of proteins by linking different parts of the protein together. AI can directly generate surface electrostatic potential maps of proteins by training deep learning models (e.g., 3D convolutional neural networks) with protein structures as input. This approach is several orders of magnitude faster than traditional numerical computations, enabling rapid screening of massive numbers of mutants. This strategy focuses on the charged amino acids on the protein surface and can improve the thermal stability of enzymes without changing their enzymatic activity, showing promise for further application in industrial manufacturing.

Vidya et al.[28] have improved the thermostability of *E. coli* L-aspartase II through surface charge engineering. Replacing positively charged residues K139 and K207 with neutral alanine yielded variants with superior stability compared to both the wild-type enzyme and a charge-conserving mutant, demonstrating that surface charge optimization is an effective strategy for enhancing protein stability.

3.6. Calculation Tool

Current computational tools for enhancing protein stability primarily leverage principles such as multiple sequence alignment (MSA) and changes in protein folding energy. The effect of mutations on stability is commonly assessed by calculating the change in folding free energy (ΔΔG) using tools such as FoldX and Rosetta. These programs employ energy functions based on physicochemical force fields to simulate conformational changes induced by amino acid substitutions. MSA also contributes significantly to stability design: conserved residues are identified through phylogenetic analysis, and high-frequency residues are selected to improve stability via a "back-to-consensus" strategy. For instance, FireProt identifies evolutionarily stable mutations through sequence evolution analysis, which are subsequently validated with FoldX. Furthermore, by integrating three-dimensional structural features, including geometric vectors and inter-residue distances, with sequence information, machine learning models can be trained to predict the stability effects of both single and multiple mutations. Combining energy-based calculations with evolutionary analysis enables the screening of synergistic multipoint mutations while mitigating antagonistic effects, thereby supporting robust and effective protein stabilization.

Peccati et al.[29] have developed a computational approach combining AlphaFold structural ensembles with Rosetta-ddG to predict enzyme thermostability. This method accurately estimates ΔΔG values for mutants and accelerates protein engineering, reducing development cycles from years to weeks. Table 4 summarizes the applications of several common AI tools in enzyme thermostability engineering.

4. ENGINEERING APPROACHES FOR ENHANCING PROTEIN CATALYTIC ACTIVITY

Catalytic activity serves as a critical parameter for evaluating enzymatic performance. However, natural enzymes often exhibit limited catalytic efficiency owing to inherent structural and functional constraints, resulting in low product conversion rates and prolonged production cycles that restrict their industrial applicability. Enhancing enzymatic activity can markedly reduce reaction durations, decrease enzyme loading, improve product yields, and ultimately enable more efficient, environmentally friendly, and cost-effective biocatalytic processes.[30] A range of strategies, primarily leveraging protein structural and sequence information, has been developed to modify enzymatic catalytic activity, yielding substantial improvements in performance. AI utilizes machine learning regression models to predict the effects of point mutations on enzymatic activity for virtual screening. The section critically reviews several emerging approaches for enhancing enzyme activity. Currently, machine learning and computational models are increasingly applied in the engineering of enzyme catalysis and have demonstrated significant potential for advancing biotechnological applications. For example, using ProteinMPNN to design variants of TEVd with enhanced catalytic efficiency.[31] Future efforts may focus on leveraging emerging technologies, such as AI and quantum computing, to achieve unprecedented efficiency and atomic-level precision in the rational design of proteases.

4.1. Catalytic Active Site

Key catalytic residues within the enzyme active site are indispensable for enzymatic function. Based on the ligand-bound crystal structure, residues surrounding the ligand are systematically identified, and their contributions to activity are assessed via alanine scanning mutagenesis. Positions where mutations lead to substantial activity changes are subsequently

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

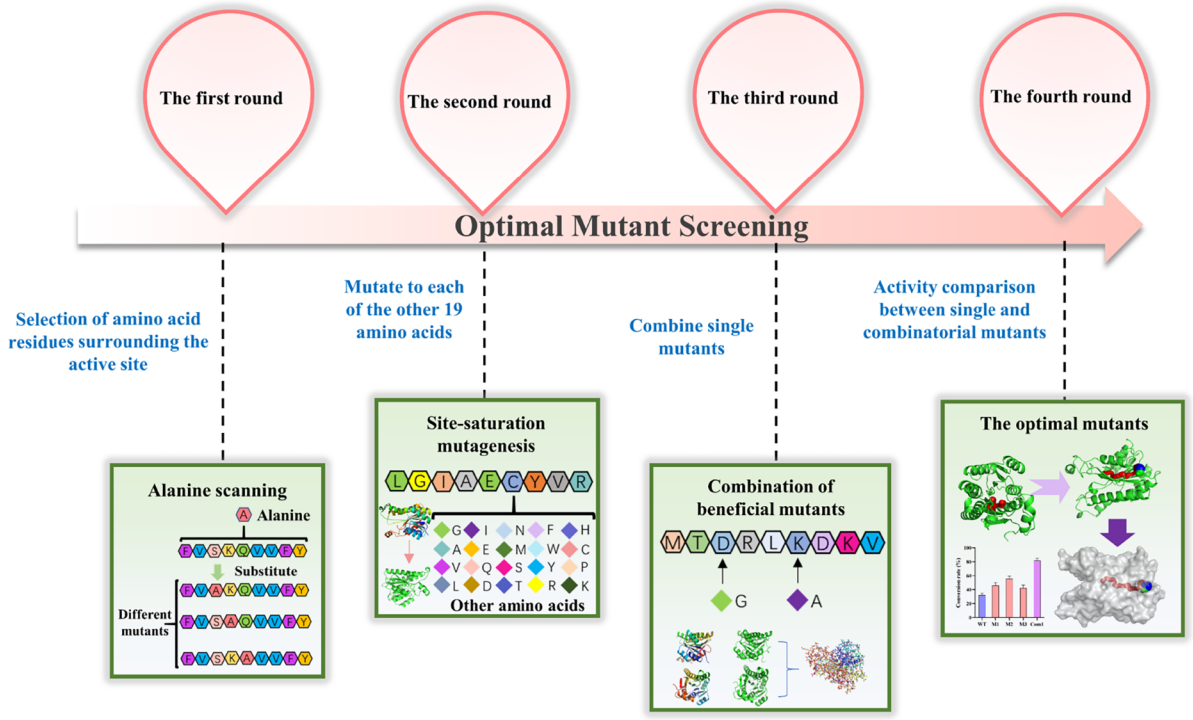


Figure 6. General workflow for enzymatic activity engineering.[151] Adapted with permission from ref 151. Copyright 2025 American Chemical Society.

subjected to saturation mutagenesis using NNK degenerate codons (N = A/C/G/T, K = G/T), which cover over 95% of all possible amino acid substitutions. For each targeted residue, approximately 100 mutant clones are typically screened, leading to the identification of variants with enhanced catalytic activity.[32]

Du et al.[33] have engineered a Baeyer−Villiger monooxygenase by mutating residues A339 and Q442 near the catalytic site. The A339E variant showed 2.4- to 3-fold higher activity toward specific cyclopentanones, while the Q442N mutant achieved a 2.7- to 3.8-fold improvement. The study highlights the need to balance activity gains with potential stability loss in enzyme engineering. Lu et al.[34] have designed hydrolase variants using a CNN-based approach, generating mutants with 3.4-fold and 29-fold higher activity than the wild-type enzyme, showcasing the potential of computational guided protein design. However, such active-site mutations often reduce enzyme stability, and the native binding pocket may be suboptimal for non-native substrates. Therefore, implementing this strategy requires carefully balancing catalytic improvement against potential destabilization. Future efforts should focus on integrated engineering approaches that simultaneously enhance activity and stability.

4.2. Conformational Regulation

The dynamic conformation of proteins is mainly explored through X-ray crystallography, cryo-electron microscopy, or time-resolved spectroscopy techniques, elucidating the conformational differences of enzymes at various catalytic stages. Substrates induce conformational changes in enzymes, making the active sites precisely match the shape of the substrate, forming the microenvironment required for catalysis. On the other hand, enzymes exist in a variety of conformational equilibria in solution, and substrate-selective binding and catalytically active conformations drive the reaction forward. By revealing the deep connection between dynamic conformation and catalytic activity, protein engineering based on dynamics and conformational regulation can be effectively applied to the modification process of enzyme catalytic activity.[35] However, excessive rigidification of enzymes may hinder product release, and distal mutations might unexpectedly affect the folding pathway or protein stability.

By examining catalytic residue dynamics across diverse enzymes, Kaczmarski et al.[36] have demonstrated that remote mutations can enhance enzyme catalysis by reshaping conformational dynamics. In the evolutionary trajectory of a dehydratase, such mutations restricted nonproductive conformations and enabled the modern enzyme to sample only catalytically competent states, highlighting the role of dynamic allostery in optimizing activity.

4.3. Sequence Alignment

In homologous protein sequence alignments, absolutely conserved residues often participate directly in catalysis, and their targeted structural and dynamic fine-tuning can enhance enzymatic activity. Sequence alignment reveals evolutionary patterns of conservation, adaptation, and coevolution, thereby identifying promising targets for catalytic optimization. Integrating structural biology with machine learning enables a synergistic combination of "local active-site optimization" and "global dynamic network reconstruction," which not only improves the efficiency of protein engineering but also facilitates the exploration of functionally relevant distant sites. AI can leverage powerful genomic databases and search tools to collect hundreds of thousands or even millions of homologous sequences, generate deep multiple sequence alignments (MSAs), subsequently extract key features from the MSA, and finally perform intelligent design based on the learned features. Nevertheless, this strategy is constrained by its reliance on high-quality homologous sequence data.

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Furthermore, mutations at evolutionarily conserved positions can inadvertently reduce enzyme stability.

The amidase from *Agrobacterium radiobacter* D3 (AmdA) catalyzes the degradation of the carcinogen ethyl carbamate in alcoholic beverages. Through targeted mutagenesis near the conserved catalytic triad of AmdA, Yao et al.[37] have developed the I97L/G195A double mutant. This variant exhibits a 3.1-fold increase in hydrolytic activity and a 1.5-fold improvement in ethanol tolerance, achieved by strengthening substrate-binding hydrogen bonds. Galmes et al.[38] have compared the active sites of two esterases and revealed that Trp104 in CalB forms a hydrophobic pocket essential for substrate positioning, analogous to phenylalanine residues in Bs2. Using a CNN model, they predicted structure-influencing mutations and demonstrated how 3D visualization of catalytic centers facilitates the design of improved catalysts.

4.4. Iterative Optimization

Iterative mutagenesis involves introducing successive rounds of mutations based on previously identified beneficial variants.[39] This strategy enables the accumulation of multiple activity-enhancing substitutions, yielding mutants with substantially improved catalytic performance. When coupled with HTS techniques, the approach facilitates the rapid identification of variants with further optimized activity. The integration of iterative mutagenesis with HTS significantly accelerates the engineering cycle, the corresponding workflow is summarized in Figure 6.

Li et al.[40] have engineered an enzyme's active site using computational iterative saturation mutagenism guided by transition-state analogs. After three rounds, a triple mutant was obtained that showed a 29.3-fold increase in catalytic activity. This TSA-guided strategy enables simultaneous optimization of multiple enzyme properties like activity and stability. However, this method necessitates the construction and screening of mutant libraries, leading to extended experimental timelines and significant resource investment. Additionally, gains in activity may sometimes coincide with reduced protein stability.

Future efforts to address this challenge may employ several integrated strategies: (1) leverage AI tools, such as DLKcat and CatPro, for preliminary screening of designed thermal stability or substrate selectivity sites to identify hotspot mutations that concurrently enhance both activity and thermostability, (2) design mutation sites related to thermal stability and substrate selectivity distal to the active pocket to minimize interference with catalytic function, and (3) systematically combine individual point mutations that independently improve activity and thermostability to generate combinatorial variants with significantly augmented performance in both attributes.

4.5. Substrate Tunnel

Substrate channel engineering is a rational protein design strategy aimed at modifying molecular transport pathways in enzymes to optimize directed substrate delivery to the active site and improve intermediate transfer efficiency. Unlike the substrate-binding pocket, the substrate access channel acts as a dedicated molecular conduit that guides ligands to the catalytic center. In natural enzymes, substrate access often depends on stochastic diffusion, which can limit reaction rates due to inefficient molecular trafficking. Through systematic optimization of these channels, substrates can be directionally guided into the active site, markedly reducing diffusion limitations and enhancing catalytic efficiency. However, the implementation of substrate channel engineering faces considerable challenges. Successful design requires the integration of computational methods, such as MD simulations and deep learning predictions, with experimental validation to precisely tailor channel architecture and function. Furthermore, structural modifications intended to alter channel properties may inadvertently compromise overall protein stability, necessitating a careful balance between catalytic efficiency and structural integrity. Progress in the practical application of this strategy will ultimately rely on the convergence of multidisciplinary expertise spanning structural biology, bioinformatics, and chemical engineering.

The spatial confinement imposed by substrate channels prevents substrate molecules from undergoing reactions at nontarget sites, such as unspecific oxidation or decomposition, thereby enhancing the regioselectivity of catalytic transformations. This spatial confinement is exemplified by cytochrome P450 enzymes, which facilitate the hydroxylation of a broad range of steroid compounds with high specificity.[41] In P450-catalyzed dihydroxylation reactions, a major constraint on synthesis efficiency arises from the accumulation of monohydroxylated intermediates at the active site. This buildup results from inefficient substrate and product transport through the access channels, hindering timely intermediate release and thereby suppressing subsequent oxidation steps. Engineering the channel architecture of P450 enzymes presents a viable strategy to alleviate this limitation and improve overall catalytic performance.[42] Deng et al.[43] have engineered a new exit channel in a P450 enzyme via a triple mutant (D182K/E143D/V178A) to facilitate intermediate release. Channel dimensions, polarity, and steric properties were key to enhancing catalytic efficiency, demonstrating a rational design strategy for enzyme pathways. Some nitrile hydratases,[44] proteases,[45] glucosidases,[46] and alcohol dehydrogenases[47] have successfully achieved activity modification engineering through substrate channels.

4.6. MD Simulation

MD simulations can identify key residues or flexible regions influencing enzymatic activity by modeling dynamic conformational changes under different conditions, thereby elucidating structure−activity relationships between enzymes and substrates. In practical applications, high-affinity mutants can be preliminarily screened through molecular docking, followed by MD simulations validation of their stability, ultimately yielding mutants with both enhanced activity and stability. The integration of AI in MD simulations for improving enzyme activity can be conceptualized as a closed-loop "design-simulate-learn-optimize" cycle. Based on MD simulations trajectories, Graph Neural Networks (GNN) analyze residue correlations (dynamic cross-correlations) and energy transfer pathways, enabling identification of "key communication pathways" from substrate binding sites to catalytic centers. However, MD simulations face challenges in integrating multiscale effects, capturing indirect regulatory roles of distal residues on active sites, and accurately modeling quantum effects or weak interactions within enzyme active sites using certain force fields. Future strategies may combine MD with other engineering approaches to achieve comprehensive and precise rational enzyme redesign.

Yin and coworkers[48] have engineered flavin-dependent enzymes through MD simulation-guided directed evolution. By structurally optimizing the flavoenzyme PtOYE from

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Parageobacillus thermantarcticus, they obtained the highly efficient variant ADes-5. This engineered catalyst exhibited a more than 70-fold enhancement in catalytic efficiency (k_{cat}/K_m) relative to the initial mutant PtOYE-Y28F, enabling the axis-selective desaturation of 1-aryl-2-tetralone. Furthermore, when integrated into a dynamic kinetic resolution process, the system achieved high-yielding synthesis of non-C2-symmetric biaryl axially chiral molecules with excellent enantioselectivity.

4.7. Additional Strategies

Researchers are increasingly adopting integrated computational toolkits for enzymatic activity engineering. By leveraging predictions of kinetic parameters, these tools enable prioritized mutant screening, significantly reducing library sizes and experimental workload. As summarized in Table 5, several key computational resources have been widely incorporated into enzyme optimization pipelines. Among them, Caver 2.0 is extensively used for mapping substrate tunnels across diverse enzyme families, enabling the identification of geometrically critical residues and providing quantitative characterization of tunnel dimensions, including surface area and volume metrics.

Zhao et al.[49] engineered a fatty acid hydroxylase by redesigning its water tunnels using Caver-guided mutagenesis. Replacing key residues with smaller or more hydrophilic ones yielded variant V456G with >13-fold higher activity, and double mutants D266A/V456A and D266T/V456G achieving 15-fold higher turnover than the wild-type enzyme, demonstrating the power of tunnel engineering in enhancing catalysis. Figure 7 presents diverse strategies for modifying enzyme activity in enzyme engineering.

5. SUBSTRATE SELECTIVITY ENGINEERING APPROACHES IN PROTEINS

Natural enzymes often exhibit promiscuous catalytic behavior, leading to undesired side reactions that necessitate selectivity engineering to minimize byproduct formation. Furthermore, industrial bioprocesses frequently require catalysis of non-native substrates, yet the inherent narrow substrate specificity of WT enzymes substantially limits their practical application.[50] Redirecting enzymatic substrate preference toward cost-effective feedstocks while preventing inadvertent catalysis of essential host metabolites represents a critical objective.[51] Substrate selectivity engineering poses significant challenges, involving multidimensional optimization of binding pocket architecture, active center geometry, substrate access tunnels, and global structural stability.[52] Engineering substrate selectivity currently involves challenges such as large mutant libraries, repetitive experimental cycles, and inconsistent catalytic results. AI models are trained on data sets of enzyme sequences, structures, and catalytic efficiency profiles to learn the complex mapping between enzyme features and substrate selectivity, enabling accurate prediction of enzymatic selectivity. With advancing accuracy in AI predictions and maturation of synthetic biology tools, precise manipulation of enzyme selectivity will transform biomanufacturing from enzyme utilization to customized biocatalyst design, unlocking new efficiencies in tailored biocatalysis. Figure 8 outlines common engineering strategies in this field.

5.1. Key Residue Design

Enzyme substrate selectivity has been engineered through strategic remodeling of the active site pocket geometry. Representative approaches include: introducing bulky side-chain residues (Y/F substitutions for G/A) to construct constricted pockets that sterically exclude large substrates, coupled with removal of sterically hindering residues (L→G mutations) to expand the cavity for accommodating bulky or branched substrates. This methodology constitutes an established paradigm for reprogramming protein substrate specificity. β-strands typically pack against the protein core on one face while exposing the opposite face to solvent. Solvent-exposed residues constitute major contributors to protein stability, with amino acid preferences at each sequence position dictated by the precise local backbone geometry and identity of adjacent residues. This structural dependency demonstrates that the most effective strategy for designing β-sheet surfaces requires integrative consideration of multiple energetic factors-including side-chain rotamer preferences, van der Waals interactions, electrostatic contributions, and desolvation effects. Under this strategy, AI enables accurate prediction of enzyme−substrate relationships by integrating cross-attention mechanisms with 3D structural information.

Loop regions play critical roles in substrate binding, guiding substrates into the active site, and facilitating conformational transitions. By engineering loops above the active site of cumene dioxygenase, Heinemann et al.[53] have created variants with modified substrate channels. These mutations generated new access pathways, enabling the enzyme to catalyze novel reactions with three distinct substrates and successfully reprogramming its specificity. However, this approach also presents certain limitations. Altering key amino acid residues can only achieve fine-tuning of substrate selectivity and fails to realize fundamental interconversion between two distinct substrates. Furthermore, modifications in substrate preference may subtly alter the geometric configuration or electrostatic environment of the active site, resulting in significant reductions in catalytic efficiency (k_{cat}). Future research could transition from single-point mutations to combinatorial multisite mutagenesis targeting mutation networks. Additionally, by integrating structural information with sequence alignment data, one could first rationally identify target regions, then perform random or semirandom mutagenesis within these regions through directed evolution, allowing the screening process to identify optimal variants.

5.2. Directed Evolution

Current directed evolution strategies for reprogramming protein substrate selectivity continue to rely on the construction of large mutant libraries. High-throughput screening (HTS) significantly mitigates the experimental burden and time investment required for library evaluation. Representative HTS platforms encompass absorbance- and fluorescence-based enzymatic activity detection, microfluidic droplet sorting for single-cell analysis, and FACS of cell-surface-displayed enzymes.

The yeast endoplasmic reticulum sequestration screening (YESS) system combines protease engineering with broad substrate profiling, facilitating substantial improvements in catalytic efficiency alongside enhanced sensitivity and dynamic range throughout the protease optimization process.[54] Through this evolutionary platform, the engineered TEV protease variant eTEV achieved a 2.25-fold increase in catalytic efficiency compared to its parent form TEV-EAV, significantly broadening its applicability in biocatalytic processes.

5.3. Dynamic and Conformational Engineering

The introduction or substitution of proline residues within active site loops can markedly influence their conformational

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 5. Advanced Tools for Enhancing Enzymatic Catalytic Performance

Tool name	Workflow	Application examples	Success rate	Ease of use	Advantage	Limitations	License	Computational resource requirements	Ref.
Caver 3.0	Visualization of substrate access tunnels in enzymes	The activity of D-amino acid oxidase (DAAO) is significantly increased[123]		User friendly, straightforward to operate	Automatic computation of static and dynamic tunnels and channels in protein structures	Failure to predict the dynamic tunnel	Open source	Locally deployed tool, integrated with PyMOL and computationally efficient	124
FuncLib	Predictive stability engineering of enzymes via single-site mutagenesis	The activity of cold-active pullulanase is significantly increased[125]	The activity engineering success rate of nonspecific peroxidase (MthUPO) reached 100%[126]	User friendly, straightforward to operate	Multiple-point mutations for the design of active sites in enzymes that are not amenable to HTS	Computations are performed without placing the substrate molecule structure, leading to the construction of a nonspecific sequence library	Open source	Web server	127
DLKcat	Prediction of k_{cat}	The activity of L-glutaminase is significantly increased[128]		User friendly, straightforward to operate	Combining protein sequence and substrate molecular structure information can predict the k_{cat} value for different enzyme and substrate combinations	It is impossible to predict enzyme-catalyzed reactions involving multiple substrates	Open source	Local deployment	129
ESP	Computational prediction of small molecule substrates		The success rate of enzyme−substrate prediction for 8 halogenases and 78 substrates reached 58.3%[130]	User friendly, 3D protein structure	When predicting enzymes with large homology differences, good predictive results can be achieved	For substrates that were not present in the training data set, prediction performance was substantially reduced	Open source	Local installation with a high-performance GPU	131
DiffDock	Simulation of ligand−receptor docking and generation of interaction complexes	The molecular docking affinity of various per- and polyfluoroalkyl substances (PFAS) for blood proteins was evaluated[132]	A 38% TOP1 success rate was achieved on PDBbind.[133]	User friendly, sequence and 3D protein structure	High accuracy in docking flexible ligands to novel targets	Lack of experimental data to provide feedback on prediction accuracy	Open source	Web server	134
CataPro	Prediction of enzymatic kinetic parameters: k_{cat}, K_m	The enzyme SsCSO, identified using CataPro, exhibited 19.53-fold higher activity than the initial enzyme (CSO_2)[135]	The predicted success rate of BH1352 from *Bacillus halodurans* is 80%[135]	Sequence and compound structure	Compared with the traditional model, the prediction accuracy is significantly improved	The catalytic reactions of enzymes and multiple substrates cannot be accurately predicted	Open source	Local deployment	135
CatPred	Prediction of enzymatic kinetic parameters: k_{cat}, K_m, and inhibition constant (K_i)		The predicted K_m value of human hexokinase for D-glucose differs by approximately 11% from the experimentally determined value[136]	The foundational knowledge in computation, the protein sequence and substrate structure	For molecules outside the training data set, there still has a high predictive performance	The kinetic parameters prediction does not consider environmental factors such as pH and temperature, requiring a larger scope and higher quality data set for pretraining	Open source	Local deployment	136
UniKP	Computational prediction of enzymatic kinetic parameters: k_{cat}, K_m		Among the 20 enzymes tested, 60% exhibited a positive correlation[137]	Requires basic computer skills, protein sequence and the substrate structure	The kinetic parameters of enzymes can be predicted solely based on enzyme sequences and substrate structures	The predicted k_{cat} values are not accurate enough	Open source	Web server	137
DeepMolecules	Computational prediction of enzymatic kinetic parameters: k_{cat} and K_m		The accuracy on the independent test data is 94.2%[138]	User friendly, protein sequence and the compound structure	Integrating predictive modeling and experimental data to predict protein-small molecule interaction forces and enzyme kinetic parameters	The prediction accuracy of non-natural substrates and protein mutants needs to be improved	Open source	Web server	138
MPEK	Computational prediction of enzymatic kinetic parameters: k_{cat} and K_m			User friendly, protein sequence and the compound structure	Prediction of K_m and k_{cat} with temperature, pH and organism information	The prediction performance for sequences with large differences from the pretraining data set is poor, and it cannot directly repre-	Open source	Web server	139

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Table 5. continued

Tool name	Workflow	Application examples	Success rate	Ease of use	Advantage	Limitations	License	Computational resource requirements	Ref.
						sent the structure of enzyme–substrate complexes			
EnzyHTP	For designing mutants with enhanced catalytic activity		The entire simulation workflow for 100 mutants was completed within 7 h using 10 GPUs and 160 CPUs	Requires basic computer skills, sequence and the compound structure	Integrating high-throughput activity testing and structural analysis, the optimal multienzyme system structure was quickly designed	The accuracy of predicting the enzyme kinetic parameters of mutants with substrates needs to be improved	Open source	Computationally intensive	140
iMARS	The synergistic contributions of multiple modules enhanced the functionality of the multienzyme system.	The production of resveratrol increased by 45.1 times, and the production of raspberry ketone increased by 11.3 times[141]		User friendly, sequences of multiple enzymes	Not only for multienzyme cascade optimization but also for the rearrangement of metabolic flux in different organisms	The accuracy of different prediction tools in the iMARS framework may affect its performance, and the pretraining data set is relatively limited	Open source	Web server	141

dynamics, thereby modulating substrate binding behavior. This strategy provides a means to functionally optimize enzymes by finely balancing stability and evolvability.[48,55] Catalytically important loops contribute to substrate guidance, binding, and conformational transitions. Computational tools including PSIPRED, JPred, and AlphaFold 2 enable accurate loop region identification. The capacity to reprogram substrate selectivity arises from reshaping the enzyme's conformational landscape, underscoring the role of structural adaptability in functional plasticity. During this process, AI utilizes known allosteric sites as training data to enable the model to learn the structure-dynamic features of allosteric sites, thereby predicting potential allosteric regulatory sites in novel proteins.

By computationally reshaping the conformational landscape of aspartate aminotransferase, St-Jacques et al.[56] have shifted its open-closed dynamic equilibrium. This multistate design approach yielded a variant with a 1900-fold change in substrate selectivity and improved catalytic efficiency for a non-native substrate. This approach has proven effective in protein selectivity engineering. Nevertheless, the inherent dynamics of enzyme structures complicate the identification and engineering of key regulatory residues, as catalysis often depends on coordinated motions across domains. Moreover, conventional design methods predominantly rely on static structural models, limiting their ability to target dynamic regulatory mechanisms. Future advances should integrate enhanced sampling techniques such as Metadynamics with machine learning methods like AlphaFold-Multimer to identify critical dynamic residues and design mutations that conformationally lock desired states.

5.4. Domain Replacement

This engineering methodology constructs hybrid enzymes by recombining distinct functional domains, including substrate recognition regions, catalytic domains, and allosteric regulatory modules. A representative implementation of this approach is domain swapping, in which native substrate recognition modules are replaced by analogous domains derived from heterologous proteins. For instance, in several recent studies, researchers have inserted fusion tags into auxiliary domains to enhance binding affinity toward specific substrates. This strategy facilitates precise reprogramming of protein substrate selectivity. However, it relies on the availability of comprehensive structural and functional data sets. Its successful implementation therefore requires close integration of experimental and computational workflows. Future advances should focus on improving the predictive accuracy of engineered variants through synergistic combination of protein structure determination and rational design methodologies.

Through targeted mutations inducing a domain displacement, Kraus et al.[57] have successfully reprogrammed sucrose phosphorylase to accept bulky polyphenolic acceptors like resveratrol and quercetin. The redesigned binding site, stabilized by new $\pi-\pi$ stacking and hydrophobic interactions, enabled efficient glycosylation of these aromatic substrates, demonstrating how structural dynamics can be harnessed to alter catalytic specificity.

5.5. Cofactor-Binding Site

Cofactor binding plays a critical role in determining enzyme substrate selectivity. Through rational design or directed evolution of the cofactor binding site, it is possible to precisely modulate cofactor affinity, optimize steric compatibility within the substrate access channel, and fine-tune the catalytic microenvironment, thereby enabling customized enhancement

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

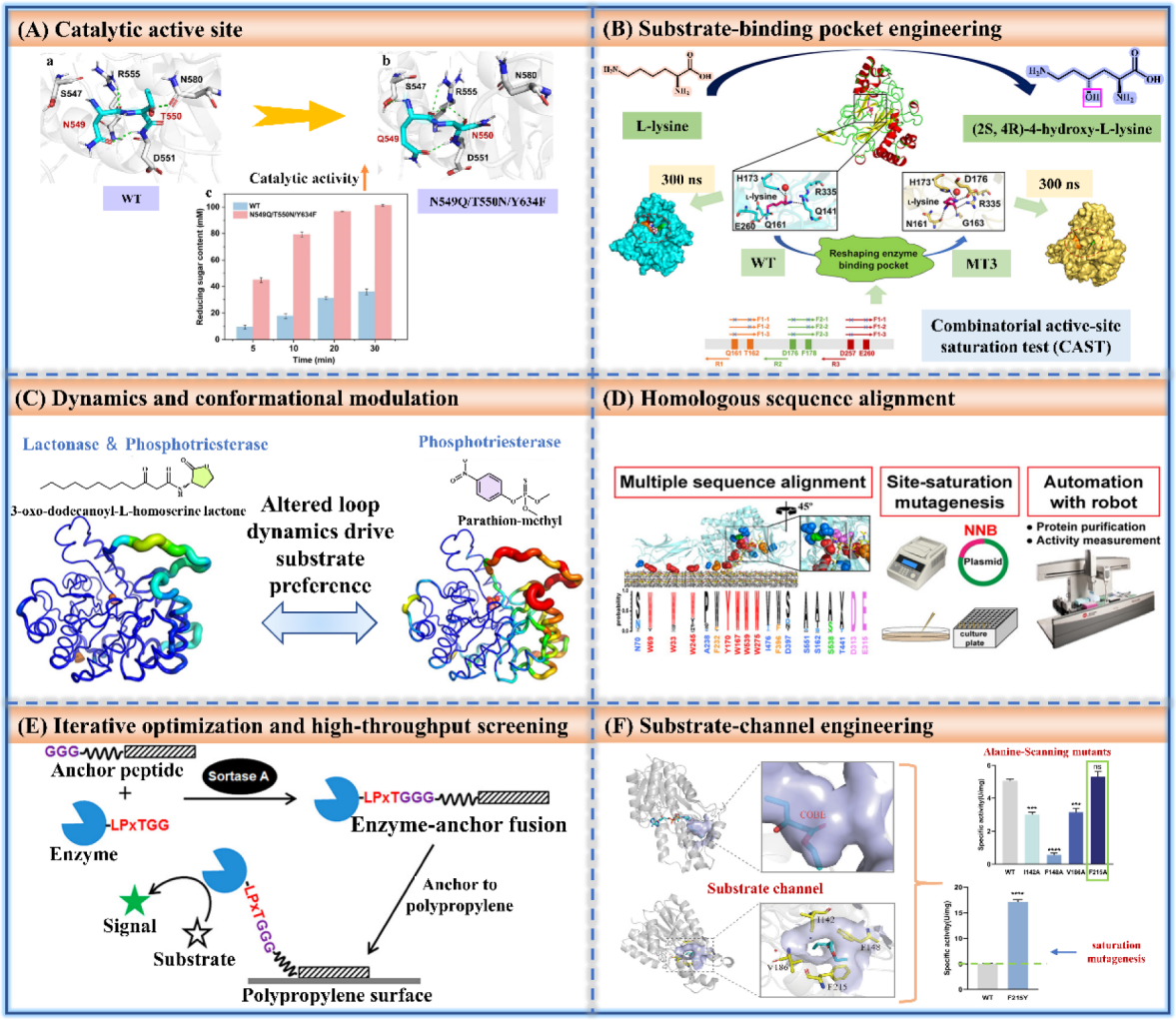


Figure 7. Workflow of enzymatic activity engineering strategies. (A) Key catalytic active sites.[152] Adapted with permission from ref 152. Copyright 2025 American Chemical Society. (B) Substrate-binding pocket engineering.[153] Reprinted with permission from ref 153. Copyright 2020 American Chemical Society. (C) Dynamics and conformational control.[154] Reprinted with permission from ref 154. Copyright 2024 American Chemical Society. (D) Homologous sequence alignment.[155] Reprinted with permission from ref 155. Copyright 2020 American Chemical Society. (E) Iterative optimization and HTS.[156] Reprinted with permission from ref 156. Copyright 2018 American Chemical Society. (F) Substrate tunnel engineering.[47] Reprinted with permission from ref 47. Copyright 2025 American Chemical Society.

of enzymatic selectivity. Cofactors, which commonly include metal ions and coenzymes, improve catalytic efficiency and specificity by stabilizing transition states and enabling accurate substrate positioning during catalytic cycles. In this process, AI employs GNN to analyze cofactor binding pockets. By integrating rapid force-field tools such as FoldX, it enables large-scale computation of the effects of thousands of virtual mutations on cofactor binding affinity (ΔΔG), thereby facilitating efficient screening of mutations that stabilize cofactor binding without compromising overall structural integrity.

Pan et al.[58] have engineered a hydroxysteroid dehydrogenase by mutating residues T15 and R16 in the cofactor-binding site. The variants T15A, R16A, and R16Q showed 7.85-fold, 2.50-fold, and 4.35-fold increases in catalytic activity, respectively, and shifted substrate preference. The enhanced performance resulted from restructured hydrogen-bonding networks and altered interactions within the catalytic triad. Ge et al.[59] have engineered *Burkholderia cepacia* lipase using HTS and AI-driven methods. By analyzing key biophysical properties, residue L167 was identified as having the greatest impact on enantioselectivity. The resulting variants have demonstrated a significant reversal of enantioselectivity from S- to R-preference. However, such selectivity enhancement often entails a trade-off with catalytic activity. This limitation may be addressed through CAST to optimize multiple residues simultaneously. Additionally, synthetic cofactors frequently exhibit steric incompatibility with natural enzyme binding pockets. One strategy to mitigate this issue involves expanding the cavity volume by replacing bulky residues with smaller ones such as glycine, thereby introducing greater flexibility and accommodating non-natural cofactors.

5.6. Substrate Channel

As critical structural determinants, substrate channels significantly contribute to enzymatic substrate selectivity and catalytic efficiency. Precision mutagenesis of channel-lining residues allows fine-tuning of pathway geometry and hydrophobicity, facilitating substrate discrimination by size or chemical properties. Coupling channel engineering with active-site redesign enables dual control over substrate specificity. A common tactic involves introducing hydrophobic

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

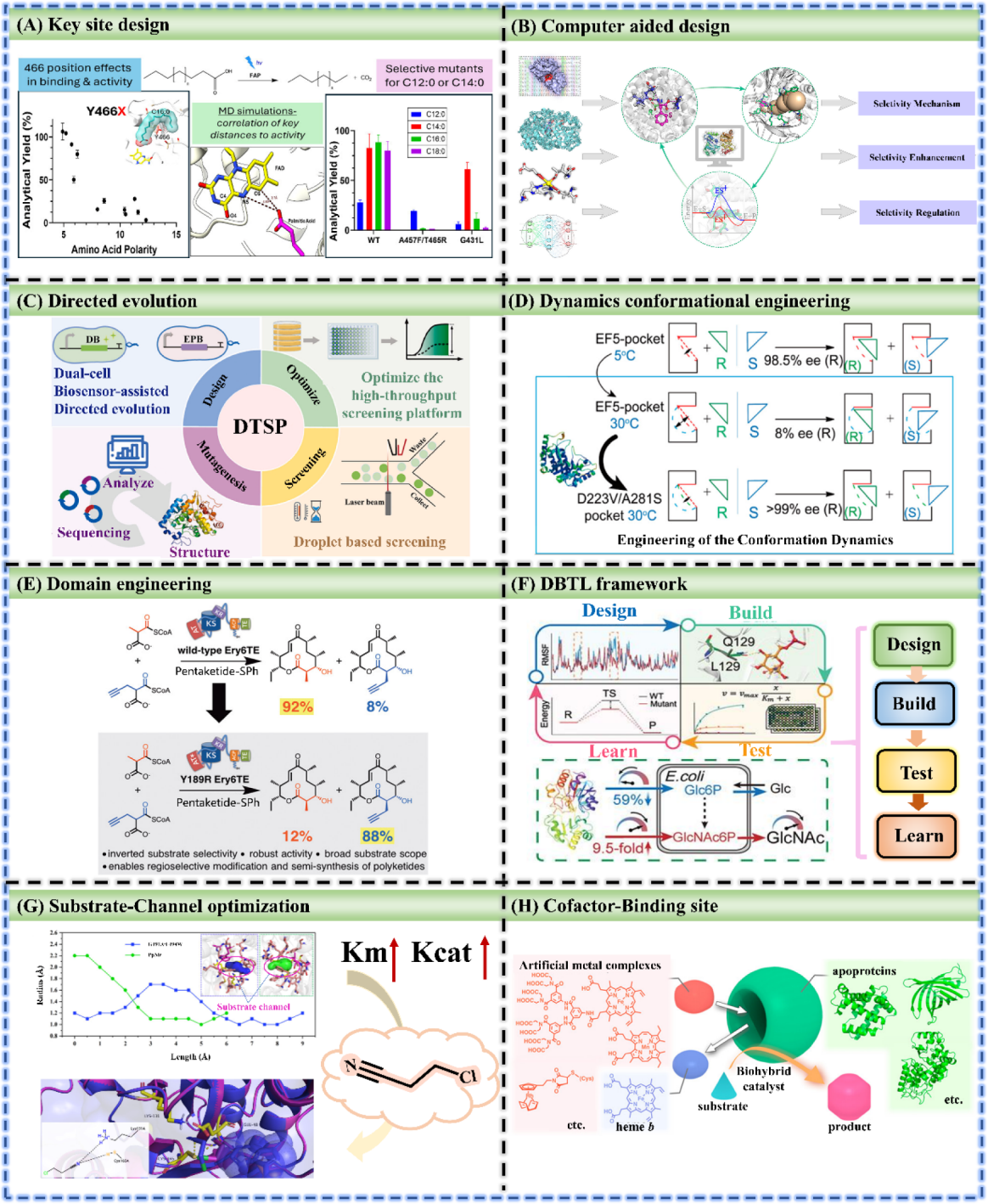


Figure 8. Substrate selectivity engineering strategies for enzymes. (A) Key residue design.[157] Reprinted with permission from ref 157. Copyright 2024 American Chemical Society. (B) Computational tool-guided prediction and optimization.[158] Reprinted with permission from ref 158. Copyright 2022 American Chemical Society. (C) Directed evolution with HTS.[159] Reprinted with permission from ref 159. Copyright 2025 American Chemical Society. (D) Dynamic control and conformational engineering.[160] Reprinted with permission from ref 160. Copyright 2017 American Chemical Society. (E) Domain engineering.[161] Reprinted with permission from ref 161. Copyright 2017 American Chemical Society. (F) Integrated DBTL framework for iterative refinement.[162] Reprinted with permission from ref 162. Copyright 2024 John Wiley and Sons. (G) Substrate channel optimization.[163] Adapted with permission from ref 163. Copyright 2025 American Chemical Society. (H) Cofactor-binding site engineering.[164] Reprinted with permission from ref 164. Copyright 2019 American Chemical Society.

residues to promote selective enrichment of nonpolar substrates like aromatic hydrocarbons. Nevertheless, such modifications risk interfering with catalytic centers or distal functional regions. Future progress may capitalize on deep learning models to evaluate substrate-channel compatibility, alongside reinforcement learning frameworks to autonomously optimize channel architecture and chemical characteristics. AI trains models using features extracted from molecular simulations or experimental data via machine learning regression models to predict the effects of mutations on

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

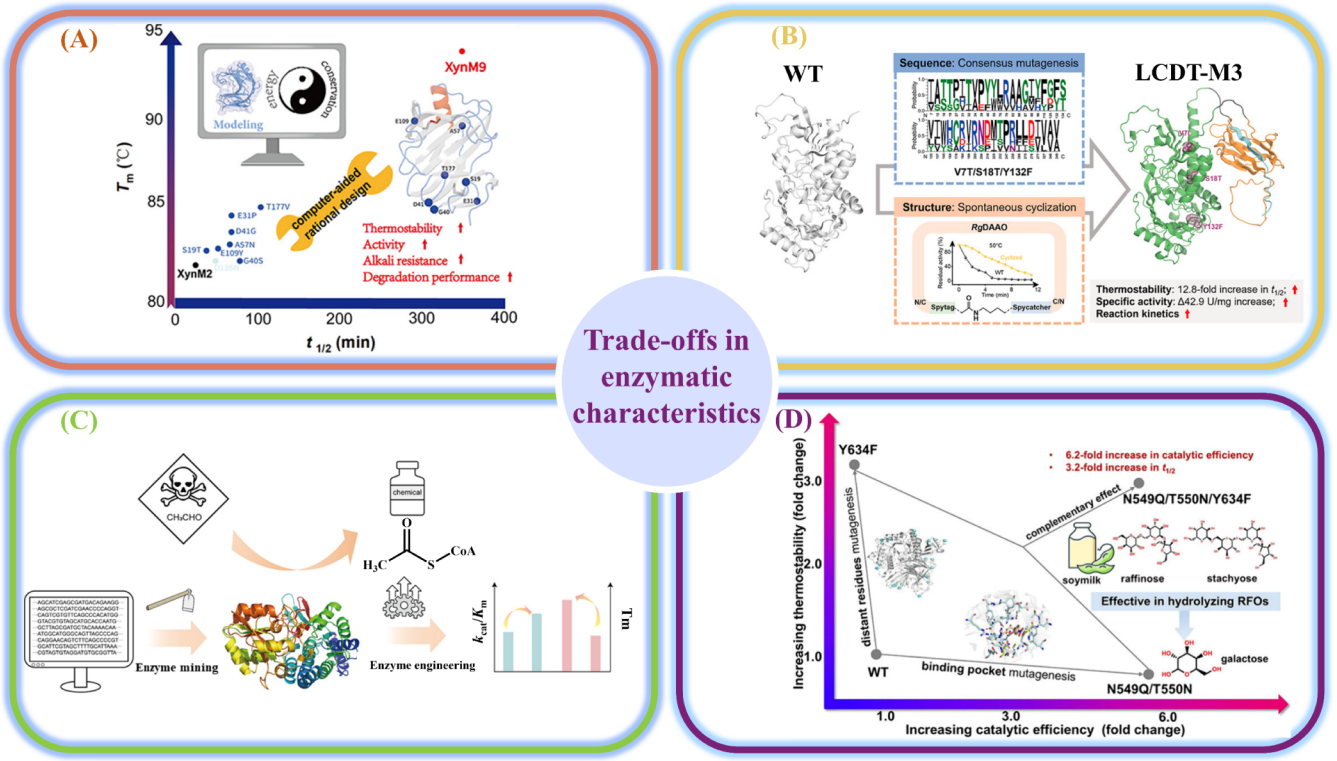


Figure 9. Trade-off strategies for engineering different enzymatic properties. (A) Enhancement of xylanase activity and thermostability through the combined analysis of folding free energy and evolutionary conservation.[165] Reprinted with permission from ref 165. Copyright 2024 American Chemical Society. (B) Improvement of D-amino acid oxidase thermostability and catalytic efficiency by integrating consensus sequence design with structural modifications.[166] Reprinted with permission from ref 166. Copyright 2025 American Chemical Society. (C) Simultaneous engineering of the thermostability and activity of a novel aldehyde dehydrogenase.[63] Reprinted with permission from ref 63. Copyright 2025 American Chemical Society. (D) Elevation of α-galactosidase catalytic activity and thermal stability by selectively targeting and engineering distal catalytic microenvironments and residues.[152] Reprinted with permission from ref 152. Copyright 2025 American Chemical Society.

channel permeability. Additionally, an end-to-end deep learning model can be trained to take the structures of WT and mutant variants as input and directly output relative permeability scores for target substrates versus interfering substrates. Nevertheless, channel engineering risks compromising catalytic turnover due to altered substrate diffusion kinetics.

Lu et al.[60] have engineered a benzoic acid decarboxylase by reshaping its substrate channel. The double mutant Q302Y/I303Y, designed using CAVER-guided analysis, retained nearly 100% activity toward benzoic acid while reducing conversion of the larger cinnamic acid by over 90%, demonstrating how channel remodeling can selectively control substrate specificity through steric hindrance.

6. OPTIMIZATION TRADE-OFFS AMONG DIFFERENT ENZYMATIC PROPERTIES

A fundamental challenge in enzyme engineering is the recurrent trade-off between activity and stability. Efforts to enhance catalytic activity often result in reduced stability, whereas strategies to improve structural stability may compromise enzymatic function. This inverse relationship stems from the intricate and often competing requirements for optimal enzyme structure and function.[61] Optimal catalytic efficiency requires precisely tuned structural flexibility that permits necessary conformational changes for substrate accommodation and transition state stabilization. However, when flexibility exceeds functional requirements, it destabilizes the catalytic architecture, leading to inefficient substrate orientation and compromised intermediate stabilization, thereby reducing enzymatic efficiency.[62] Enzymes require considerable structural rigidity to maintain their structural integrity and functional competence. While a rigid architecture contributes to stability under extreme conditions, including high temperatures, nonphysiological pH, and organic solvent environments, excessive rigidity may limit conformational flexibility and thereby impair catalytic activity. Consequently, improvements in stability through rigidification often involve a trade-off with catalytic efficiency, the molecular basis of which remains incompletely understood in protein science. To balance enzyme activity and stability, several strategies are commonly employed: (1) targeted modification of active site residues to enhance catalytic performance, (2) engineering regions distal to the active center to minimize functional disruption, (3) rational redesign of flexible structural segments to improve thermal stability without compromising catalytic function. The relevant schematic diagram is shown in Figure 9.

By integrating deep learning and rational design, Xu et al.[63] have engineered an adenosine deaminase variant, P443C, which showed 80.7% higher residual activity after thermal challenge and achieved 93.2% substrate conversion efficiency, successfully enhancing both thermostability and catalytic performance. Through semirational design, Zhao et al.[64] have developed two RrCuZnSOD double mutants, D25/A115T and A115T/S135P, which exhibited 1.2-fold and 1.6-fold longer half-lives at 80 °C, along with melting temperature

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

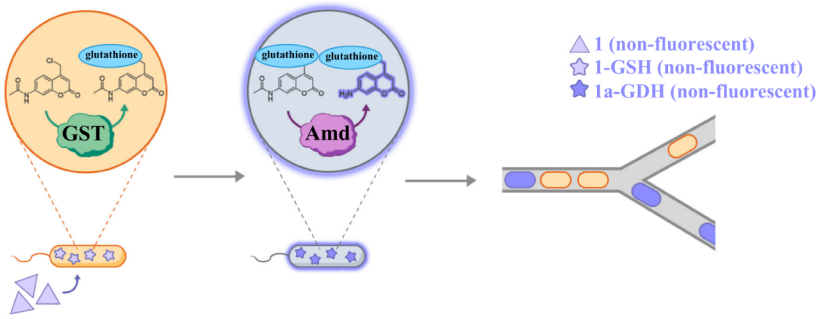


Figure 10. Overview of the FACS-based assay principle. Fluorescence histograms of *E. coli* BL21(DE3)-gold cells expressing GST only as a negative control.[167] Reprinted with permission from ref 167. Copyright 2025 American Chemical Society. 7-Acetamido-4-chloromethylcoumarin (AACMC, 1) conjugates with glutathione (GSH) via glutathione S-transferase (GST) to form the intermediate product 1-GSH, which remains intracellular. When the amide bond is cleaved by amidase, it generates the fluorescent target product 1a-GSH. Since neither the initial compound 1 nor the intermediate product 1-GSH exhibit fluorescence, this characteristic enables us to screen cells with higher amidase activity.

increases of 3.4 °C and 2.5 °C, respectively, while fully retaining wild-type level catalytic activity.

In practical enzyme engineering, strategic modulation of structural dynamics is often employed: flexibility around the active site can be increased to facilitate substrate binding, while rigidity is enhanced within the core regions or secondary structural elements to improve thermostability. This balance can be further optimized by introducing salt bridges to reinforce structural integrity, while preserving catalytic efficiency through the careful design of flexible loop regions near the catalytic pocket. Using site-saturation mutagenesis, residues surrounding the active site or other critical structural zones are systematically replaced to assess their individual contributions to enzymatic activity and stability. In certain application contexts, a marginal reduction in catalytic activity may be deemed acceptable when offset by substantial gains in operational stability.

7. HTS PLATFORMS

The comprehensive screening of large enzyme libraries with sufficient coverage demands the development of specialized protein engineering platforms. Such systems must provide dramatically higher throughput than conventional microtiter plate methods while reliably maintaining the genotype–phenotype linkage essential for directed evolution. This integrated approach can substantially decrease both the time and cost associated with mutant screening.[65] The schematic diagram of HTS in the enzyme engineering process is shown in Figure 10. Ultrahigh-throughput screening strategies have been successfully implemented for the identification and engineering of diverse enzyme classes, including esterases, PET hydrolases, carboxylases, and NAD(P)-dependent oxidoreductases, through the use of enzyme-class-specific fluorescence-based activity assays.[66] Common HTS technologies include optical screening techniques, flow cytometry FACS, and droplet microfluidics.

Droplet-based microfluidic systems have emerged as powerful HTS tools that are transforming enzyme engineering practices.[67] Droplet microfluidics technology refers to transferring traditional biochemical reactions and assays, which are conventionally conducted in test tubes or microplates, into individual microdroplets with uniform size and volume ranging from picoliters to nanoliters, where each droplet functions as a miniature reactor. Su et al.[68] have employed droplet-based microfluidics (DMFS) to screen for thermostable laccases. The DMFS platform incorporated a heating module and fluorescence-activated droplet sorting (FADS) to screen a large laccase variant library, yielding 12 variants with enhanced thermostability. This study highlights the importance of DMFS for effectively identifying functionally beneficial distal amino acid substitutions, facilitating the discovery of distal advantageous amino acid mutations by researchers. However, this technique also has certain limitations. For instance, it requires extremely high droplet stability. The droplets must remain stable throughout the entire experimental process to ensure uniform reagent concentration within each droplet. Additionally, it is not suitable for applications involving long-term cell culture or screening of intercellular interactions.

FADS has rapidly developed in recent years into a powerful ultrahigh-throughput screening platform, widely applied in the selection of enzymes, metabolites, and antibodies. A key enabling technology for expanding the scope of FADS is the use of high-sensitivity fluorescence-coupled strategies, which link enzymatic activity to detectable fluorescent signals. This coupling allows quantitative monitoring of target molecules and forms a critical technological foundation for advancing FADS implementations. Among these strategies, fluorogenic substrate assays are one of the most commonly employed methods. In this design, a fluorophore is covalently linked to a quenching group, effectively suppressing fluorescence. Upon enzymatic cleavage of the substrate, the quencher is released, restoring fluorescence emission. The resulting signal intensity correlates directly with enzymatic activity, enabling quantitative HTS.[69] For instance, Agresti et al.[70] have implemented the commercial substrate Amplex Red within a droplet microfluidic system to conduct directed evolution of horseradish peroxidase across ten distinct populations. Their screening campaign identified a variant displaying a 10-fold enhancement in catalytic efficiency relative to the WT enzyme, achieving a k_{cat}/K_m value of $2.5 \times 10^7\ M^{-1}{\cdot}s^{-1}$, a level approaching the theoretical diffusion limit.

Single-cell screening within monodisperse microdroplets generated by microfluidic devices has emerged as a powerful ultrahigh-throughput strategy for identifying superior enzyme variants. Compartmentalization of individual cells in droplets enables efficient phenotype-genotype linkage by confining extracellular target molecules, allowing rapid enrichment of desirable mutants from extensive libraries while reducing screening time and cost by orders of magnitude. However, several technical challenges remain, including the precise

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

fusion control of multicomponent droplets for reaction initiation and the critical need to prevent cross-contamination between compartments. Future advancements may employ hydrophobicity-modulated substrate delivery systems to spatially confine fluorescence signals within droplets, thereby enabling precise spatiotemporal control over reaction initiation and detection.

8. TRENDS IN RATIONAL PROTEIN ENGINEERING

Rational protein engineering has emerged as a central research discipline bridging synthetic and structural biology. Its developmental trajectory is progressively advancing from initial single-function optimization toward increasingly intelligent, dynamic, and multifunctional protein systems. The integration of machine learning with protein design methodologies is poised to fundamentally accelerate innovation in enzyme engineering. Looking forward, AI-guided rational engineering strategies are anticipated to evolve along multiple parallel pathways, driving the next generation of biocatalyst development.

(1) The deep integration of AI and machine learning will be central to advancing enzyme engineering. While current predictive models primarily analyze static protein−ligand docking poses, future efforts should prioritize simulating dynamic conformational changes before and after ligand binding. In parallel, enhancing the accuracy of enzyme sequence-function predictions will be crucial to minimize reliance on experimental trial and error, thereby streamlining the design process.

(2) The widespread implementation of automation and high-throughput experimental platforms will significantly advance enzyme engineering. Integrating HTS techniques, such as Adaptive Laboratory Evolution, with automated Workstations enables accelerated design-build-test-learn cycles and facilitates rapid mutant library screening. Future developments should prioritize advancing high-throughput methodologies specifically tailored for evaluating enzyme thermostability, substrate selectivity, and catalytic activity.

(3) A critical future direction involves systematically expanding experimental databases for diverse enzyme characteristics. Integrating enzymatic kinetic parameter databases with protein language models will enable the development of specialized AI tools for predicting enzyme kinetic properties. Such integration is expected to significantly enhance the accuracy of existing prediction models, thereby reducing the size and screening burden of mutant libraries.

9. CONCLUSIONS AND FUTURE OUTLOOK

Continuously evolving protein engineering strategies are enabling the rational modification of enzymatic properties in various proteins, facilitating their utility across multiple domains. In food processing, this technology will find further applications in precision production of specific flavor peptides, improvement of food texture, extension of shelf life, inhibition of microbial growth, efficient degradation of food contaminants and toxins. In environmental bioremediation processes, it enables efficient depolymerization of various plastic wastes and degradation of specific pollutants in industrial wastewater. During biobased material production, agricultural byproducts can be converted into high-value materials, achieving closed-loop recycling of biobased materials.

However, the current application of artificial intelligence in enzyme engineering faces several challenges that need to be addressed to broaden its practical utility. A critical factor in artificial neural networks is the number of neurons in the hidden layer, where an insufficient number leads to underfitting, while an excessive number results in overfitting due to increased model complexity. Controllable generation of functionally specific protein sequences remains technically challenging. Furthermore, the application of protein language models continues to encounter difficulties in data acquisition and modeling, particularly in functional data collection and evolutionary analysis of enzymes with multiple substrates. Future research will focus on developing protein tools that are more user-friendly and cost-effective, enabling the design of increasingly complex proteins and the exploration of novel chemical structures incorporating noncanonical amino acids. AI-assisted enzyme engineering modifications will further advance enzyme design and development of multifunctional enzyme preparations, thereby promoting technological upgrading and sustainable development in the food industry and green chemistry sector.

AUTHOR INFORMATION

Corresponding Authors

Bao Yang − *Guangdong Provincial Key Laboratory of Applied Botany, Key State Laboratory of Plant Diversity and Specialty Crops, Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China; University of Chinese Academy of Sciences, Beijing 100049, China;* orcid.org/0000-0002-5483-7960; Email: yangbao@scbg.ac.cn

Yaoguang Huang − *Guangdong Provincial Key Laboratory of Applied Botany, Key State Laboratory of Plant Diversity and Specialty Crops, Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China; University of Chinese Academy of Sciences, Beijing 100049, China;* orcid.org/0000-0001-8290-8749; Email: huangyg@scbg.ac.cn

Author

Lingrong Wen − *Guangdong Provincial Key Laboratory of Applied Botany, Key State Laboratory of Plant Diversity and Specialty Crops, Key Laboratory of National Forestry and Grassland Administration on Plant Conservation and Utilization in Southern China, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China; University of Chinese Academy of Sciences, Beijing 100049, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jafc.5c13434

Author Contributions

Y.H.: Writing—original draft. L.W.: Visualization, Supervision, Investigation. B.Y.: Writing—review and editing, Methodology, Conceptualization.

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

Funding

This work was supported by the Guangdong S&T Program (No. 2024B1111160001).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Zhang, R.; Yan, X.; Gao, L.; Fan, K. Nanozymes expanding the boundaries of biocatalysis. *Nat. Commun.* **2025**, *16* (1), 6817.
(2) Buller, R.; Lutz, S.; Kazlauskas, R. J.; Snajdrova, R.; Moore, J. C.; Bornscheuer, U. T. From nature to industry: harnessing enzymes for biocatalysis. *Science*, **2023**, *382*, 6673. .
(3) Wang, M.; Medarametla, P.; Kronenberger, T.; Deingruber, T.; Brear, P.; Figueroa, W.; Ho, P. M.; Krueger, T.; Pearce, J. C.; Poso, A.; et al. Pseudomonas aeruginosa acyl-CoA dehydrogenases and structure-guided inversion of their substrate specificity. *Nat. Commun* **2025**, *16* (1), 2334.
(4) Zhang, Q.; Chen, W.; Qin, M.; Wang, Y.; Pu, Z.; Ding, K.; Liu, Y.; Zhang, Q.; Li, D. F.; Li, X. J.; et al. Integrating protein language models and automatic biofoundry for enhanced protein evolution. *Nat. Commun.* **2025**, *16* (1), 1553.
(5) Ruffolo, J. A.; Madani, A. Designing proteins with language models. *Nat. Biotechnol.* **2024**, *42* (2), 200−202.
(6) Minton, K. Predicting variant pathogenicity with AlphaMissense. *Nat. Rev. Genet.* **2023**, *24* (12), 804.
(7) Kalifa, D.; Singer, U.; Radinsky, K. FusionProt: fusing sequence and structural information for unified protein representation learning. *bioRxiv* **2025**.
(8) Wang, D.; Pourmirzaei, M.; Abbas, U. L.; Zeng, S.; Manshour, N.; Esmaili, F.; Poudel, B.; Jiang, Y.; Shao, Q.; Chen, J.; et al. S-PLM: structure-aware protein language model via contrastive learning between sequence and structure. *Adv. Sci.* **2024**, *12*, 2404212.
(9) Zhang, L.; Tan, P.; Hong, L. Research progress in mutation effect prediction based on protein language models. *Sheng Wu Gong Cheng Xue Bao*. **2025**, *41* (3), 934−948.
(10) Meier, J.; Rao, R.; Verkuil, R.; Liu, J.; Sercu, T.; Rives, A. Language models enable zero-shot prediction of the effects of mutations on protein function *bioRxiv*. DOI: .
(11) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (15), No. e2016239118.
(12) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031−3048.
(13) Delgado, J.; Reche, R.; Cianferoni, D.; Orlando, G.; van der Kant, R.; Rousseau, F.; Schymkowitz, J.; Serrano, L. FoldX force field revisited, an improved version. *Bioinformatics* **2025**, *41* (2), btaf064.
(14) Gou, W.; Ge, W.; Tan, Y.; Li, M.; Fan, G.; Yu, H. CPE-Pro: A Structure-Sensitive deep learning method for protein representation and origin evaluation. *Interdiscip Sci.* **2025**, 1−16.
(15) Li, M.; Tan, P.; Ma, X.; Zhong, B.; Yu, H.; Zhou, Z.; Ouyang, W.; Zhou, B.; Hong, L.; Tan, Y. ProSST: protein language modeling with quantized structure and disentangled attention. *bioRxiv* **2024**.
(16) Su, J.; Han, C.; Zhou, Y.; Shan, J.; Zhou, X.; Yuan, F. SaProt: orotein language modeling with structure-aware vocabulary. *bioRxiv* **2023**.
(17) Li, L.; Liu, X.; Bai, Y.; Yao, B.; Luo, H.; Tu, T. High-Throughput screening techniques for the selection of thermostable enzymes. *J. Agric. Food Chem.* **2024**, *72* (8), 3833−3845.
(18) May, M.; Chuah, A.; Lehmann, N.; Goodall, L.; Cho, V.; Andrews, T. D. Functionally constrained human proteins are less prone to mutational instability from single amino acid substitutions. *Nat. Commun.* **2025**, *16* (1), 2492.
(19) Xu, K.; Fu, H.; Chen, Q.; Sun, R.; Li, R.; Zhao, X.; Zhou, J.; Wang, X. Engineering thermostability of industrial enzymes for enhanced application performance. *Int. J. Biol. Macromol.* **2025**, *291*, 139067.
(20) Ban, X.; Wang, T.; Fan, W.; Li, C.; Gu, Z.; Cheng, L.; Hong, Y.; Li, Z. Thermostability and catalytic ability enhancements of 1,4-α-glucan branching enzyme by introducing salt bridges at flexible amino acid sites. *Int. J. Biol. Macromol.* **2023**, *224*, 1276−1282.
(21) Yu, K.; Chen, L.; Tang, Y.; Ma, A.; Zhu, W.; Wang, H.; Tang, X.; Li, Y.; Li, J. Enhanced thermostability of nattokinase by rational design of disulfide bond. *Microb. Cell Fact.* **2025**, *24* (1), 51.
(22) Mican, J.; Jaradat, D. S. M. M.; Liu, W.; Weber, G.; Mazurenko, S.; Bornscheuer, U. T.; Damborsky, J.; Wei, R.; Bednar, D. Exploring new galaxies: perspectives on the discovery of novel PET-degrading enzymes. *Appl. Catal., B* **2024**, *342*, 123404.
(23) Li, L. X.; Bai, Y. G.; Zhang, Y. F.; Qin, X.; Wang, X. L.; Zhang, J.; Wang, Y. R.; Luo, H. Y.; Yao, B.; Tu, T. Rational design of an engineered zearalenone lactone hydrolase with enhanced thermostability and acid tolerance for zearalenone degradation in food. *Food Chem.* **2025**, *497*, 147005.
(24) Li, T.; Cui, N.; Fan, J.; Liu, X.; Guan, J.; Zhang, W.; Li, Y.; Gao, Q.; Liu, Y.; Xu, Y. T.; Wang, L.; Guo, C. J.; Wang, Y. Engineering β-agarase Aga0917 from *Pseudoalteromonas fuliginea* YTW1−15−1 to improve the thermostability and enzyme activity. *ACS Sustainable Chem. Eng.* **2022**, *10* (47), 15437−15449.
(25) Hu, D.; Wang, Y.; Jiang, H.; Pan, H.; Bian, Y.; Ren, W.; Chen, H.; Guo, Z.; Wang, Y. Effect of Additional Terminal Residues on the Folding and Unfolding Dynamics of Cold Shock Protein. *Adv. Sci.* **2025**, *12*, No. e01369.
(26) Wang, Y.; Liu, W.; Peng, S.; Chen, Y.; Chen, F.; Zhang, A.; Chen, K. Enhancing thermostability of tryptophan hydroxylase via protein engineering and its application in 5-hydroxytryptophan production. *Int. J. Biol. Macromol.* **2024**, *264*, 130609.
(27) Tian, F.; Yang, C.; Wang, C.; Guo, T.; Zhou, P. Mutatomics analysis of the systematic thermostability profile of Bacillus subtilis lipase A. *J. Mol. Model.* **2014**, *20* (6), 2257.
(28) Vidya, J.; Ushasree, M. V.; Pandey, A. Effect of surface charge alteration on stability of l-asparaginase II from *Escherichia sp. Enzyme Microb. Technol.* **2014**, *56*, 15−19.
(29) Peccati, F.; Alunno-Rufini, S.; Jiménez-Osés, G. Accurate Prediction of enzyme thermostabilization with Rosetta using AlphaFold ensembles. *J. Chem. Inf. Model.* **2023**, *63* (3), 898−909.
(30) Yu, H.; Wang, Y.; Yang, Z.; Ying, J.; Guan, F.; Liu, B.; Miao, M.; Mohamed, A.; Wei, X.; Yang, Y. J.; Liu, X.; Sun, L. F.; Jiang, Z. Q.; Yang, S. Q.; Xin, F. J. Enhancing the synthesis efficiency of galacto-oligosaccharides of a β-galactosidase from *Paenibacillus barengoltzii* by engineering the active and distal sites. *Food Chem.* **2025**, *483*, 144208.
(31) Nieuwkoop, T.; Claassens, N. J.; van der Oost, J. Improved protein production and codon optimization analyses in Escherichia coli by bicistronic design. *Microb. Biotechnol.* **2019**, *12* (1), 173−179.
(32) Wang, F.; Ma, X.; Sun, Y.; Guo, E.; Shi, C.; Yuan, Z.; Li, Y.; Li, Q.; Lu, F.; Liu, Y. Structure-Guided engineering of a protease to improve its activity under cold conditions. *J. Agric. Food. Chem.* **2023**, *71*, 12528−12537.
(33) Du, Y.; Lv, X.; Siebenmorgen, T.; Popowicz, G.; Feng, C.; Ma, Y.; Wang, Y. Enhancing catalytic activity of a baeyer−villiger monooxygenase from Oceanicola granulosus: simultaneous engineering of the distal site and active site. *J. Agric. Food. Chem.* **2025**, *73* (23), 14453−14466.
(34) Lu, H.; Diaz, D. J.; Czarnecki, N. J.; Zhu, C.; Kim, W.; Shroff, R.; Acosta, D. J.; Alexander, B. R.; Cole, H. O.; Zhang, Y.; Lynd, N. A.; Ellington, A. D.; Alper, H. S. Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* **2022**, *604* (7907), 662−667.
(35) Zheng, L. E.; Barethiya, S.; Nordquist, E.; Chen, J. Machine learning generation of dynamic protein conformational ensembles. *Molecules* **2023**, *28* (10), 4047.
(36) Kaczmarski, J. A.; Mahawaththa, M. C.; Feintuch, A.; Clifton, B. E.; Adams, L. A.; Goldfarb, D.; Otting, G.; Jackson, C. J. Altered

conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme. *Nat. Commun.* **2020**, *11* (1), 5945.
(37) Yao, X.; Kang, T.; Pu, Z.; Zhang, T.; Lin, J.; Yang, L.; Yu, H.; Wu, M. Sequence and structure-guided engineering of urethanase from *Agrobacterium tumefaciens* d3 for improved catalytic activity. *J. Agric. Food. Chem.* **2022**, *70* (23), 7267−7278.
(38) Galmés, M.; Nödling, A. R.; Luk, L.; Świderek, K.; Moliner, V. Combined Theoretical and Experimental Study to Unravel the Differences in Promiscuous Amidase Activity of Two Nonhomologous Enzymes. *ACS Catal.* **2021**, *11* (14), 8635−8644.
(39) Raps, F. C.; Jin, C. A.; Brown, A. C.; Sorigué, D.; Hyster, T. K. Stereoselective photoenzymatic hydroarylation for the construction of quaternary stereocenters. *J. Am. Chem. Soc.* **2025**, *147* (44), 40547−40553.
(40) Li, J.; Wang, S.; Liu, C.; Li, Y.; Wei, Y.; Fu, G.; Liu, P.; Ma, H.; Huang, D.; Lin, J.; Zhang, D. Going beyond the local catalytic activity space of Chitinase using a simulation-based iterative saturation mutagenesis strategy. *ACS Catal.* **2022**, *12* (16), 10235−10244.
(41) Yi, G.; Zou, H.; Long, T.; Osire, T.; Wang, L.; Wei, X.; Long, M.; Rao, Z.; Liao, G. Novel cytochrome P450s for various hydroxylation of steroids from *filamentous fungi*. *Bioresour. Technol.* **2024**, *394*, 130244.
(42) He, P.; Li, H.; Sun, J.; Zhang, X.; Gong, J.; Shi, J.; Xu, Z. Identification of a fungal cytochrome P450 with steroid two-step ordered selective hydroxylation characteristics in Colletotrichum lini. *J. Steroid Biochem. Mol. Biol.* **2022**, *220*, 106096.
(43) Deng, Q.; Lu, Z. M.; Yuan, Z.; Feng, Y.; Zhang, L.; Shi, J.; Xu, Z.; Kofas, M. A. G.; Li, H. Rationally designing P450BM3-H to excavate a novel channel for product exit and enhance overall performance. *Int. J. Biol. Macromol.* **2025**, *307*, 142162.
(44) Li, S. F.; Gao, Y. C.; Xu, H. B.; Xu, C. L.; Wang, Y. J.; Liu, Z. Q.; Zheng, Y. G. Substrate access tunnel engineering of a Fe-type nitrile hydratase from *Pseudomonas fluorescens* ZJUT001 for substrate preference adjustment and catalytic performance enhancement. *Bioorg. Chem.* **2024**, *152*, 107744.
(45) Ma, X.; Wang, L.; Chen, J.; Guo, E.; Zheng, H.; Zhao, L.; Lu, F.; Liu, Y. Surfactant-tolerance evolution of *Bacillus clausii* protease for enhancing activity and stability by reshaping the substrate access tunnel. *Colloids Surf., B* **2025**, *245*, 114289.
(46) Wang, J.; Chen, J.; Lv, K.; Gao, Z.; Li, J.; Wu, B.; He, B.; Schenk, G. Structure-Guided engineering unveils deeper substrate channel in processive endoglucanase EG5C-1 contributing to enhanced catalytic efficiency and processivity. *ACS Synth. Biol.* **2024**, *13* (12), 4131−4142.
(47) Zhou, P.; Wu, M.; Ma, L.; Li, Y.; Liu, X.; Chen, Z.; Zhao, Y.; Li, Z.; Zheng, L.; Sun, Y.; Xu, Y. B.; Liu, Y. P.; Li, H. Engineering alcohol dehydrogenase for efficient catalytic synthesis of Ethyl (R)-4-Chloro-3-hydroxybutyrate. *J. Agric. Food. Chem.* **2025**, *73* (18), 11146−11156.
(48) Yin, H.-N.; Chen, Z.; Zhao, X.; Liu, Z.; Yu, F.; Zheng, S.; Huang, N.; Liu, Z. Molecular-dynamics-simulation-guided directed evolution of flavoenzymes for atroposelective desaturation. *Nat. Synth.* **2025**, *4*, 1621−1629.
(49) Zhao, P.; Kong, F.; Jiang, Y.; Qin, X.; Tian, X.; Cong, Z. Enabling peroxygenase activity in cytochrome P450 monooxygenases by engineering hydrogen peroxide tunnels. *J. Am. Chem. Soc.* **2023**, *145* (9), 5506−5511.
(50) Zou, S.; Chen, J.; Lee, Y. Y.; Wang, Y.; Zhang, Z. Candida antartica lipase-catalyzed esterification for efficient partial acylglycerol synthesis in solvent-free system: Substrate selectivity, molecular modelling and optimization. *Bioresour. Technol.* **2024**, *412*, 131368.
(51) Huang, A.; Song, X. H.; Song, L. L.; Pan, S. S.; Tian, J. X.; Wang, P. F.; Ju, S. Y.; Wang, Y. J. Reshaping the substrate-binding pocket of ene-reductase for enhanced and inverted stereoselectivity: a conciseaccess to the stereocomplementary chiral GABA derivatives. *ACS Catal.* **2025**, *15* (3), 1961−1972.
(52) Ashley, B.; Mathew, S.; Sajjad, M.; Zhu, Y.; Novikovs, N.; Baslé, A.; Marles-Wright, J.; Campopiano, D. J. Rational engineering of a thermostable α-oxoamine synthase biocatalyst expands the substrate scope and synthetic applicability. *Commun. Chem.* **2025**, *8* (1), 78.
(53) Heinemann, P. M.; Armbruster, D.; Hauer, B. Active-site loop variations adjust activity and selectivity of the cumene dioxygenase. *Nat. Commun.* **2021**, *12* (1), 1095.
(54) Denard, C. A.; Paresi, C.; Yaghi, R.; McGinnis, N.; Bennett, Z.; Yi, L.; Georgiou, G.; Iverson, B. L. YESS 2.0, a Tunable Platform for Enzyme Evolution, Yields Highly Active TEV Protease Variants. *ACS Synth. Biol.* **2021**, *10* (1), 63−71.
(55) Qu, G.; Bi, Y.; Liu, B.; Li, J.; Han, X.; Liu, W.; Jiang, Y.; Qin, Z.; Sun, Z. Unlocking the stereoselectivity and substrate acceptance of enzymes: proline-induced loop engineering test. *Angew. Chem., Int. Ed. Engl.* **2022**, *61* (1), No. e202110793.
(56) St-Jacques, A. D.; Rodriguez, J. M.; Eason, M. G.; Foster, S. M.; Khan, S. T.; Damry, A. M.; Goto, N. K.; Thompson, M. C.; Chica, R. A. Computational remodeling of an enzyme conformational landscape for altered substrate selectivity. *Nat. Commun.* **2023**, *14* (1), 6058.
(57) Kraus, M.; Grimm, C.; Seibel, J. Switching enzyme specificity from phosphate to resveratrol glucosylation. *Chem. Commun.* **2017**, *53* (90), 12181−12184.
(58) Pan, Y.; Zhu, L.; Tan, J.; Lou, D.; Wang, B. Engineering the cofactor binding site of 7α-hydroxysteroid dehydrogenase for improvement of catalytic activity, thermostability, and alteration of substrate preference. *Int. J. Biol. Macromol.* **2024**, *258*, 128847.
(59) Ge, F.; Chen, G.; Qian, M.; Xu, C.; Liu, J.; Cao, J.; Li, X.; Hu, D.; Xu, Y.; Xin, Y.; Wang, D. L.; Zhou, J.; Shi, H.; Tan, Z. B. Artificial intelligence aided lipase production and engineering for enzymatic performance improvement. *J. Agric. Food. Chem.* **2023**, *71* (41), 14911−14930.
(60) Lu, L.; Wang, X.; Wang, T.; Shen, X.; Sun, X.; Tian, P.; Yan, Y.; Nielsen, J.; Wang, J.; Yuan, Q. A bacterial platform for producing aromatic esters from glycerol. *Nat. Chem. Eng.* **2024**, *1* (12), 751−764.
(61) Luo, J.; Song, C.; Cui, W.; Han, L.; Zhou, Z. Counteraction of stability-activity trade-off of Nattokinase through flexible region shifting. *Food Chem.* **2023**, *423*, 136241.
(62) Vanella, R.; Küng, C.; Schoepfer, A. A.; Doffini, V.; Ren, J.; Nash, M. A. Understanding activity-stability tradeoffs in biocatalysts by enzyme proximity sequencing. *Nat. Commun.* **2024**, *15* (1), 1807.
(63) Xu, K.; Chen, Q.; Fu, H.; Chen, Q.; Gu, J.; Wang, X.; Zhou, J. Simultaneous Engineering of the thermostability and activity of a novel aldehyde dehydrogenase. *ACS Catal.* **2025**, *15* (3), 1841−1853.
(64) Zhao, Y.; Chen, K.; Yang, H.; Wang, Y.; Liao, X. Semirational design based on consensus sequences to balance the enzyme activity-stability trade-off. *J. Agric. Food. Chem.* **2024**, *72* (12), 6454−6462.
(65) Song, S.; Wang, C.; Bao, W.; Xu, Z.; Wu, J.; Lin, L.; Zhao, Y. High-Throughput 19F NMR chiral analysis for screening and directed evolution of imine reductases. *ACS Cent. Sci.* **2025**, *11* (7), 1094−1102.
(66) Li, C.; Gao, X.; Qi, H.; Zhang, W.; Li, L.; Wei, C.; Wei, M.; Sun, X.; Wang, S.; Wang, L. Y.; et al. Substantial improvement of an epimerase for the synthesis of D-Allulose by biosensor-based high-throughput microdroplet screening. *Angew. Chem., Int. Ed. Engl.* **2023**, *62* (10), No. e202216721.
(67) Nauwynck, W.; Faust, K.; Boon, N. Droplet microfluidics for single-cell studies: a frontier in ecological understanding of microbiomes. *FEMS Microbiol. Rev.* **2025**, *49*, fuaf032.
(68) Su, X.; Yang, J.; Yuan, H.; Liu, C.; Tu, R.; Liu, P.; Wang, Q.; Zhu, L. Directed evolution of laccase for improved thermal stability facilitated by droplet-based microfluidic screening system. *J. Agric. Food. Chem.* **2022**, *70* (42), 13700−13708.
(69) Jiang, J.; Yang, G.; Ma, F. Fluorescence coupling strategies in fluorescence-activated droplet sorting (FADS) for ultrahigh-throughput screening of enzymes, metabolites, and antibodies. *Biotechnol. Adv.* **2023**, *66*, 108173.
(70) Agresti, J. J.; Antipov, E.; Abate, A. R.; Ahn, K.; Rowat, A. C.; Baret, J.-C.; Marquez, M.; Klibanov, A. M.; Griffiths, A. D.; Weitz, D. A. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *PNAS* **2010**, *107* (9), 4004−4009.

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210

(71) Mohammadi, M.; Shahedi, M.; Ahrari, F.; Mostafavi, M.; Habibi, Z.; Yousefi, M. Isocyanide-based multi-component reactions for carrier-free and carrier-bound covalent immobilization of enzymes. *Nat. Protoc.* **2023**, *18* (5), 1641−1657.
(72) Van Stappen, C.; Deng, Y.; Liu, Y.; Heidari, H.; Wang, J. X.; Zhou, Y.; Ledray, A. P.; Lu, Y. Designing artificial metalloenzymes by tuning of the environment beyond the primary coordination sphere. *Chem. Rev.* **2022**, *122* (14), 11974−12045.
(73) Lan, Y.; Zhang, C.; Tang, C.; Ye, Y.; Zhang, R.; Sheng, X.; Liao, C. Semirational protein engineering of a decarboxylative aldolase for the regiodivergent and stereodivergent synthesis of cyclic imino acids. *Angew. Chem., Int. Ed. Engl.* **2025**, *64* (11), No. e202500080.
(74) Nguyen, T. T.; Jiang, Z.; Nguyen, V. N.; Le, N. Q. K.; Chua, M. C. H. Integrating ESM-2 and graph neural networks with AlphaFold-2 structures for enhanced protein function prediction. *ACS Omega* **2025**, *10* (33), 38103−38111.
(75) Ghazikhani, H.; Butler, G. Ion channel classification through machine learning and protein language model embeddings. *J. Integr Bioinform.* **2024**, *21* (4), 20230047.
(76) Heinzinger, M.; Weissenow, K.; Sanchez, J. G.; Henkel, A.; Steinegger, M.; Rost, B. ProstT5: bilingual language model for protein sequence and structure. *NAR:Genomics Bioinf.* **2023**, *6*, lqae150.
(77) Madani, A.; Krause, B.; Greene, E. R.; Subramanian, S.; Mohr, B. P.; Holton, J. M.; Olmos, J. L. J.; Xiong, C.; Sun, Z. Z.; Socher, R.; et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **2023**, *41* (8), 1099−1106.
(78) Cheng, J.; Novati, G.; Pan, J.; Bycroft, C.; Žemgulytė, A.; Applebaum, T.; Pritzel, A.; Wong, L. H.; Zielinski, M.; Sargeant, T.; et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **2023**, *381* (6664), No. eadg7492.
(79) Rao, R.; Liu, J.; Verkuil, R.; Meier, J.; Canny, J. F.; Abbeel, P.; Sercu, T.; Rives, A. MSA Transformer. *bioRxiv* **2021**.
(80) Zhang, Y.; Qin, Y.; Pourmirzaei, M.; Shao, Q.; Wang, D.; Xu, D. Enhancing structure-aware protein language models with efficient fine-tuning for various protein prediction tasks. *Methods Mol. Biol.* **2025**, *2941*, 31−58.
(81) Xie, R.; Wei, H.; Feng, L.; An, B. GearNet: stepwise dual learning for weakly supervised domain adaptation. *Proceedings Of The AAAI Conference On Artificial Intelligence*. AAAI, 2022; Vol., *36*; pp. 8717−8725. DOI: .
(82) Wang, Z.; Combs, S. A.; Brand, R.; Calvo, M. R.; Xu, P.; Price, G.; Golovach, N.; Salawu, E. O.; Wise, C. J.; Ponnapalli, S. P.; Clark, P. M. LM-GVP: an extensible sequence and structure informed deep learning framework for protein property prediction. *Sci. Rep.* **2022**, *12* (1), 6832.
(83) Hayes, T.; Rao, R.; Akin, H.; Sofroniew, N. J.; Oktay, D.; Lin, Z.; Verkuil, R.; Tran, V. Q.; Deaton, J.; Wiggert, M.; Badkundri, R.; Shafkat, I.; Gong, J.; Derry, A.; Molina, R. S.; Thomas, N.; Khan, Y. A.; Mishra, C.; Kim, C.; Bartie, L. J.; Nemeth, M.; Hsu, P. D.; Sercu, T.; Candido, S.; Rives, A. Simulating 500 million years of evolution with a language model. *Science* **2025**, *387* (6736), 850−858.
(84) Su, J.; Zhou, X.; Zhang, X.; Yuan, F. ProTrek: navigating the protein universe through tri-modal contrastive learning. *bioRxiv* **2024**.
(85) Zhou, G.; Zhao, Y.; He, S.; Bo, X. SST-ResNet: A sequence and structure information integration model for protein property prediction. *Int. J. Mol. Sci.* **2025**, *26* (6), 2783.
(86) Deng, M.; Wang, B.; Zhou, J.; Dong, J.; Ni, Y.; Han, R. Ancestral sequence reconstruction and semirational engineering of glycosyltransferase for efficient synthesis of rare ginsenoside Rh1. *J. Agric. Food. Chem.* **2025**, *73* (13), 7944−7953.
(87) Sun, J.; Cui, Y.; Wu, B. Chapter Ten - GRAPE, a greedy accumulated strategy for computational protein engineering. *Methods Enzymol.* **2021**, *648*, 207−230.
(88) Kunka, A.; Marques, S. M.; Havlasek, M.; Vasina, M.; Velatova, N.; Cengelova, L.; Kovar, D.; Damborsky, J.; Marek, M.; Bednar, D.; Prokop, Z. Advancing enzyme's stability and catalytic efficiency through synergy of force-field calculations, evolutionary analysis, and machine learning. *ACS Catal.* **2023**, *13* (19), 12506−12518.
(89) Wang, C.; Li, X.; Bai, Y.; Li, X.; Jin, Z.; Long, J. Multienzyme cascade coimmobilization on ZIF-8-Coated magnetic nanoparticles for efficient d-Allulose synthesis. *J. Agric. Food. Chem.* **2025**, *73* (38), 24302−24313.
(90) Niu, Y.; Chen, B.; Zhang, H.; Zheng, W.; Wu, J.; Yang, L.; Yang, M.; Yu, H. Computational design of a thermostable and highly active terminal deoxynucleotidyl transferase for synthesis of long de novo DNA molecules. *ACS Catal.* **2025**, *15* (9), 7201−7216.
(91) Xu, S. Y.; Chu, R. L.; Liu, H. T.; Weng, C. Y.; Wang, Y. J.; Zheng, Y. G. Computer-directed rational design enhanced the thermostability of carbonyl reductase LsCR for the synthesis of ticagrelor precursor. *Biotechnol. Bioeng.* **2024**, *121* (5), 1532−1542.
(92) Manyumwa, C. V.; Zhang, C.; Jers, C.; Mijakovic, I. Alpha carbonic anhydrase from nitratiruptor tergarcus engineered for increased activity and thermostability. *Int. J. Mol. Sci.* **2024**, *25* (11), 5853.
(93) Cao, H.; Wang, J.; He, L.; Qi, Y.; Zhang, J. Z. DeepDDG: Predicting the stability change of protein point mutations using neural networks. *J. Chem. Inf. Model.* **2019**, *59* (4), 1508−1514.
(94) Zhang, P.; Mao, X. A.; Kong, X. L.; Gong, J. S.; Wang, Z. K.; Su, C.; Li, H.; Zhang, X. M.; Rao, Z. M.; Xu, Z. H.; Shi, J. S. Rational design of phospholipase D to enhance thermostability and catalytic activity for efficient biocatalytic phosphatidylserine synthesis. *Food Chem.* **2025**, *474*, 143212.
(95) Charoenwongpaiboon, T.; Wangpaiboon, K.; Puangpathanachai, M.; Pongsawasdi, P.; Pichyangkura, R. Energy- and evolution-based design of inulosucrase for enhanced thermostability and inulin production. *Appl. Microbiol. Biotechnol.* **2023**, *107* (22), 6831−6843.
(96) Musil, M.; Stourac, J.; Bendl, J.; Brezovsky, J.; Prokop, Z.; Zendulka, J.; Martinek, T.; Bednar, D.; Damborsky, J. FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Res.* **2017**, *45* (W1), 393−399.
(97) Wu, T.; Du, Z.; Li, H.; Jiang, Z.; Zheng, M.; Li, Z.; Hong, T.; Du, X.; Ni, H.; Zhu, Y. A disulfide bond mutant of Pseudoalteromonas porphyrae κ-carrageenase conferred improved thermostability and catalytic activity and facilitated its utilization in κ-carrageenan industrial waste residues recycling. *Int. J. Biol. Macromol.* **2024**, *280* (Pt 1), 135573.
(98) Zhao, G.; Zhang, J. Molecular dynamics studies of disulfide bonds for enhancing the stability of serine protease PB92. *New J. Chem.* **2024**, *48* (4), 1614−1622.
(99) Yi, X. M.; Yu, H. R.; Yu, H. W.; Ye, L. D. Enhancing activity and stability of transaminase through integrated machine learning, rational design, and directed evolution approaches. *ACS Catal.* **2025**, *15* (16), 14150−14160.
(100) Suplatov, D.; Timonina, D.; Sharapova, Y.; Švedas, V. Yosshi: a web-server for disulfide engineering by bioinformatic analysis of diverse protein families. *Nucleic Acids Res.* **2019**, *47*, W308−W314.
(101) Huang, A.; Chai, C.; Zhang, J.; Zhao, L.; Lu, F.; Liu, F. Engineered N57P variant of ulvan lyase with improvement of catalytic efficiency and thermostability via reducing loop flexibility and anchoring substrate. *ACS Sustainable Chem. Eng.* **2021**, *9* (48), 16415−16423.
(102) Chi, H.; Wang, Y.; Xia, B.; Zhou, Y.; Lu, Z.; Lu, F.; Zhu, P. Enhanced Thermostability and Molecular Insights for l-Asparaginase from *Bacillus licheniformis* via Structure- and Computation-Based Rational Design. *J. Agric. Food. Chem.* **2022**, *70* (45), 14499−14509.
(103) Liu, Z.; Liang, Q.; Wang, P.; Kong, Q.; Fu, X.; Mou, H. Improving the kinetic stability of a hyperthermostable β-mannanase by a rationally combined strategy. *Int. J. Biol. Macromol.* **2021**, *167*, 405−414.
(104) Luo, X.; Wang, Y.; Zheng, W.; Sun, X.; Hu, G.; Yin, L.; Zhang, Y.; Yin, F.; Fu, Y. Simultaneous improvement of the thermostability and activity of lactic dehydrogenase from Lactobacillus rossiae through rational design. *RSC Adv.* **2022**, *12* (51), 33251−33259.
(105) Sun, J.; Shi, W.; Xing, Z.; Fan, G.; Sun, Q.; Wu, L.; Ma, J.; Cui, Y.; Wu, B. GRAPE-WEB: An automated computational redesign web

server for improving protein thermostability. *Mlife* **2024**, *3* (4), 527−531.
(106) The UniProt Consortium UniProt: the universal protein knowledgebase in 2025*Nucleic Acids Res.***2025***53*609−617
(107) Zhang, Y.; Zhao, Y. F.; Zhu, X. L.; Peng, Y. Z.; Khurshid, M.; Herman, R. A.; Lv, X.; Li, J.; Ling, M. X.; Wang, J.; You, S. Bioengineering GH10 xylanase through rational design and MD simulation: Enhanced thermal stability and bioenergy synergies with cellulase. *Ind. Crops Prod.* **2025**, *234*, 121606.
(108) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: an online force field. *Nucleic Acids Res.* **2005**, *33* (Web Server), W382−W388.
(109) Teng, R.; Zhang, J.; Tu, Z.; He, Q.; Li, Y. Computer-Aided design to improve the thermal stability of rhizomucor miehei Lipase. *Foods* **2024**, *13* (24), 4023.
(110) Gao, S.; Zhu, S.; Huang, R.; Li, H.; Wang, H.; Zheng, G. Engineering the enantioselectivity and thermostability of a (+)-γ-Lactamase-Lactamase from microbacterium hydrocarbonoxydans for kinetic resolution of vince lactam (2-Azabicyclo[2.2.1]hept-5-en-3-one). *Appl. Environ. Microbiol.* **2018**, *84* (1), No. e01780−01717.
(111) Li, S.; Cao, L.; Yang, X.; Wu, X.; Xu, S.; Liu, Y. Simultaneously optimizing multiple properties of β-glucosidase Bgl6 using combined (semi-)rational design strategies and investigation of the underlying mechanisms. *Bioresour. Technol.* **2023**, *374*, 128792.
(112) Aalbers, F. S.; Fürst, M. J.; Rovida, S.; Trajkovic, M.; Gómez Castellanos, J. R.; Bartsch, S.; Vogel, A.; Mattevi, A.; Fraaije, M. W. Approaching boiling point stability of an alcohol dehydrogenase through computationally-guided enzyme engineering. *Elife* **2020**, *9*, No. e54639.
(113) Meng, Q.; Capra, N.; Palacio, C. M.; Lanfranchi, E.; Otzen, M.; van Schie, L. Z.; Rozeboom, H. J.; Thunnissen, A. M. W. H.; Wijma, H. J.; Janssen, D. B. Robust ω-transaminases by computational stabilization of the subunit interface. *ACS Catal.* **2020**, *10* (5), 2915−2928.
(114) Guo, Y.; Alvigini, L.; Trajkovic, M.; Alonso-Cotchico, L.; Monza, E.; Savino, S.; Marić, I.; Mattevi, A.; Fraaije, M. W. Structure- and computational-aided engineering of an oxidase to produce isoeugenol from a lignin-derived compound. *Nat. Commun.* **2022**, *13* (1), 7195.
(115) Sumida, K. H.; Núñez-Franco, R.; Kalvet, I.; Pellock, S. J.; Wicky, B. I. M.; Milles, L. F.; Dauparas, J.; Wang, J.; Kipnis, Y.; Jameson, N.; et al. Improving protein expression, stability, and function with ProteinMPNN. *J. Am. Chem. Soc.* **2024**, *146* (3), 2054−2061.
(116) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **2022**, *378* (6615), 49−56.
(117) Go, S.-R.; Lee, S. J.; Ahn, W. C.; Park, K. H.; Woo, E. J. Enhancing the thermostability and activity of glycosyltransferase UGT76G1 via computational design. *Commun. Chem.* **2023**, *6* (1), 265.
(118) Thieker, D. F.; Maguire, J. B.; Kudlacek, S. T.; Leaver-Fay, A.; Lyskov, S.; Kuhlman, B. Stabilizing proteins, simplified: A Rosetta-based webtool for predicting favorable mutations. *Protein Sci.* **2022**, *31* (10), No. e4428.
(119) Korbeld, K. T.; Fürst, M. Enriching stabilizing mutations through automated analysis of molecular dynamics simulations using BoostMut. *Protein Sci.* **2025**, *34* (11), No. e70334.
(120) Cai, X.; Jiang, H.; Zhang, T.; Jiang, B.; Mu, W.; Miao, M. Thermostability and specific-activity enhancement of an arginine deiminase from *Enterococcus faecalis* SK23.001 via semirational design for l-Citrulline production. *J. Agric. Food. Chem.* **2018**, *66* (33), 8841−8850.
(121) Kunlawatwimon, T.; Promdonkoy, B.; Boonserm, P. Rational protein design to improve the thermal stability of a *Bacillus thuringiensis* Vip3A protein. *J. Invertebr. Pathol.* **2025**, *213*, 108429.
(122) Pucci, F.; Kwasigroch, J. M.; Rooman, M. Protein thermal stability engineering using HoTMuSiC. In *Structural Bioinformatics: methods and Protocols*; Gáspári, Z.; ed., Springer: US. pp. 59−73, 2020.
(123) Tang, H.; Zhu, H. L.; Zhao, J. Q.; Wang, L. Y.; Xue, Y. P.; Zheng, Y. G. Through virtual saturation mutagenesis and rational design for superior substrate conversion in engineered d-amino acid oxidase. *Biotechnol. J.* **2024**, *19* (7), No. e2400287.
(124) Pavelka, A.; Sebestova, E.; Kozlikova, B.; Brezovsky, J.; Sochor, J.; Damborsky, J. CAVER: algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2016**, *13* (3), 505−517.
(125) Wang, X.; Wang, Z.; Zhang, X.; Zhang, Y.; Zhang, W.; Zhang, Y.; Zhang, X.; Xiao, Y.; Zhang, Y.; Fang, W. Bioinformatics-assisted mining and design of novel pullulanase suitable for starch cold hydrolysis. *J. Biotechnol.* **2025**, *398*, 106−116.
(126) Münch, J.; Soler, J.; Gildor-Cristal, O.; Fleishman, S. J.; Garcia-Borràs, M.; Weissenborn, M. J. Computationally designed peroxygenases that exhibit diverse and selective terpene oxy-functionalization. *ACS Catal.* **2025**, *15* (15), 12741−12755.
(127) Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D. L.; Prilusky, J.; Amengual-Rigo, P.; Guallar, V.; Tawfik, D. S.; Fleishman, S. J. Automated design of efficient and functionally diverse Eenzyme repertoires. *Mol. Cell* **2018**, *72* (1), 178−186.
(128) Qiu, W.; Yang, P.; Ye, J.; Zhou, J.; Liu, S. Unveiling highly active and stable l-glutaminase through ancestral sequence reconstruction and turnover number prediction. *J. Agric. Food. Chem.* **2025**, *73* (9), 5353−5362.
(129) Li, F.; Yuan, L.; Lu, H.; Li, G.; Chen, Y.; Engqvist, M. K. M.; Kerkhoven, E. J.; Nielsen, J. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* **2022**, *5* (8), 662−672.
(130) Cui, H.; Su, Y.; Dean, T. J.; Yu, T.; Zhang, Z.; Peng, J.; Shukla, D.; Zhao, H. Enzyme specificity prediction using cross-attention graph neural networks. *Nature* **2025**, *647* (8090), 639−647.
(131) Kroll, A.; Ranjan, S.; Engqvist, M. K. M.; Lercher, M. J. A general model to predict small molecule substrates of enzymes based on machine and deep learning. *Nat. Commun.* **2023**, *14* (1), 2787.
(132) Fortela, D. L. B; Mikolajczyk, A. P.; Carnes, M. R.; Sharp, W.; Revellame, E.; Hernandez, R.; Holmes, W. E.; Zappi, M. E. Predicting molecular docking of per- and polyfluoroalkyl substances to blood protein using generative artificial intelligence algorithm DiffDock. *Biotechniques* **2024**, *76* (1), 14−26.
(133) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. DiffDock: diffusion steps, twists, and turns for molecular Docking. *ArXiv* **2022**.
(134) Fortela, D. L. B.; Mikolajczyk, A. P.; Carnes, M. R.; Sharp, W.; Revellame, E.; Hernandez, R.; Holmes, W. E.; Zappi, M. E. Predicting molecular docking of per- and polyfluoroalkyl substances to blood protein using generative artificial intelligence algorithm Diffdock. *BioTechniques* **2024**, *76* (1), 14−26.
(135) Wang, Z.; Xie, D.; Wu, D.; Luo, X.; Wang, S.; Li, Y.; Yang, Y.; Li, W.; Zheng, L. Robust enzyme discovery and engineering with deep learning using CataPro. *Nat. Commun.* **2025**, *16* (1), 2736.
(136) Boorla, V. S.; Maranas, C. D. CatPred: a comprehensive framework for deep learning *in vitro* enzyme kinetic parameters. *Nat. Commun.* **2025**, *16* (1), 2072.
(137) Yu, H.; Deng, H.; He, J.; Keasling, J. D.; Luo, X. UniKP: a unified framework for the prediction of enzyme kinetic parameters. *Nat. Commun.* **2023**, *14* (1), 8211.
(138) Kroll, A.; Rousset, Y.; Spitzlei, T.; Lercher, M. J. DeepMolecules: a web server for predicting enzyme and transporter−small molecule interactions. *Nucleic Acids Res.* **2025**, *53* (W1), W213−W218.
(139) Wang, J.; Yang, Z.; Chen, C.; Yao, G.; Wan, X.; Bao, S.; Ding, J.; Wang, L.; Jiang, H. MPEK: a multitask deep learning framework based on pretrained language models for enzymatic reaction kinetic parameters prediction. *Briefings Bioinf.* **2024**, *25* (5), bbae387.

(140) Shao, Q.; Jiang, Y.; Yang, Z. J. EnzyHTP: A high-throughput computational platform for enzyme modeling. *J. Chem. Inf. Model.* **2022**, *62* (3), 647−655.
(141) Wang, J.; Ouyang, X.; Meng, S.; Zhao, B.; Liu, L.; Li, C.; Li, H.; Zheng, H.; Liu, Y.; Shi, T.; Zhao, Y. L.; Ni, J. Rational multienzyme architecture design with iMARS. *Cell* **2025**, *188* (5), 1349−1362.
(142) Zhang, L.; Zhou, X.; Liang, L.; Lin, J. PPAC: Predicting Protein−Protein affinity changes induced by amino acid mutations using protein large language models. *ACS Omega* **2025**, *10* (45), 54895−54902.
(143) Guan, B.; Guo, L.; Lam, T. K.-Y.; Xiong, Z.; Luo, S.; Sun, Y.; Chen, K. ProtPSP: Determination of protein phosphorylation sites with the protein large language model. *ACS Omega* **2025**.
(144) Ong, G.; Kong, K. W.; Poh, S. E.; Wong, F. T.; Seow, Y.; Koh, W. Design of a labile RNase a using protein language models. *ACS Synth. Biol.* **2025**.
(145) Xu, S.; Onoda, A. Accurate and fast prediction of intrinsically disordered protein by multiple protein language models and ensemble learning. *J. Chem. Inf. Model.* **2024**, *64* (7), 2901−2911.
(146) Ekrami, M.; Shokrollahi Yancheshmeh, B.; Roshani-Dehlaghi, N.; Mobahi, N.; Emam-Djomeh, Z.; Mohammadifar, M. Next-generation smart and safe foods: Artificial intelligence -driven strategies for 4D food pre-printing challenges. *Trends In Food Sci. Technol.* **2025**, *165*, 105317.
(147) Joy, A.; Biswas, R. Significance of the disulfide bridge in the structure and stability of metalloprotein azurin. *J. Phys. Chem. B* **2024**, *128* (4), 973−984.
(148) Liu, J.; Han, L.; Li, J.; Du, G.; Zhang, G. Modification of flexible regions for enhanced thermal stability of alkaline amylase. *J. Agric. Food. Chem.* **2025**, *73* (16), 9973−9982.
(149) Chen, J.; Chen, D.; Chen, Q.; Xu, W.; Zhang, W.; Mu, W. Computer-Aided targeted mutagenesis of thermoclostridium caenicola d-Allulose 3-Epimerase for improved thermostability. *J. Agric. Food. Chem.* **2022**, *70* (6), 1943−1951.
(150) Heiby, J. C.; Goretzki, B.; Johnson, C. M.; Hellmich, U. A.; Neuweiler, H. Methionine in a protein hydrophobic core drives tight interactions required for assembly of spider silk. *Nat. Commun.* **2019**, *10* (1), 4378.
(151) Yao, L. L.; Xue, B.; Ye, Y. F.; Wang, Z. X.; Li, Y. Y.; Zheng, B. F. C.; Ju, S. Y.; Wang, Y. J. Structure-Guided engineering of carbonyl reductase LbCR to simultaneously enhance catalytic activity and thermostability toward bulky ketones. *J. Agric. Food. Chem.* **2025**, *73* (17), 10470−10481.
(152) Huang, Z.; Zhou, J.; Wang, J.; Xu, S.; Cheng, C.; Ma, J.; Gao, Z. Complementary distant and active site mutations simultaneously enhance catalytic activity and thermostability of α-Galactosidase. *J. Agric. Food. Chem.* **2025**, *73* (6), 3635−3644.
(153) Wang, F.; Zhu, M.; Song, Z.; Li, C.; Wang, Y.; Zhu, Z.; Sun, D.; Lu, F.; Qin, H. M. Reshaping the binding pocket of lysine hydroxylase for enhanced activity. *ACS Catal.* **2020**, *10* (23), 13946−13956.
(154) Jacquet, P.; Billot, R.; Shimon, A.; Hoekstra, N.; Bergonzi, C.; Jenks, A.; Chabrière, E.; Daudé, D.; Elias, M. H. Changes in active site loop conformation relate to the transition toward a novel enzymatic activity. *JACS Au* **2024**, *4* (5), 1941−1953.
(155) Visootsat, A.; Nakamura, A.; Wang, T. W.; Iino, R. Combined approach to engineer a highly active mutant of processive Chitinase hydrolyzing crystalline chitin. *ACS Omega* **2020**, *5* (41), 26807−26816.
(156) Zou, Z.; Mate, D. M.; Rübsam, K.; Jakob, F.; Schwaneberg, U. Sortase-Mediated High-Throughput screening platform for directed enzyme evolution. *ACS Comb. Sci.* **2018**, *20* (4), 203−211.
(157) Chanquia, S. N.; Bittner, J. P.; Santner, P.; Szabó, L. K.; Madsen, J. S.; Øhlenschlæger, M. L.; Sarvari, A. G.; Merrild, A. H. J.; Fønss, K. G.; Jaron, D.; et al. Active-Site mutagenesis of fatty acid photodecarboxylase: experimental and computational insight into substrate chain-length specificity. *ACS Catal.* **2024**, *14* (21), 15837−15849.
(158) Wu, L.; Qin, L.; Nie, Y.; Xu, Y.; Zhao, Y. L. Computer-aided understanding and engineering of enzymatic selectivity. *Biotechnol. Adv.* **2022**, *54*, 107793.
(159) Li, H.; Zhang, W.; Han, Y.; Tang, G.; Lu, F.; Qin, H. M. Programming a bacterial biosensor for directed evolution of tryptophan hydroxylase via high-throughput droplet sorting. *Biosens. Bioelectron.* **2025**, *271*, 117072.
(160) Yang, B.; Wang, H.; Song, W.; Chen, X.; Liu, J.; Luo, Q.; Liu, L. Engineering of the conformational dynamics of lipase to increase enantioselectivity. *ACS Catal.* **2017**, *7* (11), 7593−7599.
(161) Koryakina, I.; Kasey, C.; McArthur, J. B.; Lowell, A. N.; Chemler, J. A.; Li, S.; Hansen, D. A.; Sherman, D. H.; Williams, G. J. Inversion of extender unit selectivity in the erythromycin polyketide synthase by acyltransferase domain engineering. *ACS Chem. Biol.* **2017**, *12* (1), 114−123.
(162) Lu, J.; Lv, X.; Yu, W.; Zhang, J.; Lu, J.; Liu, Y.; Li, J.; Du, G.; Chen, J.; Liu, L. Reshaping phosphatase substrate preference for controlled biosynthesis using a "Design-Build-Test-Learn" framework. *Adv. Sci.* **2024**, *11* (22), No. e2309852.
(163) Bian, S. Q.; Wang, Z.; Gong, J. S.; Su, C.; Li, H.; Xu, Z. H.; Shi, J. S. Enhancing the substrate specificity of nitrilase toward aliphatic nitriles based on substrate channel design. *ACS Sustainable Chem. Eng.* **2025**, *13* (14), 5332−5344.
(164) Oohora, K.; Onoda, A.; Hayashi, T. Hemoproteins reconstituted with artificial metal complexes as biohybrid catalysts. *Acc. Chem. Res.* **2019**, *52* (4), 945−954.
(165) Lu, M.; Li, Z.; Zhuang, H.; Yang, S.; Zhao, X.; Feng, R.; Shen, H.; Kovalevsky, A.; Zhang, S.; Xie, Z. P.; Li, X.; Shen, Q. R.; Wan, Q. Dual enhancement of thermostability and activity of Xylanase through computer-aided rational design. *ACS Sustainable Chem. Eng.* **2024**, *12* (41), 15114−15124.
(166) Li, M.; Zhuang, W.; Zhang, J.; Zhang, K.; Xu, J.; Wang, Z. Improving the thermostability and catalytic activity of RgDAAO by a combinatorial strategy using sequence consensus design and SpyTag/SpyCatcher self-cyclization. *J. Agric. Food. Chem.* **2025**, *73* (38), 24211−24221.
(167) Somvilla, I.; Meinert, H.; Cziegler, C.; Gökler, T.; Berner, C. F.; Wolfgramm, H.; Branson, Y.; Conibear, A. C.; Völker, U.; Badenhorst, C. P. S.; Bayer, T.; Bornscheuer, U. T. Ultrahigh-Throughput activity engineering of promiscuous amidases through a fluorescence-activated cell sorting assay. *ACS Catal.* **2025**, *15* (11), 8902−8912.

https://doi.org/10.1021/acs.jafc.5c13434
J. Agric. Food Chem. 2026, 74, 185−210