

Ancestral Sequence Reconstruction for Designing Biocatalysts and Investigating their Functional Mechanisms

Published as part of JACS Au special issue "Biocatalysis in Asia and Pacific".

Kridsakorn Prakinee, Suppalak Phaisan, Sirius Kongjaroon, and Pimchai Chaiyen*

Cite This: JACS Au 2024, 4, 4571–4591

Read Online

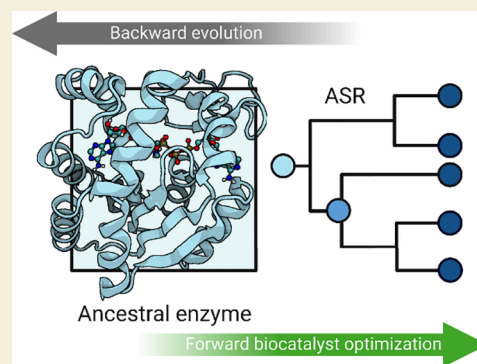
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Biocatalysis has emerged as a green approach for efficient and sustainable production in various industries. In recent decades, numerous advancements in computational and predictive approaches, including ancestral sequence reconstruction (ASR) have sparked a new wave for protein engineers to improve and expand biocatalyst capabilities. ASR is an evolution-based strategy that uses phylogenetic relationships among homologous extant sequences to probabilistically infer the most likely ancestral sequences. It has proven to be a powerful tool with applications ranging from creating highly stable enzymes for direct applications to preparing moderately active robust protein scaffolds for further enzyme engineering. Intriguingly, it can also provide insights into fundamental aspects that are challenging to study with extant (current) enzymes. This Perspective discusses a practical strategy for guiding enzyme engineers on how to embrace ASR as a practical or associated protocol for protein engineering and highlights recent examples of using ASR in various applications, including increasing thermostability, expanding promiscuity, fine-tuning selectivity and function, and investigating mechanistic and evolution aspects. We believe that the use of the ASR approach will continue to contribute to the ongoing development of the biocatalysis field. We have been in a "golden era" for biocatalysis in which numerous useful enzymes have been developed through many waves of enzyme engineering via advancements in computational methodologies.

KEYWORDS: ancestral sequence reconstruction, ancestral enzyme, biocatalysis, enzyme engineering, enzyme design



INTRODUCTION

Enzymes or biocatalysts are powerful agents that deliver exquisite catalytic functions and selectivity. They are excellent tools for driving green industries.¹ Biocatalysts have been used in various industries, ranging from the production of commodity chemicals to the synthesis of complex pharmaceuticals. Employing enzymes in production methods contributes to the development of environmentally friendly and sustainable processes.^{2,3} Generally, biocatalysts are obtained from nature through the identification of their unique roles in selective transformation in natural reactions and pathways. Unfortunately, not many wild-type enzymes can meet industrial requirements and be used directly in synthesis applications due to their limited stability and low production yield. Owing to scientific and technological advancements during the past two decades, native enzymes can be engineered to achieve properties and functions desired for *in vivo* and *in vitro* applications.⁴ The approach of protein engineering and design has accelerated the pace of biocatalyst improvement. Directed evolution is a rapid and efficient method to improve biocatalysts; however, it requires screening of large libraries, which is challenging for most reactions. Rational and

semirational designs for enzyme engineering have also been demonstrated to be useful, requiring less screening effort and more specific alteration; however, these approaches require comprehensive knowledge regarding the enzyme's structures and functions to help guide the strategy.^{5–8} Effective methods to engineer enzymes of interest to meet the demands of applications with small libraries remain a challenge within the field.

Ancestral sequence reconstruction (ASR) has emerged among other techniques in protein engineering and design as an efficient approach for obtaining enzymes with robust stability and broad activities.⁹ ASR is an evolution-based strategy that uses phylogenetic relationships among homologous proteins to infer ancestral sequences based on a

Received: July 19, 2024
Revised: October 9, 2024
Accepted: October 9, 2024
Published: October 25, 2024



probabilistic approach.¹⁰ While directed evolution relies on the construction of large libraries and screening of more than 10^3 to 10^6 samples, ASR screening efforts could be performed with a very small set of enzymes of less than ten candidates. Enzyme candidates obtained from this approach are generally perceived as stable and can subsequently serve as ideal scaffolds for further studies. While the principle of ASR does not directly aim to achieve soluble protein expression, we nevertheless found that most cases discussed in this Perspective reported that ancestral sequences could express well in heterologous systems such as *Escherichia coli*. In general, ASR contributes to the improvement of protein properties with regard to increased stability¹¹ and crystallizability.¹² These findings thus highlighted the ability of the ASR approach to deliver good results using low experimental screening effort.

Because ancestral enzymes are generally assumed to be more thermostable and capable of catalyzing broader reactions than their extant descendants,^{13,14} initial work in ASR development was dedicated to testing these properties. Currently, ASR is widely used to investigate the evolutionary functions of enzymes toward a variety of substrates and for investigating catalytic functions. A comprehensive understanding of evolution-related enzymes is useful for deciphering the missing gap of enzyme catalysis through natural evolution. Upon realizing the robustness and broad activities of ancestral enzymes as mentioned earlier, many research groups have used ASR to construct ancestral enzymes as starting templates for further optimization through rational engineering or directed evolution campaigns.¹⁵ Inspired by the substantial progress of ASR usage in biocatalysis, we put together this Perspective to discuss the latest developments, particularly during the past five years, in areas ranging from resurrecting highly stable enzymes for industrial applications, creating robust protein scaffolds for further fine-tuning of specificity and functions, and finally the use of ASR for mechanistic studies¹² to understanding catalysis through evolution (Figure 1).^{16,17} As several excellent publications regarding the ASR methodology are

already available, including those focusing on principles of methods and algorithms^{18,19} and bioinformatic tools,²⁰ our Perspective thus only briefly presents workflows and currently available resources for resurrecting ancestral sequences through practical aspects of protein engineering prior to discussion of insights and key findings from study cases. A major goal of this review is to provide a summary of the ASR approach across diverse biocatalysis contexts for biochemists and chemists who are interested in employing ASR as a tool to optimize or characterize their biocatalysts or to acquire novel and robust enzymes.

WORKFLOW AND ASSOCIATED COMPUTATIONAL TOOLS

The ASR process is based on statistical principles to predict sequences of putative ancestors based on amino acid sequences of modern or known proteins. In general, the strategy begins with statistical inference of ancestral protein sequences, followed by gene synthesis, expression and characterization of the physical and functional properties of ancestral proteins.¹⁷ Generation of ancestral sequences thus consists of multiple steps in which individual steps can influence the characteristics of resultant enzymes. Therefore, the choice of proper methodology and appropriate bioinformatic resources is an essential primer for initiating the ASR campaign. In this section, we briefly summarize a necessary workflow for creating ancestral sequences starting from collecting the input sequences (Figure 2), as well as present software packages that can be used at each step (Table 1).

Typically, the generation of ancestral sequences employs four main steps including collection of extant enzyme sequences, sequence alignment and curation, construction of phylogenetic trees, and inferencing ancestral sequences from the trees.^{19,21} The first step is crucial, as it can strongly influence the quality of sequence inference. Ideally, selecting as large a number of extant sequences as possible would strengthen the quality of the tree construction and increase the sequence reconstruction accuracy. The number of extant sequences can be varied from ten to one hundred sequences. Important factors that can influence the precision and efficiency of the reconstruction process are the quantity and distribution of employed sequences (Figure 2A). Selection criteria of extant sequences depend on the project objectives; representative sequences should belong to members of the desired phenotypes or phylogenetic lineages. For instance, researchers may choose extant sequences that belong to thermophilic species if the project aim is to improve protein stability or include extant sequences with known functions and activities for functional engineering purposes. Commonly, homologous sequences can be retrieved from several databases such as GenBank of the NCBI or UniProtKB of the EBI using different search tools (Table 1). Among the available tools, BLAST is the most common tool that is used in sequence collection.

The second step is to organize the sequence pool to remove erroneous or misaligned extant sequences. Typically, a multiple sequence alignment (MSA) software such as Clustal Omega is used to select a set of appropriate sequences (Figure 2B). It is important for the MSA process to remove sequences likely containing errors while maintaining the broadest possible coverage of the extant sequence space. The undesired sequences, such as very short sequence fragments, transcription errors, wrongly assigned insertions or deletions, etc., can lead

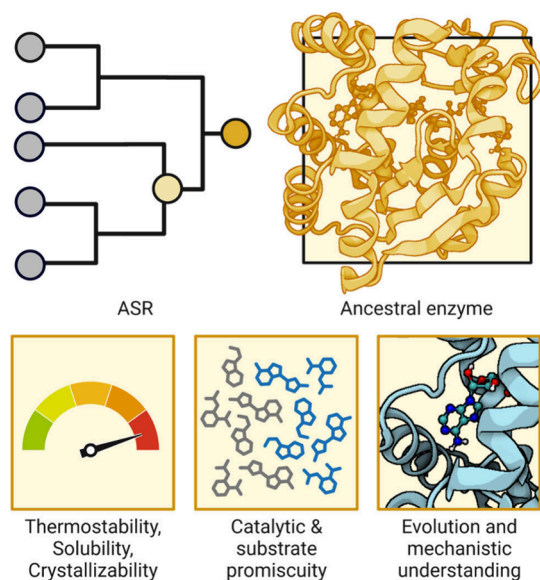


Figure 1. Applications of ancestral sequence reconstruction (ASR) range from improving biophysical properties and catalytic function to serving as an exploratory tool for understanding enzyme evolution and function.

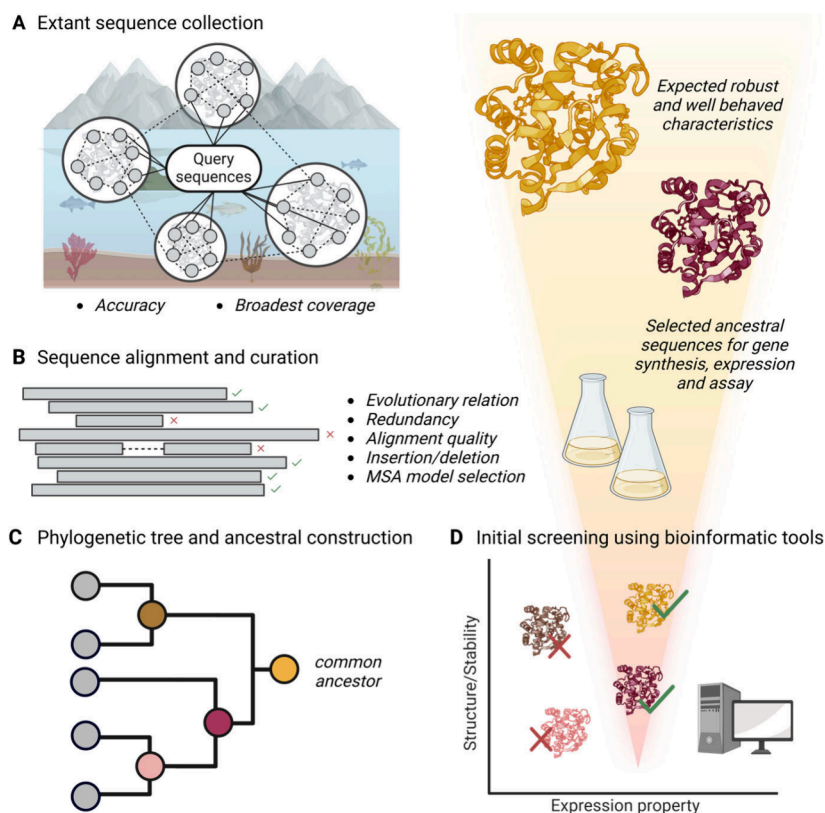


Figure 2. Workflow used for ancestral sequence reconstruction and evaluation. (A) Searching extant sequences from the database and selecting for good variations of ancestral sequences. (B) Curating sequences and performing multiple sequence alignment. (C) Constructing phylogenetic trees and analyzing for ancestral sequence inference. (D) *In silico* screening of ancestral enzymes.

to unfolding or nonfunctional resultant proteins.¹⁹ In recent years, several algorithms have been developed to attain good quality sequence alignment (Table 1), and the choice of software may depend on the user's preferences.

The third step is to organize the evolutionary relationships between extant sequences to obtain a phylogenetic tree construction. Usually, there are two methods for tree construction, including the distance-based and character-based methods. The distant-matrix method calculates the tree from a mismatched position of MSA. The closely related sequences are included in the same interior node, and the branch length represents the genetic distance between sequences. The most popular distant matrix method is the neighbor-joining method. For the character-based method, three approaches are available, including maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). All approaches use statistical algorithms to construct major frameworks for the ASR process. Currently, the majority of ASR construction uses the probabilistic approach with either ML and/or BI algorithms to infer ancestral sequences from phylogeny (Table 1).

After the MSA and phylogenetic tree have been refined, the two data sets are used for the ancestral sequence inference process (Figure 2C); three methods are commonly employed for this step, including MP, ML, and BI.²² Each method provides a different result, and there is no definitive correct prediction of inference sequences. The method selection should depend on the nature of the data set, the ultimate goal of work, and computing speed. The MP method minimizes a number of evolutionary changes and ignores the mutation rate and selection pressure of evolution by assuming that

evolutionary rates are constant in every branch of a phylogenetic tree.^{23,24} Therefore, the MP method may not represent true evolution if the evolutionary process is complex. The ML method considers a variety of evolutionary factors including selection pressures, mutation rate, and indel occurrences.^{25,26} In a data set in which the evolutionary process is complex, ancestral sequences inferred by ML would be more accurate than the MP method.²⁷ Moreover, the ML method also applies substitution models in the calculations, such as Dayhoff, Whelan and Goldman (WAG), Le and Gascuel (LG), and Jones–Taylor–Thornton (JTT). For the reason mentioned above, ML-based methods are frequently employed in both phylogenetic and ancestral sequence reconstruction steps. With more factors considered in calculations, the BI method is usually considered as more comprehensive than the ML method because it integrates previous knowledge and uncertainty into the analysis.²⁸ Two steps of estimation are generally used for the BI method. The first one is an empirical BI that performs calculations of probability distributions but does not account for uncertainty in the calculation. The latter step is a hierarchical BI that incorporates the uncertainty into the calculation and averages the possibility of the tree and evolution model before generating the final results. However, the BI method is intensive and requires significant computing power for analysis. Available software packages currently available for inferring ancestral sequences are presented in Table 1.

As an alternative to the mentioned protocols, the whole workflow can be integrated into automatic tools. In 2016, a web portal, PhyloBot, was developed as an easy-to-use sequence interface software.²⁹ The server requires the input

Table 1. Available Tools for the ASR Process

tool	description	access link	ref
BLAST	Sequence similarity search based on many databases such as nonredundant protein sequences (nr), UniProtKB/Swiss-Prot (swissprot), patented protein sequences (pataa), and Protein Data Bank (PDB). It includes various search algorithms including BLASTp for protein–protein sequence similarity searches and PSI-BLAST for iterative searching.	https://blast.ncbi.nlm.nih.gov/Blast.cgi	38
UniProt	Sequence similarity search from the UniProt database. Sequences in UniProt are classified as reviewed and nonreviewed.	https://www.uniprot.org/blast	39
EnzymeMiner	Sequence homology search of query sequence(s) in the NCBI nr database based on a PSI-BLAST algorithm. Essential residues can be used to filter the homology search. The sequences are reported with solubility expression scores.	https://loschmidt.chemi.muni.cz/enzymeminer/	40
ESI-EST	The input sequence is used as the query for a homology search of the UniProt, UniRef90, or UniRef50 database using BLAST. A sequence similarity network (SSN) can be visualized of relationships among protein sequences.	https://efi.igb.illinois.edu/efi-est/	41
KEGG	The Japanese database resource with supporting sequence similarity search (BLAST/FASTA). The tool offers useful information such as enzyme EC numbers and origin species of enzymes.	https://www.genome.jp/kegg/	42
tool	multiple sequence alignment tool	access link	ref
Clustal Omega	MSA is constructed using progressive alignment algorithms.	https://www.ebi.ac.uk/jdispatcher/msa/clustalo	43
MAFFT	Multiple Alignment using Fast Fourier Transform. MAFFT provides high accuracy using various methods including progressive and iterative based methods.	https://www.ebi.ac.uk/jdispatcher/msa/mafft	44
MUSCLE	Multiple Sequence Comparison by Log-Expectation. MUSCLE combines information from sequence and structural information.	https://www.ebi.ac.uk/jdispatcher/msa/muscle	45
T-Coffee	Tree-based Consistency Objective Function for alignment Evaluation. T-Coffee provides consistency-based multiple sequence alignment and combines output from other algorithms.	https://www.ebi.ac.uk/jdispatcher/msa/tcofee	46
tool	phylogenetic tree construction tool	access link	ref
FastTree	Rapid construction of large phylogenetic trees based on ML methods. The method is appropriate for sizable data sets comprising thousands of sequences.	http://www.microbesonline.org/fastree/	47
RAxML	Randomized Axelerated Maximum Likelihood. Fast and efficient tool based on a ML method.	https://github.com/stamatak/standard-RAxML	48
PhyML	Phylogenetic construction using Maximum Likelihood. A user-friendly tool of ML-based method with options for branch support estimation, providing options for bootstrapping and approximate likelihood ratio tests.	http://www.atgc-montpellier.fr/phyml/	49
MrBayes	Bayesian Inference of Phylogeny. A BI-based method using Markov chain Monte Carlo algorithm, allowing estimation of posterior probabilities for tree nodes.	https://nbisweden.github.io/MrBayes/download.html	50
BEAST 2	Bayesian Evolutionary Analysis Sampling Trees. A BI-based method that integrates molecular clock models and coalescent theory. The program supports complex models of sequence evolution, demographic history, molecular clock calibration, and estimation of divergence times and population parameters.	https://www.beast2.org/	51
tool	ASR inferring tool	access link	ref.
GRASP	Graphical Representation of Ancestral Sequence Predictions. An inferring ancestral sequence based on ML method. The method supports insertion or deletion of amino acids. An alignment file and phylogenetic tree file are needed as inputs. The evolution model can be selected by user.	http://grasp.scmh.uq.edu.au/	52
Ancescon	A software based on distant-based phylogenetic inference. This tool can observe the variation of evolutionary rate between positions leading to more precise evolution.	http://proddata.swmed.edu/ancescon/ancescon.php	53
PAML	Phylogenetic Analysis by Maximum Likelihood. A tool for constructing a phylogenetic tree and inferring ancestral sequence based on ML method.	http://abacus.gene.ucl.ac.uk/software/paml.html	54
MEGA 11	Inferring ancestral sequence based on ML or MP methods.	https://megasoftware.net/dload_wm_beta	55
FastML	Inferring ancestral sequence based on ML method that implements a large array of evolutionary models.	http://fastml.tau.ac.il/	56

Table 1. continued

tool	description	access link	ref.
ProtASR2	Inferring ancestral sequences based on protein structures using ML method. The method provides a more realistic inference of ancestral protein in terms of folding stability.	https://github.com/miguelarenas/protasr	57
PhyloBot	A web portal that can automatically generate ancestral sequences and allow for the exploration of mutational trajectories related to protein evolution.	http://www.phylobot.com	29
FireProt ^{ASR}	A fully automatic ASR tool that covers steps of extant sequence searching, multiple sequence alignment, phylogenetic tree construction, and ASR inferring.	https://loschmidt.chemi.muni.cz/fireprotasr/	30

of a FASTA-formatted text file containing a collection of protein sequences. The program reads the sequences and launches its automated analysis pipeline, which includes sequence alignment, phylogenetic model fitting, tests of branch support, and ancestral sequence reconstruction, by using different external software packages. Nevertheless, it is currently unclear whether the server remains active at the time of the preparation of this Perspective. More recently, a web-based suite FireProt^{ASR} was reported in 2021 for a fully automated ASR process based on a single query sequence.³⁰ The software runs the whole workflow using externally available and in-house (Jiří Damborský's group) developed algorithms. The web tool also offers user-defined practices by defining catalytic residues to identify biologically relevant sets of homologous sequences. Users can also initiate the ASR process with their own set of sequences to construct MSA or a phylogenetic tree. These fully automated workflows allow inexperienced users to obtain ancestral sequences based on a simple input of only sequence queries. In our opinion, the automated tools may be more suitable for enzyme engineers who want to obtain ASR constructs with improved robustness compared to the extant enzymes. On the other hand, for researchers with specific investigational aims, ancestral sequence construction using a step-by-step approach would be more suitable because it can also provide understanding of different catalytic or biophysical properties among enzymes in the evolutionary family.

Perhaps prior to gene synthesis and performing experimental validation, researchers can analyze the biophysical properties of ancestral candidates *in silico* using bioinformatic tools to predict the success of the ASR campaign. There have been significant developments in computational tools, and many have been designed with easy-to-use interfaces for general users to allow them to investigate protein biophysical properties and structures. This *in silico* analysis may save a lot of resources and experimental effort (Figure 2D). For predicting protein expression levels in *E. coli*, the SoluProt web server may be employed.³¹ Properties such as hydrophobic and charged residues, which are related to stability, can be analyzed based on sequences using EMBOSS Pepstats.³² Analysis for complexed data such as hydrogen bonds, hydrophobic clusters, and salt bridges can be analyzed by ProteinTools,³³ while substrate access tunnels can be identified by CaverWeb.³⁴ These web-based tools require protein structures as the input, which can be prepared by structure prediction tools such as ESMFold.³⁵ As a recent development, the protein complex and interactions with ligands can be predicted using state-of-the-art tools such as RoseTTAFold All-Atom³⁶ and AlphaFold 3.³⁷ Although these tools may not reflect the actual characteristics of the proteins, the additional *in silico* screening process is cost-free and can be practically implemented as a rational-based decision factor before initiating experimental investigations.

■ USE OF ASR FOR CREATING THERMOSTABLE BIOCATALYSTS

The general belief that ancestral enzymes are more thermostable than their extant ones has been driving many research programs to create thermostable enzymes using ASR. In the early 2010s, success in employing ASR approaches to improve the thermal stability of enzymes was demonstrated in selected enzymes such as 3-isopropylmalate dehydrogenase,⁵⁸ elongation factor proteins,⁵⁹ and glycyl-tRNA synthetase.⁶⁰ Since then, there has been an increasing number of reports using

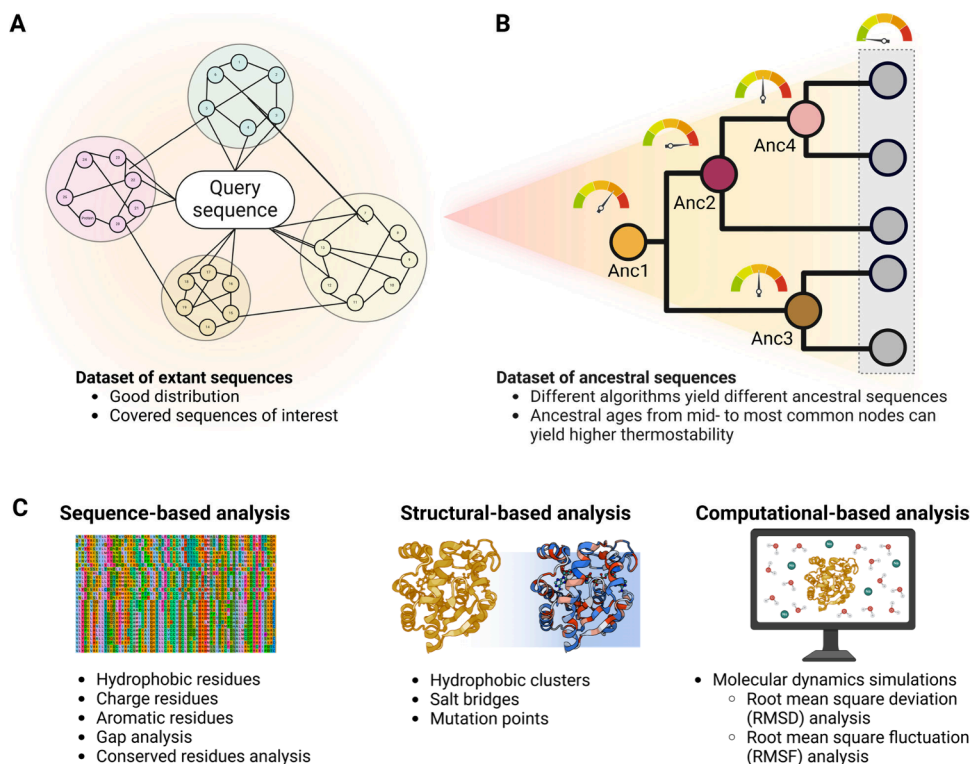


Figure 3. Key processes for improving and analyzing enzyme stability. (A) Collection of extant sequences. (B) Ancestral enzymes generally show better thermostability compared to extant enzymes. Anc stands for ancestral enzyme. (C) Tools for the analysis of origin in improvement of ancestral enzyme thermostability.

ASR to obtain thermostable ancestral proteins. The use of ASR for obtaining thermostable ancestral enzymes can follow the process described in Figure 2. In most cases, the successful reconstruction to obtain thermostable ancestors is attributed to the step of selecting the extant sequences (Figure 3A). Early studies indicate that more thermotolerant enzymes are likely to be found in ancestral sequences with older ancestral ages,^{61,62} however, recent studies reported that younger ancestral enzymes could also yield higher thermostability than older ones (Figure 3B).^{63–65} Numerous tools for the analysis of protein sequences and structures have recently been employed to comprehend insights of the superior stability in ancestral enzymes such as sequence-based analysis (e.g., hydrophobic and charge residues), structure-based analysis (e.g., salt bridges, hydrophobic clusters), and computational analysis (e.g., molecular dynamic simulations) (Figure 3C).^{66,67} In this section, we discuss noteworthy cases of enzyme stability improvements and present an analysis of key molecular features likely contributing to the improved properties.

P450 enzymes are among the major biocatalysts that can catalyze the functionalization of unactivated C–H bonds. Their reactions can be used to produce high-value chemicals such as drug metabolites or for the synthesis of drugs in the pharmaceutical industry. However, most applications of P450 enzymes are limited by their poor stability. In 2018, the Gillam group unveiled an exquisite study of resurrecting ancestral P450s with critical improvements in thermal tolerance. Ancestors of the CYP3 family of P450s were created using 138 extant CYP3 sequences from different vertebrate species, with the hypothesis to resurrect enzymes existing during the pre-Cambrian era.⁶⁸ The last vertebrate common ancestor denoted CYP3_N1 was characterized with a significant

improvement in thermostability compared to other extant CYP3s. For instance, its optimum temperature was elevated by ~ 20 °C compared to the extant CYP3A4, and the temperature at which half of the protein remains folded after 60 min ($^{60}T_{50}$) in CYP3_N1 was 30 °C greater than that of the extant CYP3A4. Analysis of the thermal stability at 50–60 °C indicates that CYP3_N1 has a half-life of 10 h, while other extant CYP3s have half-lives of less than 5 min. Moreover, CYP3_N1 was also noted for its enhanced solvent tolerance by displaying 2- and 15-fold increased stability toward 10% (v/v) methanol and 10% (v/v) acetonitrile treatment as compared to CYP3A4, respectively. Structural analysis indicates that CYP3_N1 contains more hydrophobic residues in the core region than CYP3A4, which may affect thermostability and solvent tolerance properties. As improvements in thermostability and solvent tolerance are often correlated,⁶⁹ this study has shown that the resurrected ancestral CYP3 enzyme is very useful for improving both properties with only a single operation.

The creation of ancestral enzymes with expanded working temperatures and pHs was also recently demonstrated with endoglucanase, an important enzyme for lignocellulosic biomass hydrolysis.⁷⁰ The bacterial endoglucanase sequence from *Bacillus subtilis* (Bs_EG) was selected as a query for ASR because of its potential applications in the biotechnology industry. Extant sequences for ancestral reconstruction were selected from three different phyla that diverged 3 billion years ago. The phylogenetic tree was constructed after ensuring that every sequence had conserved residues necessary for the enzyme function. Three ancestral enzymes (LFCA, last *Firmicutes* common ancestor; LCCA, last *Clostridia* common ancestor; LACA, last *Actinobacteria* common ancestor) were

selected for biochemical investigations. Among the ancestral enzymes, LFCA showed the best thermostability because it could maintain activity in a range of 30–90 °C (defined as meso- to hyper-thermophiles), which is higher than the extant Bs_EG. The phylogenetic tree analysis indicates that LFCA is the oldest ancestral (~2.8 billion years ago) enzyme among the three ancestral enzymes and shares 73% sequence identity with Bs_EG. Structural analysis did not reveal any substantial structural changes, and only a small number of mutations and displacements in α -helices and loops were revealed in the crystal structures.

In another example, ASR of carboxylic acid reductase (CAR), which catalyzes the reduction of aliphatic or aromatic acids to form aldehydes, was investigated. CARs are attractive for biotechnology applications, but the native enzymes generally lack sufficient stability for industrial applications. ASR of CAR is also challenging because the enzyme contains three domains with different structures. In 2019, Nicholas Harmer's group used 48 extant sequences of the CAR1 family, which were identified as CAR1 from different genera including *Mycobacteria*, *Nocardia*, *Streptomyces*, and *Tsukamurella* (out-group), to construct candidate ancestral enzymes.⁷¹ This study employed different ASR methods, including FastML, PAML, and Ancescon, at the ancestral inferring step to create various ancestral sequences. The four ancestral enzymes, including AncCAR-A by Ancescon, AncCAR-F by FastML, AncCAR-PA by PAML with gaps reconstructed by cross-mapping from AncCAR-A, and AncCAR-PF by PAML with gaps reconstructed by cross-mapping from AncCAR-F as the last common nodes of *Mycobacteria*, *Nocardia*, and *Streptomyces*, were identified and investigated. The results showed that AncCAR-A, AncCAR-PA, and AncCAR-PF retained good activities, while AncCAR-F was inactive even though it shared a high sequence identity (95% ID) with AncCAR-PF. These results indicate that the real properties of ancestral enzymes, although derived from predictions from similar tools, can be quite different. In this study, the most thermostable enzyme was AncCAR-A, which maintained 50% of its activity after incubation at 70 °C for 30 min; this stability is comparable to that of the most stable extant CAR enzyme from *Mycobacterium phlei* (MpCAR). AncCAR-PA and AncCAR-PF showed thermostability higher than those of extant CARs, retaining 50% activity after incubation at 65.1 and 65.4 °C for 30 min, respectively. Moreover, AncCAR-PA and AncCAR-PF showed half-lives at 37 °C that were 168 and 216 h longer than those of MpCAR, which was less than 24 h. Structural analysis showed that a major difference between AncCAR-PF and MpCAR is around the surface loop areas of domain A, which is shorter in CAR-PF than in MpCAR, possibly leading to less motion and more thermostability of the CAR-PF. This work demonstrated that significant differences in outcomes could be obtained using different sequence-inferring approaches. To improve the biophysical properties of predicted ancestral enzymes, it would be beneficial to compare the predicted ASR sequences using various techniques.

Another example of ASR that yielded ancestral enzymes with a large improvement in thermostability was reported for the case of diterpene cyclase.⁶³ In this work, the template utilized was the enzyme ent-copalyl diphosphate synthase from *Streptomyces platensis* (PtmT2). The top 250 sequences with highest similarity to the template enzyme were chosen for ASR. It was noted that among the 250 selected extant sequences, half of them were predicted as prenyltransferases containing

the DxDD motif of class II terpene cyclases, while the other half were not. After ASR candidates were obtained, the first four ancestral nodes closest to the template, including Anc01, Anc02, Anc03, and Anc04, were selected for further studies. The results showed that Anc03 and Anc04 could not be expressed in a soluble form when utilizing *E. coli* as a host, whereas Anc01 and Anc02 could be expressed and possessed melting temperatures around 4.9 and 39.7 °C higher than the extant enzyme (PtmT2), respectively. Further analysis of Anc01 showed that the enzyme contains most of the same conserved catalytic residues as PtmT2 and differed only on the protein surface. Anc02 showed a few variations in its active site region. These variations likely caused the lowering of Anc02 activity to around 6-fold as compared to PtmT2. MD simulations identified the flexible parts of PtmT2 and Anc01 and identified residues 344–353 as gate keepers of the active site cavity. At 343 K, the loop regions of Anc01 and PtmT2 showed high fluctuation and collapse to block the active site, possibly preventing substrate binding. On the other hand, Anc02 showed less fluctuation in the same loop region, explaining its ability to maintain activity at high temperatures. This work demonstrated an instance of ASR contributing additional thermostable candidate enzymes, although the native enzymes remained relatively unexplored. This would be beneficial for future investigations of structures and functions of diterpene cyclase in general.

Although many studies showed that the resurrected ancestral forms were more stable than the extant enzymes, a recent work published in 2024 showed that this trend may not be consistent.⁶⁵ The ASR investigation of CYP116B showed that among the 11 resurrected ancestral enzymes (N0, N1, N3, N23, N70, N104, N106, N115, N118, N213, and N374) derived from 467 extant sequences, ancestral N106, N115, and N374 showed the highest ¹⁵T₅₀ values among ancestral enzymes. However, the ¹⁵T₅₀ values of N106, N115, and N374 were lower than those of the known thermophilic CYP116B including P450-TT, P450-AX, and P450-TB; the ⁶⁰T₅₀ of N104 was also lower than those of P450-TT and P450-AX. It was interpreted that these ancestral enzymes were still mesophilic because the majority of the extant sequences used were mesophilic enzymes. The structure and sequence analysis of ancestral CYP116B compared with P450-TT revealed that the major mutations in ancestral CYP116B occurred in the loop region. These mutations likely impact enzyme stability, resulting in lower thermostability for ancestral CYP116B compared to P450-TT. Notably, the ancestral enzymes in this work were observed to perform different regioselective hydroxylation reactions and may serve as potential useful scaffolds for further exploring and improving new functions (see the next section).

The last example case under this category is the work by Livada et al. from a Pfizer research group in 2023, which presents a comparison of statistical data between ancestral and extant enzyme populations.⁶⁴ A panel of 56 ancestral enzymes of ene-reductases was constructed and compared for protein expression, thermostability, and reactivity against 57 extant enzymes. With regard to thermostability, the average melting temperature of the ancestral group (56 enzymes) was significantly higher than that of the extant group (57 enzymes) by approximately 9 °C. When comparing between a pair of closest ancestral and extant enzymes to minimize bias arising from mutations, the results indicated that the ancestral enzyme showed a *T_m* higher than that of the extant enzyme by

Table 2. Improvements of the Thermostability and Other Biophysical Properties in Selected Ancestral Enzymes

enzymes	extant sequences	selected ancestors	thermostable properties ^a	other properties (e.g., solubility, pH, or solvent)	ref
cytochrome P450 (CYP3)	vertebrate species, e.g., CYP3A4	CYP3_N1	<ul style="list-style-type: none"> •⁶⁰T₅₀ is 66 °C (+30 °C of CYP3A4) •Half-life at 50 °C is ~10 h (improved >100 fold of CYP3A4) 	increase in solvent tolerance	68
endoglucanase	<i>Bacillus subtilis</i> (Bs_EG)	LFCA	<ul style="list-style-type: none"> •³⁰T₅₀ is 79 °C (+11 °C) 	increase in pH tolerance and expression yield	70
carboxylic acid reductase	<i>Mycobacterium phlei</i> (MpCAR)	AncCAR-A, AncCAR-PA, AncCAR-PF	<ul style="list-style-type: none"> •³⁰T₅₀ is 65.1 to 70.0 °C (+15.1 to 20.0 °C) •T_m 67–68 °C (+35 °C) •half-life at 37 °C is 41–216 h (improved 9–48 fold) 	increase in pH tolerance and expression yield	71
diterpene cyclase	<i>Streptomyces platensis</i> (PtmT2)	Anc01, Anc02	<ul style="list-style-type: none"> •T_m 50.7 °C (+4.7 °C) and 85.7 °C (+39.7 °C) for Anc01 and Anc02, respectively 		63
ene-reductase	<i>Candida albicans</i> (EBP1)	56 ancestral enzymes	<ul style="list-style-type: none"> •average ΔT_m + 9 °C 	slightly higher expression on average	64
cytochrome P450 (CYP116B)	Bacterial CYP116s	N106, N115, N374	<ul style="list-style-type: none"> •¹⁵T₅₀ is 50 °C (+10 to 12 °C) 	increase in expression yield of N106 and N115	65
L-amino acid oxidase	<i>Pseudoalteromonas piscicida</i> (PpAROD)	AncLAAO-N4	<ul style="list-style-type: none"> •¹⁰T₅₀ is ~65 °C •T_{opt} is 50 °C 	high activity (>80%) at pH 6.5–8.0	75
phenylalanine/tyrosine ammonia-lyase	<i>Rhodotorula glutinis</i> (RgPAL) and <i>Trichosporon cutaneum</i> (TcPAL)	MEGA_A1	<ul style="list-style-type: none"> •T_m is 70.9 °C (+4.7 °C of RgPAL, + 9.2 °C of TcPAL) •70% activity remaining after treatment at 37 °C for 2 weeks 		76
diaryl alcohol dehydrogenase	<i>Kluyveromyces marxianus</i> (D10)	A64	<ul style="list-style-type: none"> •¹⁵T₅₀ is 57.5 °C (+15.1 °C) •T_m is 61.7 °C (+14.9 °C) 	optimum pH shift from 5.5 (D10) to 7.5	77
phenolic acid decarboxylase	<i>Bacillus subtilis</i> (BsPAD)	N31	<ul style="list-style-type: none"> •T_m is 78.1 °C (+23.6 °C) •half-life at 60 °C is 45 h 	increase in expression yield	78
PETase	<i>Ideonella sakaiensis</i> PETase (IsPETase)	GrAnc8, GrAnc6	<ul style="list-style-type: none"> •GrAnc8 T_m is 63.2 °C (+20.0 °C) •GrAnc6 T_m is 50.9 °C (+7.7 °C) 	soluble expression increases by 2.8 fold (GrAnc8) and 26.8 fold (GrAnc6)	79

^aImprovements compared to the extant enzymes are indicated in circle brackets. Abbreviations; T_m, melting temperature; ^xT₅₀, temperatures at which half the population of proteins remains active after being exposed at specific time, where x is the time in minutes; T_{opt}, optimal temperature.

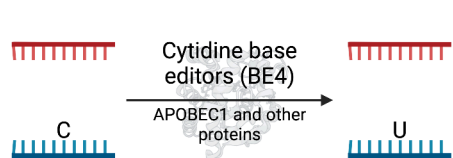
approximately 0.16 °C per point of mutation. However, T_m was not directly correlated with the ancestral age. This study is particularly useful in that it provides a comparison of statistical data, which should be useful for understanding the difference between ancestral and extant groups.

In summary, the reconstructed ancestral proteins often exhibit increased thermostability. The thermal stability improvement is likely linked to evolutionary driving factors. For instance, the most thermostable endoglucanase LFCA likely emerged during the Hadean and Archean eons when ocean temperatures were approximately 60–70 °C.⁷⁰ Two other ancestral endoglucanases also displayed correlations between their thermostability and the predicted seawater temperatures during their respective time periods. Nevertheless, the correlation between evolutionary processes and environmental factors could not be observed clearly in some other cases. In the case of CYP3, the oldest node, N1, corresponding to the earliest vertebrates exhibited higher thermal tolerance than the extant enzymes despite the fact that ancestral and extant enzymes are found under similar environmental conditions.⁶⁸ For CYP116B, the ancestral version did not exhibit increased thermal stability, and there was no correlation with ancestral age.⁶⁵ The lack of correlation is primarily due to the characteristics of extant sequences, which mostly originated from mesophiles. A recent hypothesis proposed that the stability of constructed ancestral enzymes may involve a consensus landscape with a preference toward residues encoding native (low-energy) conformations before evolving to modern proteins.⁷² Consequently, the correlation between the thermostability and time origin of ASR remains unclear. As ASR can only infer the most probable ancestor

based on the inputs provided and the use of algorithms, it is plausible that an inherent systematic bias may arise based on the quality of inputs and algorithm characteristics.^{19,73} Therefore, the experimental investigation of alternative ancestral sequences should be expanded or ancestral sequences should be inferred by different methods to assess the reliability of the observed ancestral properties before drawing any evolutionary conclusions. Although predicting thermal properties based on evolutionary data is currently challenging, some common parameters nevertheless have been found in highly stable ancestral enzymes such as specific mutations in the loop region,^{63,70} lengthening or shortening of the surface loop,⁷¹ and the number of of charge residues or hydrophobic residues.^{68,74} These factors should be considered in conjunction with considerations of ancestral age and environmental gaps when selecting ancestral sequences for reconstruction.

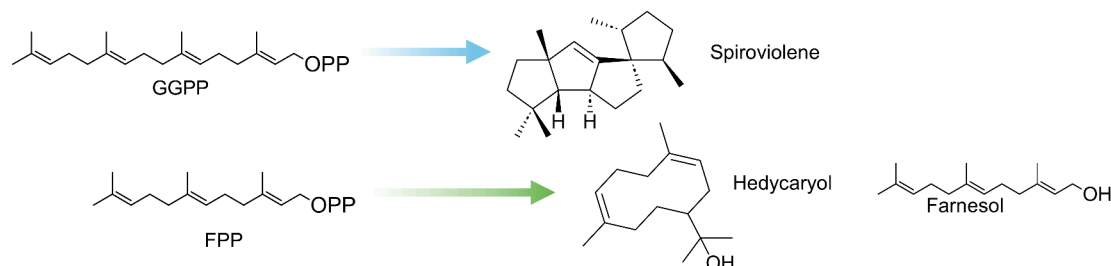
Overall, this section presents the overview of ancestral enzymes that can deliver thermostable enzymes superior to extant ones. Previous investigations have shown that in addition to thermostability improvement, other properties such as solvent tolerance, pH tolerance, or protein expression could also be improved in the mentioned studies. The increase in enzyme robustness demonstrates that ancestral enzymes are useful starting templates for further activity fine-tuning to broaden applications of biocatalysts. The improvements in thermostability as well as notable interesting biocatalyst properties of selected ancestral enzymes are summarized in Table 2 to highlight the ongoing successful cases of ASR in creating robust enzymes. Altogether, all results demonstrate

A Improvement of base-editing efficiency



Cytidine base editors	BE4	AncBE4max
APOBEC variant	Wildtype	Ancestral
%Relative mRNA level	1	5.2
Base editing at three genomic loci in HEK293T cell lines	48 ± 8.0%	90 ± 1.5%

B Engineering specificity of terpene cyclases



Variants	Substrate condition	Approx. product percentage ratio		
		Spiroviolene	Hedycaryol	Farnesol
SvS-wt	Single	40	40	20
SvS-A2		73	18	9
SvS-wt		67	25	8
SvS-A2	Competition	76	20	4
SvS-A2-A224I		-	~100	-
SvS-A2-W156Y		~100	-	-

Figure 4. Uses of ASR on enzyme activity and specificity engineering. (A) Improvement of base-editing efficiency in cytidine base editor (BE4) systems. (B) Reaction schemes and specificities of wild-type, ancestral, and variant terpene cyclases.

that ASR is a powerful technique for creating thermostable enzymes.

ENGINEERING ENZYME ACTIVITY, PROMISCUITY, AND FUNCTIONS

Besides showing potential in improving enzyme stability, ASR has been a promising tool for enhancing enzyme activity and widening substrate utilization. By dating back to ancestral sequences, the resurrected enzymes are hypothesized to possess robustness in terms of stability and can utilize a wider range of substrates compared to those of the extant enzymes. Thus, the use of ASR for improving enzyme activity, substrate and catalytic promiscuity, and applicational efficiency has been illustrated in several studies.

In 2018, ASR was used to improve the single-nucleotide base-editing efficiency by improving the expression level of APOBEC1 cytidine deaminase, which was the bottleneck of the process.⁸⁰ Five ancestral cytidine deaminases were reconstructed based on the extant APOBEC1, and all enzymes were tested for their cytidine base conversion efficiency. Anc689 and Anc687 demonstrated higher activities in base editing in HEK293T cell line systems compared to the extant enzyme activity. The base-editing activities were found to vary among different plasmid doses, genomic loci, and cell line types. The highest activity was obtained from ancestral Anc689 with a codon-optimized system (AncBE4max). The mRNA and protein expression level were increased by 5.2-fold compared to that of the wild-type enzyme in the HEK293T cell line system, contributing to the increase in editing efficiency application as presented in Figure 4A.

The Per-Olof Syrén group demonstrated the use of ASR on the enzyme spiroviolene synthase, which catalyzes the class I

cyclization of diterpenes.⁸¹ The spiroviolene synthase from *Streptomyces violens* (SvS) was used as an extant sequence for the ASR process. Interestingly, most ancestral enzymes in this case could overcome the stability–activity trade-off. The ancestral enzymes showed improvements in stability, *i.e.*, a 7–13 °C increase in T_m and a 2–4-fold increase in soluble protein expression compared to the extant SvS.⁸¹ Some ancestral enzymes also exhibited higher activity and broader substrate utilization, such as the ability to use farnesyl pyrophosphate (FPP) in addition to the native substrate geranylgeranyl pyrophosphate (GGPP) of the extant enzyme. Due to having the highest thermostability, the ancestral SvS-A2 was chosen for further engineering based on guidance from crystal structure analysis, molecular docking, and mechanistic studies to obtain variant terpene cyclases with highly improved specificity (Figure 4B).⁸² In the case of using either FPP or GGPP as a single substrate, the approximate percentage ratio of spiroviolene/hedycaryol/farnesol product formation was 40:40:20 in wild-type SvS (SvS-wt), whereas the percentage ratio of product formation was 73:18:9 in SvS-A2. In the case of using mixed substrates of FPP and GGPP, the ratios were 67:25:8 (wild-type SvS) and 76:20:4 (SvS-A2). The data indicate that SvS-A2 is more specific, whereas wild-type SvS is more promiscuous because more hedycaryol and farnesol products were generated in the reactions of the wild-type enzyme compared to those of SvS-A2. Intriguingly, engineering of SvS-A2 resulted in several highly specific variants such as variant A224I, which is an FPP-specific variant with almost 100% hedycaryol formation, while variant W156Y is a GGPP-specific variant with almost 100% spiroviolene formation. This work demonstrated that ancestral enzymes can provide useful scaffolds for further enzyme engineering processes. Molecular

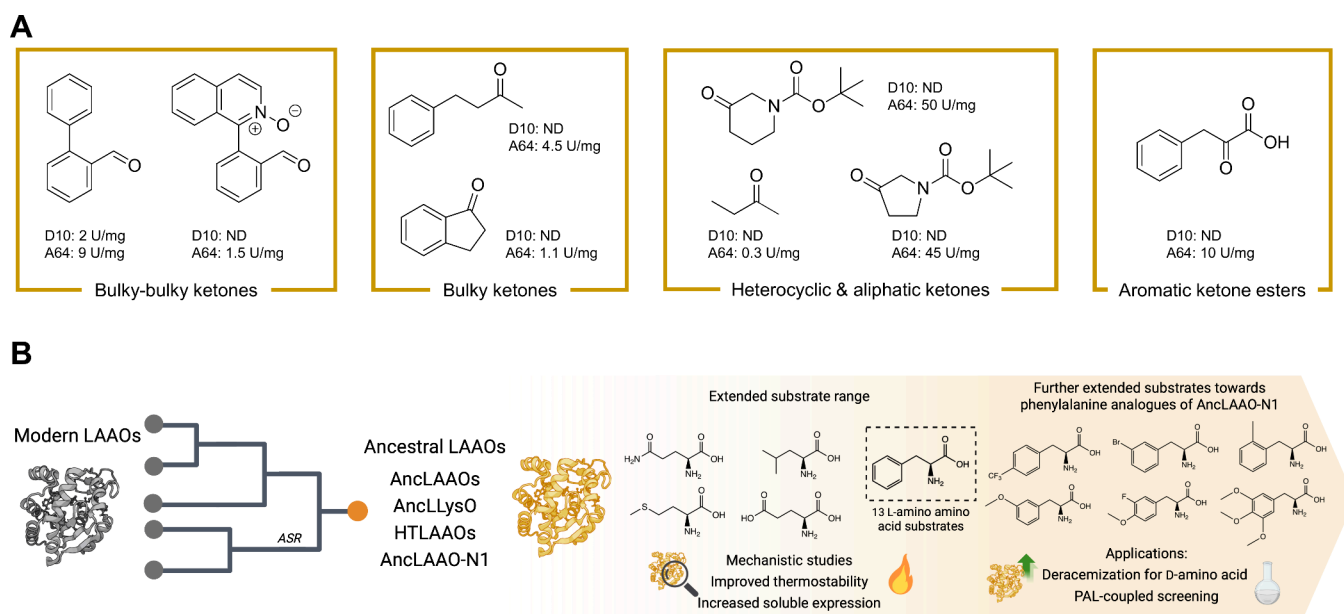


Figure 5. ASR for expanding the substrate utilization scope. (A) Substrate utilization profiles of ancestral alcohol dehydrogenase A64 compared to those of extant D10. Specific activity is reported for each compound. ND: not detected/not determined. (B) Ancestral LAAOs and extended substrates.

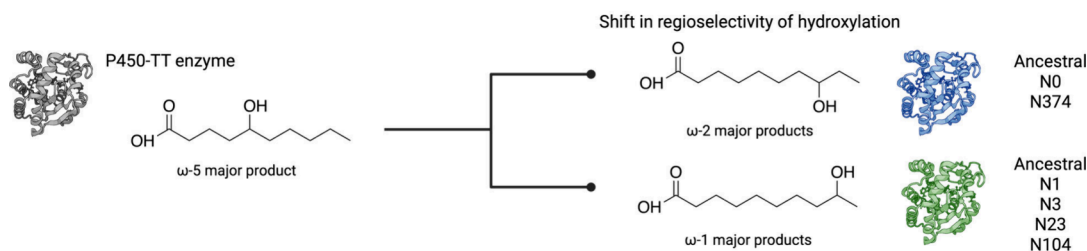


Figure 6. Regioselectivity shift in ancestral P450s.

docking information and an in-depth understanding of enzyme catalysis are instrumental in helping researchers succeed in engineering campaigns to improve the substrate specificity of ancestral enzymes compared with the extant enzymes while maintaining thermostability.

Resurrected enzymes from ASR often demonstrate an expansion of substrate promiscuity. One example showcasing broadening of the substrate range is the reconstruction of ancestral alcohol dehydrogenase.⁷⁷ This work started from genome mining analysis using a diaryl alcohol dehydrogenase (*KpADH*) as a query sequence to obtain a range of ADHs that were then submitted for ancestral reconstruction. The descendent D10 was identified as the most efficient ADH with desired ketone reduction activity, and its ancestral ADHs from evolutionary branch were chosen for characterization. A64 was identified as the most promising ancestral enzyme based on expression and reaction analysis. The A64 enzyme has a wider substrate utilization range, including bulky ketone, heterocyclic ketone, and aliphatic ketone derivatives. Substrates that are preferred by A64 compared to the extant enzyme D10 are shown in Figure 5A. In addition, A64 was also observed to have shifted pH-activity profiles and significantly increased thermostability, as previously mentioned (Table 2). Overall, this study demonstrates a strategy to create promiscuous and robust enzymes by combining enzyme mining and ASR.

Another example of expanding enzyme promiscuity was reported for L-amino acid oxidases (LAAOs). Selected LAAOs are highly specific to their amino acid substrates, such as those found in the reactions of L-phenylalanine oxidase, L-arginine oxidase, and L-lysine oxidase.⁸³ However, some LAAOs can catalyze the oxidation of various amino acids.^{84,85} Intrigued by the variety of specificity and promiscuity of LAAOs in nature, several studies led by Nakano et al. demonstrated the use of ASR on LAAOs to expand or regulate substrate promiscuity.^{75,86,87} Ancestral LAAOs (AncLAAOs) with improved thermostability were identified and exploited in one-pot deracemization reactions for D-amino acid production.⁸⁸ One of the AncLAAOs was recently reported to have expanded substrate profiles toward oxidation of *ortho*-, *meta*-, and *para*-monosubstituted, disubstituted, and trisubstituted phenylalanine.⁸⁹ Monosubstituted phenylalanine with either an electron-donating or electron-withdrawing group in the *meta*- and *para*- locations was shown to be the most efficient substrate for the enzyme. The AncLAAO could also be coupled with a phenylalanine ammonia lyase (PAL)-based screening platform, demonstrating the use of the ancestral LAAO as a tool for high-throughput screening.⁸⁹ ASR of LAAOs and their extended usable substrates are summarized in Figure 5B.

In the previous section, we discussed significant improvement in expression and thermal stability of ancestral P450s

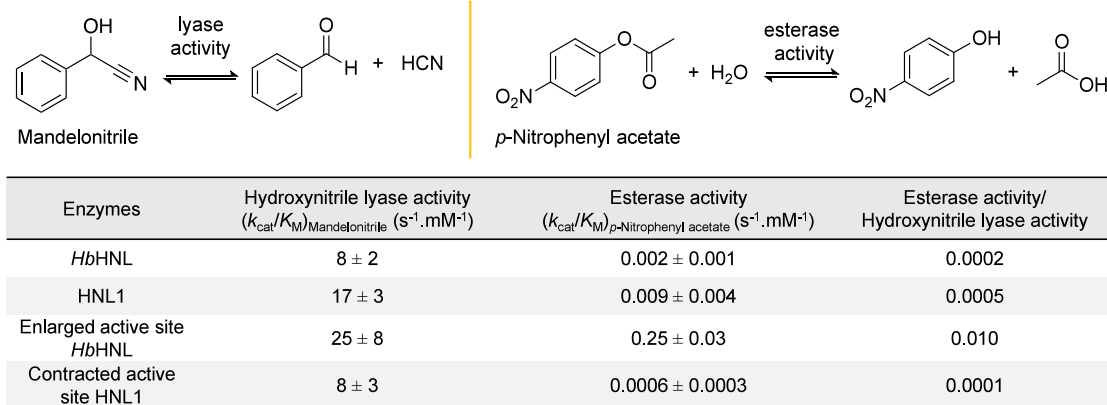


Figure 7. Reaction schemes and comparison of the catalytic promiscuity of ancestral and variant HNLs.

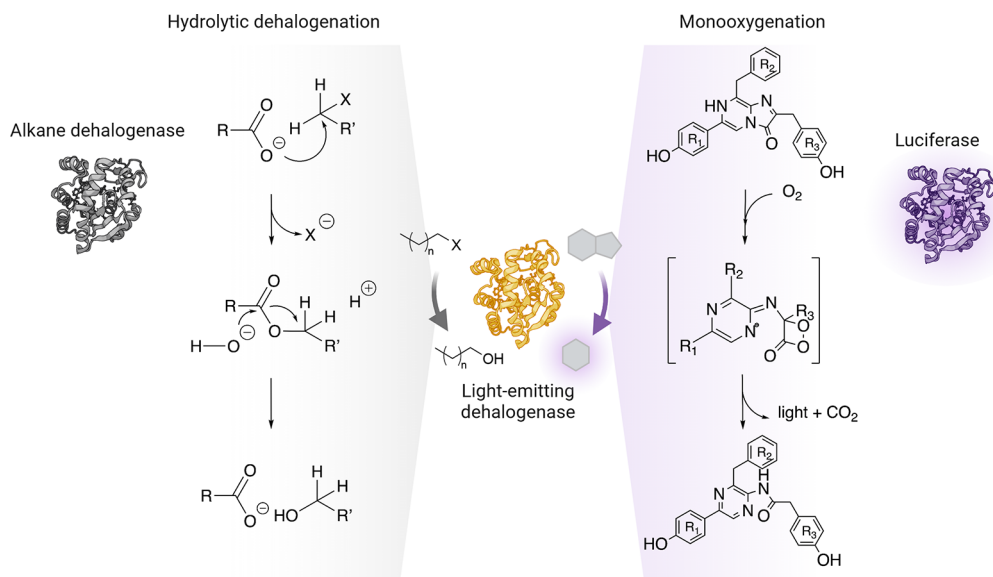


Figure 8. Dual function of AnCHLD-RLuc: hydrolytic alkane dehalogenase and monooxygenase (luciferase) activity.

from the CYP116B families.⁶⁵ The constructed ancestors were also investigated for their regioselectivity in the fatty acid hydroxylation of multidomain self-sufficient P450 monooxygenases and compared to the extant CYP116B46 (P450-TT), which favors the ω -5 midchain hydroxylation. Although the ancestral CYP116Bs were catalytically active with decanoic acid, only N70 showed activities comparable with the extant enzymes, while other enzymes required 5-fold enzyme concentrations in bioconversion experiments. The regioselectivity of hydroxylation was different among the ancestral enzymes. The ω -2 hydroxylation was the major product of N0 and N374, whereas the ω -1 hydroxylation product was found in N1, N104, N3, and N23 as the major activity. Thus, the ancestral enzymes showed a shift in selectivity from ω -5 to ω -1 and ω -2 as shown in Figure 6. Sequence alignment of the active sites of ancestral enzymes and extant P450-TT showed differences at residues 205 and 206. To understand the shift in regioselectivity, A205T and F206W substitutions were introduced to the extant P450-TT to create single and double mutations. The results showed that although the variants gave ω -5 hydroxylation as a major product, an increase in the production of ω -3 and ω -4 products was also observed, particularly in the double mutation variant. It was explained that this was due to increased polarity at the active site.

One of the studies on expanding both catalytic promiscuity (the ability to catalyze different kinds of reactions) and substrate utilization *via* ASR was demonstrated in the reaction of hydroxynitrile lyases from *Hevea brasiliensis* (HbHNL).⁹⁰ A reconstructed ancestral HNL1 enzyme taken from the evolutionary path of esterases and hydroxynitrile lyases exhibited increases in both activity and promiscuity. Mandelonitrile and *p*-nitrophenyl acetate were used to compare the catalytic activities between (hydroxynitrile) lyase activity and esterase activity, respectively (Figure 7). The HNL1 exhibited a 2.5-fold increase in lyase activity (mandelonitrile cleavage) and a 4.5-fold increase in esterase activity (*p*-nitrophenyl acetate hydrolysis), illustrating that the catalytic promiscuity toward esterase activity was increased by 2-fold for HNL1 compared to the extant HbHNL (Figure 7). The reaction scope of HNL1 was also expanded toward bulkier compounds because lyase activity can use 2-nitro-1-phenylethanol and several cyanohydrins, while esterase activity can use naphthyl acetate as substrates. The X-ray structural analysis showed that the active site of the ancestral HNL1 was larger by 60% compared with that of the extant enzyme, explaining how HNL1 could utilize larger hydroxyl nitrile substrates. Moreover, site-directed mutagenesis to enlarge the active site of HbHNL to mimic that of HNL1 was performed to study the

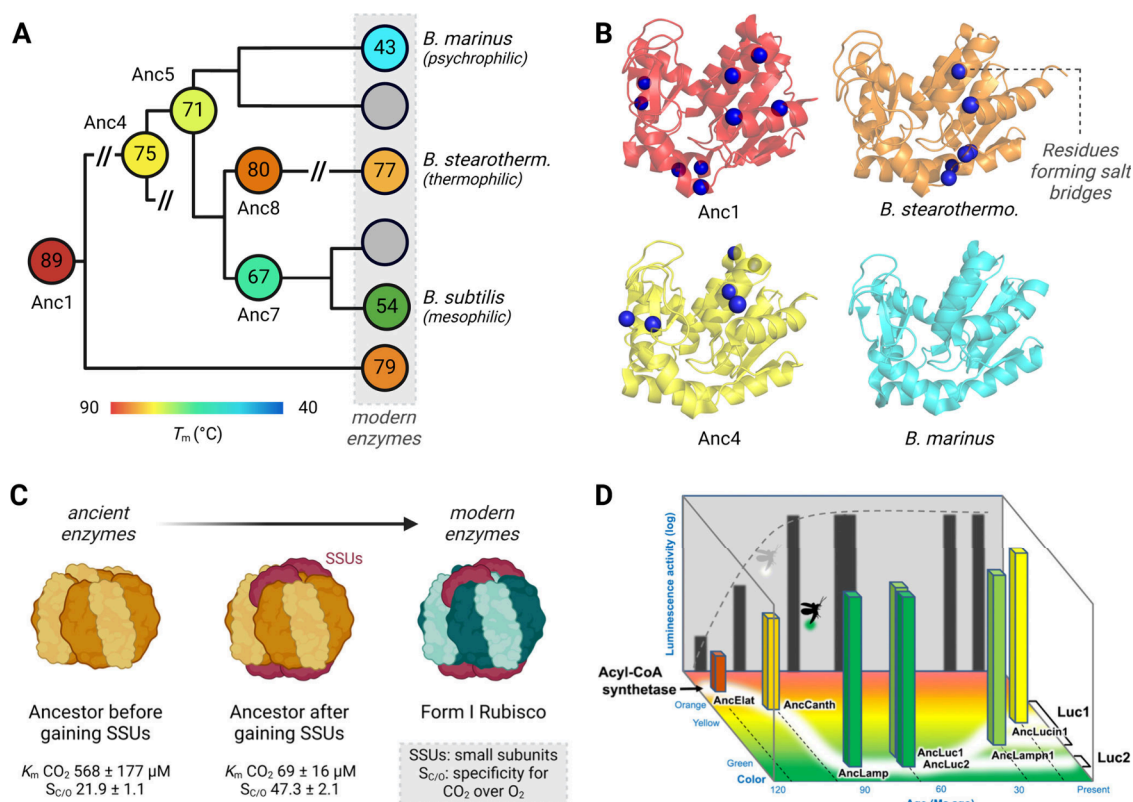


Figure 9. ASR for understanding the evolution of biocatalyst property and functions. (A, B) Evolution of thermostability and key structural features for thermoadaptation in adenylate kinases, respectively. T_m is labeled for each enzyme. (C) Molecular assembly evolution of Form I Rubiscos for CO_2 specificity. (D) Scheme of the evolution of firefly bioluminescence. Reproduced with permission from Oba et al.¹⁰⁴ Copyright 2020 American Association for the Advancement of Science.

effects of a larger active site. The 3-point mutation enlarging the active site of *HbHNL* led to an increase in promiscuity. Likewise, the replacement of the three residues in HNL1 with the corresponding residues from *HbHNL* resulted in a decrease in the promiscuity. This work suggests that the lyase activity of *HbHNL* might have evolved from the esterase. The overall results also imply that understanding how ancestral enzymes evolved should be useful to increase the promiscuity and activity of modern enzymes.

Another study by Chaloupkova and Damborsky et al. employed ASR to create an enzyme exhibiting dual function: light-emission and alkane dehalogenation (Figure 8).⁹¹ The group reconstructed a common ancestor of haloalkane dehalogenases (HLDs) and *Renilla* luciferase (RLuc), which both possess different catalytic activities, although they belong to the same HLD subfamily. It is intriguing that RLuc is the only characterized monooxygenase among the clustered HLD enzymes. It was thus attractive to elucidate the evolution pathway of this enzyme family. The created AncHLD-RLuc was found to possess both hydrolase (haloalkane dehalogenation) and monooxygenase (bioluminescence) activities despite having much lower activity than its descendant enzymes. In addition, the ancestral enzyme could use a wider range of substrates for dehalogenation activity as compared to extant HLDs. Structural analysis revealed that a catalytic pentad of the descendant dehalogenase and luciferase was highly conserved in the AncHLD-RLuc active site. Nevertheless, careful analysis of the ancestral and extant enzyme structures led to the identification of key residues adjacent to the catalytic active site that can promote an activity exchange between

luciferase and halogenase. This work demonstrated the concept of exchanges of functions between two different enzyme commission (EC) classes during the course of evolution and paved a way to construct multifunctional biocatalysts.

RECONSTRUCTING ANCIENT ENZYMES FOR EVOLUTION AND MECHANISTIC STUDY

Understanding the evolutionary track of enzymes can help elucidate how enzymes attain their characteristic functions such as thermal adaptation, enzyme function, and selectivity during the evolutionary process. As enzymes in early evolution are assumed to be promiscuous and stable, it is useful to resurrect them to decipher the origin of enzyme structure and function. In the previous section, we discussed that the ancestral diterpene cyclase is active and solubilized well, leading to a successful crystallization that could not typically be achieved with the wild-type enzymes.⁸² The ASR study could successfully provide structural insights and computational analyses, which were then employed to acquire mechanistic insights into cyclase-catalyzed cyclization and could be further guided through engineering to obtain specific and active terpene cyclase variants. This finding as well as other mentioned cases has shown that ancestral sequences are beneficial for obtaining optimally stable enzymes for structural elucidation and further in-depth studies of enzyme catalysis.

Understanding how natural evolution resulted in improved enzyme stability and activity in response to marked changes in environmental features is crucial to harnessing the full potential of enzymes for biocatalysis. For instance, it has been proposed

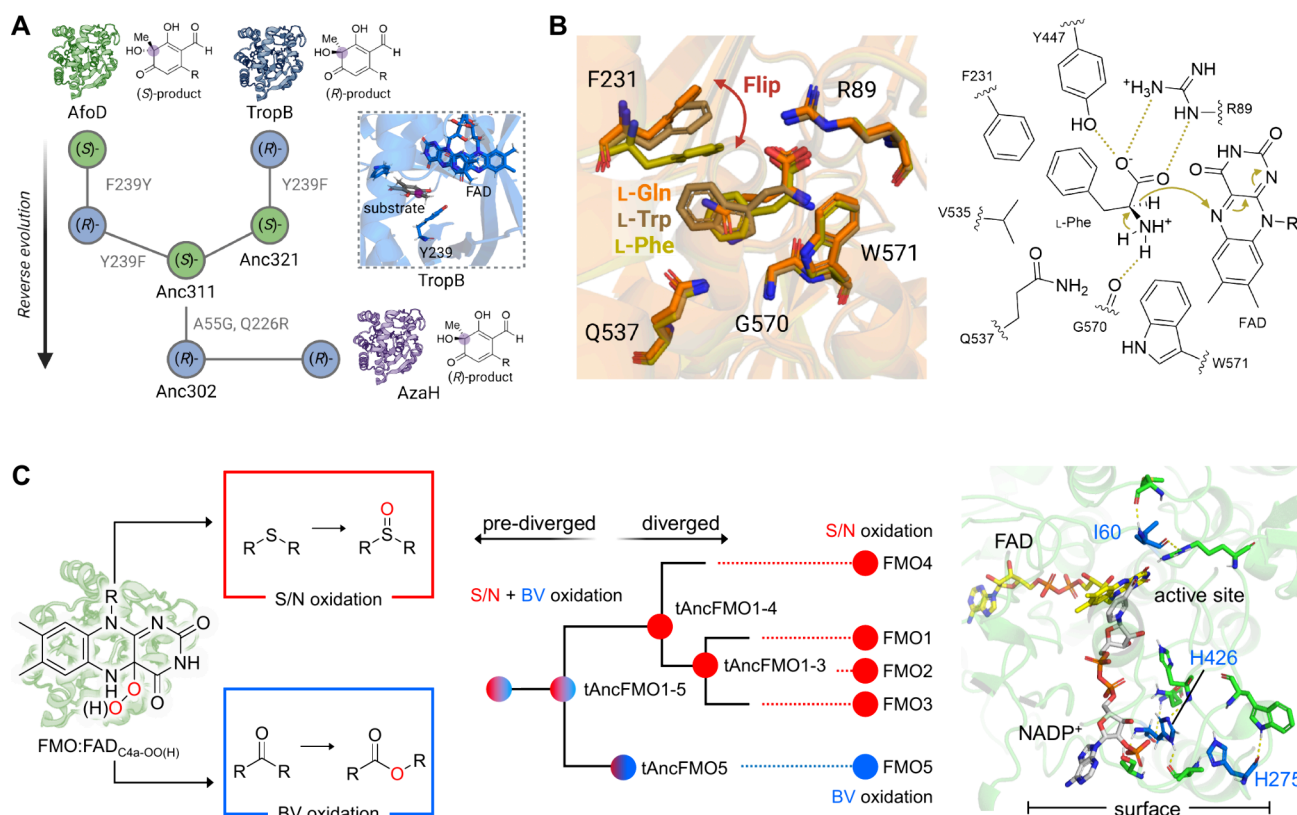


Figure 10. ASR for understanding structural features controlling enzymatic stereoselectivity, substrate specificity, and reaction mechanisms. (A) Deciphering the evolution of FDMO stereoselectivity. The structure of TropB was generated from PDB ID 6NET. (B) Ancestral LAAO (AncLAAO-N5) reveals substrate recognition and reaction mechanisms. The structures were generated from PDB IDs 7C4L, 7C4M, and 7C4N. (C) A functional diversification along evolution in FMOs and key interactions responsible for modulating the oxidation activities. The tAncFMO1-4 variant complex with FAD and NADP⁺ was created using AlphaFold 3. Mutated residues are shown in blue.

that prehistoric life on Earth consisted of a hot environment (known as the “hot-start”) before cooling of the Earth required adaptation to cooler temperatures. As early studies on the reconstruction of ancestral proteins showed that the oldest nodes in phylogenetic trees are often the most thermophilic,¹⁴ the observed results sparked a question of how the temperature adaptation affects enzyme evolution. A study in 2017 by Dorothee Kern’s group accordingly used ASR to trace the evolution of adenylate kinase activity and stability from the hot-start toward various modern enzymes (Figure 9A).⁶⁶ Results of the study identified that salt bridges in the ancestral kinases are responsible for high thermostability (Figure 9B). This key structural feature sequentially disappears during evolution toward colder environments and then reappears in species that subsequently adapt to hot places. Interestingly, one of the ancestral enzymes was found to oppose the commonly assumed activity–stability trade-off concept by showing both thermal tolerance and preserved catalytic activity at low temperatures. Adenylate kinase is an essential enzyme found in nearly every life form for maintaining cellular nucleotide concentrations and is also important for biocatalytic ATP regeneration.⁹² This finding highlighted that this strategy can be used by evolution in different ways to increase the biocatalyst fitness, in addition to revealing the molecular features determining thermostability.

In a related theme, ASR can be used to illuminate the emergence of adaptations in enzyme specificity for investigating the structural features defining the selective catalytic function of biocatalysts. Ribulose-1,5-bisphosphate carboxy-

lase/oxygenase (Rubisco) is the enzyme with the highest abundance in nature, and its ability to capture CO₂ is key to a sustainable carbon-neutral bioeconomy.^{93–95} Despite its promising activity, this enzyme is not particularly specific to CO₂ over molecular oxygen (O₂), which is an inhibitor of CO₂ assimilation activity.⁹⁶ The competitive binding between O₂ and CO₂ to Rubisco severely hinders the efficiency of carboxylation activity. However, the mechanism of enzyme selectivity and differentiation among the two gases is poorly understood. It has been proposed that noncatalytic small subunits of the enzyme are responsible for the high CO₂ specificity of modern-day form I Rubiscos.⁹⁷ In the recent work by Schulz et al., the authors inferred ancestral sequences to recapitulate the evolution of form I Rubiscos before and after gaining the small subunits and investigated effects of different molecular assemblies on the catalytic function.⁹⁸ When small subunits were introduced into the ancestral enzymes, which consist mainly of large subunits, the results showed that gaining small subunits did increase the CO₂/O₂ specificity of the ancestral enzymes (Figure 9C). Structural analysis showed that the assembly of small subunits does not result in observable active site rearrangements; thus, the small subunits were not directly responsible for an increase in specificity. The study concludes that Rubiscos have evolved to recruit the binding of a small subunit to improve their catalysis through allosteric effects and to obtain high specificity. As Rubisco is a slow-evolving enzyme in nature, this study sheds light on how molecular assembly has been achieved to drive the specific evolution of enzyme.

Fireflies are an ancient species that can be dated back to 100 million years ago. Their bioluminescence has attracted great interest as detection systems, such as useful diagnostic tools in biomedical research⁹⁹ and for the detection of pesticides in food and agricultural residue.¹⁰⁰ Naturally, the luminescence colors of different firefly families can vary widely between species despite the commonality in their enzymatic reactions.¹⁰¹ It has been proposed that firefly bioluminescence originally functioned as an aposematic warning display toward predators and later acquired a role in sexual communication.¹⁰⁰ There have been two major gene types of modern firefly luciferases isolated from various species, and gene analyses suggested that these two major families of firefly luciferases might have evolved from a common origin before the diversification of luciferases.^{102,103} Because the color of ancestral bioluminescence cannot be elucidated from fossil records, it is difficult to predict the original function of firefly light. To shed light on the original function of the firefly light, Oba and Shirai's groups reconstructed seven ancestral firefly luciferase genes using the maximum likelihood method and characterized their bioluminescent properties.¹⁰⁴ Measurement of the luminescence spectra of the seven ancestors showed that the perceived colors can vary from orange-red to green-yellow regions (Figure 9D). By mapping the geological ages with the bioluminescent wavelengths of the resurrected ancestral luciferases, the study concluded that ancient firefly luciferases emitted green light for warning display purposes, while some species evolved their yellowish light for complex sexual communications. Additionally, the study also found that one of the firefly lineages evolved from a nonluminescent ancestor. This finding has led to the hypothesis that beetle luciferases originated from another enzyme class (fatty acyl-CoA synthetase) before being subjected to adaptive evolution. This study has also provided a new perspective in expanding the firefly bioluminescence colors.

It is generally critical to understand the structural features that control the selectivity in biocatalytic reactions to achieve tunable selectivity. For this point, the ASR approach has been used to navigate a historical sequence space of inherent selective preferences in biocatalysts. Flavin-dependent monooxygenases (FDMOs) are attractive for their mild reaction conditions with low catalyst loadings and utilization of molecular oxygen as the stoichiometric oxidant^{105,106} and because they have evolved as highly stereoselective catalysts. A recent study was conducted to decipher the stereoselectivity evolution of FDMOs.¹⁰⁷ Three FDMOs including TropB, AfoD, and AzaH, which are responsible for the biosynthesis of azaphilone through enzymatic oxidative dearomatization, were subjected to evolution and mechanistic studies. These three FDMOs display interesting features in that they display conserved regioselectivity; however, they have different stereoselectivity to form either (*R*)- or (*S*)-products. Ancestral enzymes were created and tested to probe the impact of each residue on the stereochemical outcome of the dearomatization reaction. The ancestral enzymes also showed desirable properties, such as good expression solubility and high thermostability. Analysis of the experimental results and the phylogenetic tree of FDMOs revealed that stereoselectivity switches are observed as two steps. First, a switch between tyrosine and phenylalanine residues in the ancestors of the same clades can strongly influence the formation of either the (*R*)- or (*S*)-product (Figure 10A). This mechanism was observed only in ancestral TropB and AfoD, suggesting that

they have a common ancestor. The other mechanism was found in the AzaH clade in that multiple active site residues are involved in stereoselectivity control (Figure 10A). Eventually, the ancestral enzymes with a superior stereoselectivity for producing the (*S*)-product and high thermal stability were acquired, demonstrating the application of ASR for understanding enzyme mechanisms and promoting engineering stereoselectivity efforts.

Another successful case to rationalize the enzyme selectivity discovered a key residue defining the substrate specificity. Intrigued by the high substrate promiscuity of a vanillyl alcohol oxidase from *Diplodia corticola* (DcVAO), Eggerichs et al. employed ASR in a recent study to perform evolutionary analysis of the substrate scope of the fungal VAO family.¹⁰⁸ DcVAO and its homologues are from different clades of fungal VAOs. Three ancestral enzymes, including the last common and two middle nodes, were constructed and subjected to a substrate profiling investigation. Results identified that the middle-node ancestors had substrate profiles more diverse than those of the last common ancestor. Ancestral model-guided mutagenesis analyses were conducted in DcVAO. The study showed that amino acid substitutions in the catalytic center proximal to substrate binding of DcVAO were responsible for diversifying or specializing substrate profiles and the enzyme catalytic function.

Obtaining ancestral enzymes can also be helpful for developing a structural and mechanistic understanding of enzymes at the molecular level. The efforts to understand the enzyme catalytic mechanisms and structural basis are difficult to achieve when enzymes of interest are not robust and poorly expressed. A study by Nakano et al. demonstrated how the ancestral *L*-amino acid oxidases (LAAOs) specify substrate recognition and catalyze their reactions.⁷⁵ By using the ASR approach, stable and good soluble expression yield of ancestral LAAOs could be obtained.⁸⁶ The structure of one ancestral enzyme was elucidated with distinct active-site structures from the homologous modern LAAOs.⁷⁵ Cocrystals of the ancestral enzyme AncLAAO-N5 with various *L*-amino acids were also obtained and revealed a key phenylalanine residue, which may control the conformational dynamics important for substrate binding (Figure 10B). This feature enables the ancestral enzyme to recognize a broad range of substrates. The structural analysis and site-directed mutagenesis also identified conserved residues that are crucial for the activity of ancestral and other LAAOs (Figure 10B). Finally, the insights from structural analysis of the ancestors were used to increase the substrate specificity toward a non-native *L*-valine substrate. Further ancestral enzyme reconstruction by the same group was extended to uncharacterized LAAO sequences in an attempt to identify new LAAOs.¹⁰⁹ The generated ancestral enzyme was shown to have unprecedented high specificity toward *L*-lysine and it was a thermophilic enzyme. Structural analysis of the *L*-lysine-specific ancestral enzyme in comparison to other LAAOs suggested that the ancestral enzyme has a narrow substrate binding pocket, and its active site conformation can be changed upon substrate binding, thus exhibiting high specificity toward *L*-lysine.

ASR was also used to extend evolutionary studies by resurrecting ancestral activities to identify sequence determinants of enzyme functional diversification. Four flavin-monooxygenase (FMO) ancestral nodes were reconstructed to study the divergence of the specialized activities—sulfide and amine (*S*/*N*) or heteroatom and Baeyer–Villiger (BV)

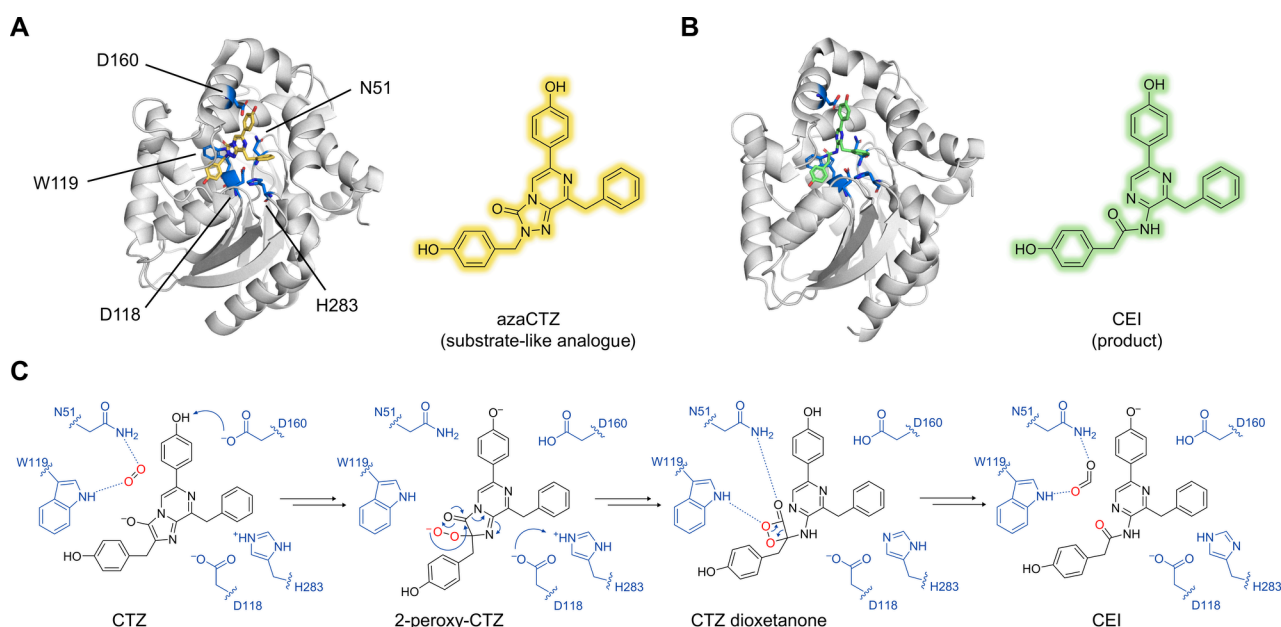


Figure 11. Elucidating the mechanism of *Renilla*-type luciferases using the X-ray structures of ancestral enzyme cocrystallized with ligands. (A, B) Cocrystal structures of azaCTZ- (PDB ID 7QXR) and CEI-bound AncFT luciferase (PDB ID 7QXQ), respectively. CTZ, coelenterazine; CEI, coelenteramide. (C) Catalytic mechanisms proposed from structural analysis of the ancestral complexes. The complete proposed catalytic reaction mechanism is described in Schenkmyerova et al.¹¹¹

oxidations among FMO1–FMO5 paralogous proteins (Figure 10C).¹¹⁰ The resurrected prediverged tAncFMO1-5 showed both S/N and BV oxidations, whereas the diverged ancestral enzymes showed their specialized activities (Figure 10C). The tAncFMO1-4 and tAncFMO1-3 showed only S/N oxidation activity, while the tAncFMO5 showed both oxidation activities with less reactivity toward S/N oxidation, suggesting that the descendant enzymes evolved from a common bifunctional ancestor. The sequence alignment and mutational analysis further identified the sequence determining the BV and S/N functionalities; for instance, introducing three substitutions from a sequence of the promiscuous tAncFMO1-5 to the S/N-selective tAncFMO1-4 could reinstall the BV oxidation activity in tAncFMO1-4. Structural analysis of the tAncFMO1-4 variant revealed the three substitutions engaged in epistatic interactions responsible for mediating the BV oxidation activity *via* coupled interactions of the FMO with the NADP⁺ and FAD cofactors (Figure 10C). This study sheds light on the pivotal role of the nucleotide cofactors FAD and NADP⁺ in that their binding during the catalytic cycle could influence the type of oxygenation reactions. FMOs are enzymes that catalyze selective oxygenations. Therefore, understanding the structural features governing diversification or specialization of functions is essential for developing biocatalytic tools in the future.

An additional interesting approach to harnessing the ancestral sequence to unlock catalytic mechanisms is an investigation of the ancestral enzyme reaction with substrate analogues. The catalytic mechanism of a widely exploited bioluminescent system, *Renilla* luciferase (RLuc), was just recently reported based on the ancestral enzyme structural analysis, even though this enzyme has been studied for more than 40 years.¹¹¹ The RLuc mechanism remained elusive, mostly due to its intrinsic difficulty in crystallization and its native substrate not being stable.¹¹² These obstacles could be overcome by enzyme engineering of the ancestral RLuc, which provided a more rigid structure compared to extant

enzymes.¹¹³ The study by Schenkmyerova and co-workers then cocrystallized the engineered ancestral RLuc (AncFT) with a developed nonoxidizable substrate-like analogue and the bioluminescence product (Figure 11A and B). The enzyme X-ray structures co-complexed with both ligands in catalytically favored conformations were solved. The data were used to obtain the binding modes of native-substrate coelenterazine (CTZ) and short-lived intermediates 2-peroxy-CTZ and CTZ dioxetanone. The combined results of complementary kinetics and molecular dynamics studies were used to successfully unravel the underlying catalytic mechanisms (Figure 11C).¹¹¹ This work clearly illustrates the usefulness of ancestral scaffolds in improving protein crystallizability and evolvability, providing mechanistic insights into these enzymatic reactions. Altogether, this section summarizes examples from recent years in which ancestral enzymes could be resurrected to create desirable scaffolds that can contribute to the fundamental understanding and be combined with enzyme engineering to create useful biocatalysts.

■ POTENTIAL OF ANCESTRAL ENZYMES FOR INDUSTRIAL APPLICATIONS

Because the development of new and efficient biocatalysts is important for green chemistry, the use of ASR to create new enzymes thus has a high potential for supporting industrial applications. For instance, ancestral ω -transaminases with improved specific activities can be used for the production of 12-aminododecanoic acid, a non-natural compound with industrial significance as a constituent building block of nylon-12 (Figure 12A). These ancestral enzymes are also highly active toward a range of other ω -amino acids and α,ω -diamines, which can lead to the expansion of polyamide synthesis.¹¹⁴ The recent ancestral reconstruction of thermo-stable ene-reductases (EREDs) by Pfizer Inc. also showcased the potential asymmetric reduction of activated alkenes by these enzymes for producing pharmaceutically relevant

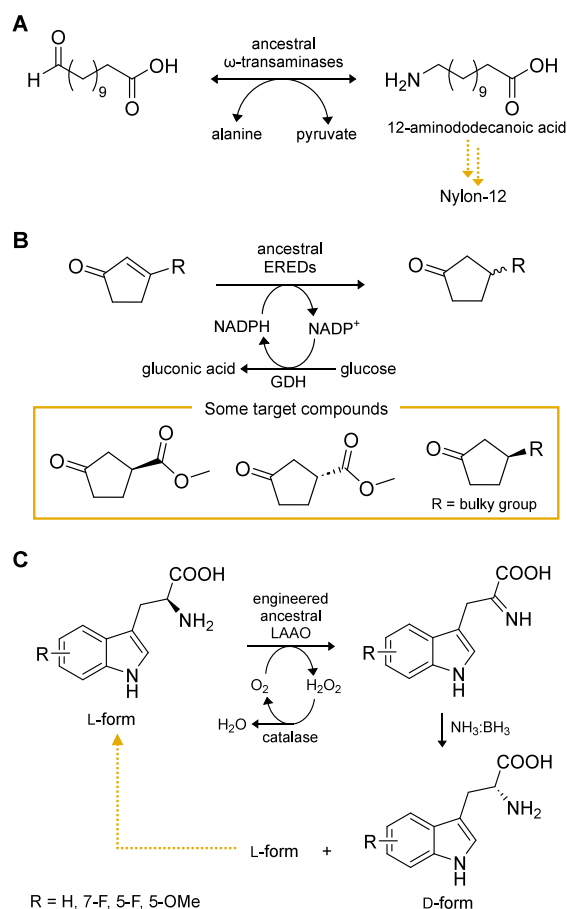


Figure 12. Reactions of ancestral enzymes with the potential for industrial applications. (A) Production of 12-aminododecanoic acid by ω -transaminases. (B) ERED-catalyzed reduction of cyclopentenones. GDH, glucose dehydrogenase. (C) Chemoenzymatic production of D-tryptophan derivatives.

optically active compounds (Figure 12B).⁶⁴ Ancestral reconstruction and enzyme engineering by the Ito and Nakano groups to obtain thermostable and high-specificity L-amino acid oxidase (LAAO) activity toward L-tryptophan has also showcased the ability to engineer the ability to perform chemoenzymatic synthesis of enantiopure D-tryptophan derivatives at a preparative scale (Figure 12C).¹¹⁵ D-Amino acids are essential precursors for manufacturing therapeutic peptides. The development of synthetic methods that are green, use renewable and abundant feedstocks such as L-amino acids, and produce specific products is important for sustainable industrial applications. In addition to the above examples, other enzymes such as lipases, acyltransferases, halogenases, polyketide synthases, etc., as well as cofactors and enzymes used in substrate recycling/regeneration systems that have already been in use in various industries, may also receive benefit from the use of ASR.^{116,117}

CURRENT LIMITATIONS

Despite the fact that ASR has been successfully applied in various studies to reconstruct ancient enzymes with enhanced stability, activity, and novel functionalities, researchers often face challenges from several factors. The first thing to be considered in commencing the ASR process is sequence collection. Sequence collection is an important factor influencing the quality of ancestral inference, as the inclusion

of erroneous extant sequences can propagate errors through the ASR process. Sequences containing transcription errors and miscalled exons, introns, insertions, deletions, and frameshifts are increasingly common in current sequence databases due to inaccurate enzyme sequence determination.^{19,118} The rise in poor-quality sequences could be a downside of increasing speed in sequencing technology. The rapid expansion of sequence databases has resulted in an increasing dependence on automated annotation, which has been shown to significantly increase the likelihood of errors in various record annotations.¹¹⁸ Aside from mistakes in protein sequences, errors such as taxonomic misclassification can cause problems at the subsequent stage of tree generation. While experimental characterization is still a bottleneck to elevate the accuracy of the database, it is crucial to carefully organize the sequence alignment to eliminate sequences containing errors while maintaining the broadest possible coverage of the extant sequence space.¹⁹ In addition to the mistaken sequence sources, limited sequence data for certain protein families can also restrict the accuracy of the ASR.

Phylogenetic tools and sequence analysis are essential to identifying the key molecular features, *i.e.*, amino acid substitutions and gain/loss of small sequences in evolutionary branches, that are likely to alter or enhance protein properties.¹⁹ The quest to fully understand the evolutionary relationship of protein sequences, structures, and functions lies at the heart of ASR research, as it is crucial to validate the resurrected sequences and explain how natural evolution produces modern proteins. For instance, the trends in lineages leading to thermophiles have been suggested to relate mostly to the evolution of biological contexts, *i.e.*, meso- to thermophile adaptation.⁶⁵ This means that it may not be possible to create a more thermostable enzyme simply by inferring and resurrecting the ancestors of a group of extant sequences without considering the possible evolutionary pathway in the context of the biological niche of the extant enzymes. Nevertheless, evolutionary analysis has not yet entered the mainstream biocatalysis field. A considerable hurdle that nonevolutionary enzyme engineers may face when using ASR is the choice of reliable phylogeny and sequence inferring methodology. Among numerous methods for inferring phylogenetic trees and ancestral sequences, the ML method is currently the preferred because it generally gives good accuracy compared to other methods with practical computing effort.¹¹⁹ Although varying the algorithms may not be necessary to obtain phylogeny reconstructions,^{120,121} it should be noted that the ML algorithm does not account for uncertainty in the reconstruction and thus can lead to errors in the sequence inference.¹⁹ Therefore, the inferred phylogeny and sequences need to be carefully assessed to draw evolutionary conclusions or gain the best approximation. The resulting reconciled species tree needs to be critically evaluated according to the systematic relationships among the taxa under examination; however, no current software is a substitute for expert knowledge in evaluating how realistic the gene family tree is in light of the species phylogeny.⁷³ It has been suggested that the recent large-scale phylogenomic investigations across all kingdoms of life may provide a guide to systematically assess the organismal relationships among taxa and provide support for phylogenetic inferences.⁷³

Perhaps a more critical factor for reliable ancestral reconstructions is to obtain a high quality MSA. Previous studies have shown that the accuracy of the phylogenetic tree

describing the evolutionary relationship of the protein family has been found to have less impact on ASR accuracy than the alignment.^{122,123} The MSA errors could promote significant biases in evolutionary reconstructions; however, it remains unclear how MSA methodological approaches impact ASR.¹²² The conventional two-dimensional arrays of aligned residues might involve implicit judgments about which residues in an alignment are homologous.¹⁹ In practice, it can be quite challenging to determine these relationships in highly variable regions of an alignment or between distantly related proteins. It has been shown that integrating uncertainty algorithms (a heuristic approach that reconstructs ancestral residues and gap states by integrating information from several alignment methods) during the sequence alignment improves ASR accuracy and the accuracy of downstream structural and functional inferences.¹²³ The development of probabilistic modeling of insertion and deletion events such as GRASP has the potential to radically improve ASR accuracy when the model reflects the true underlying evolutionary history.¹²⁴ Further studies are still required to thoroughly evaluate the reliability of these approaches. Consequently, the MSA methods should be carefully selected to accurately determine ancestral states with confidence.

Another considerable challenge to successful ancestral reconstruction could be at the downstream pipeline level, such as the ability to overexpress ancestral enzymes in a soluble form and the ability to use modern substrates. Although many examples presented in this Perspective were cases in which the ancestral enzymes were expressed well, this may be simply due to the insoluble enzymes not being mentioned much in the literature. Successfully expressing and obtaining properly folded ancient proteins in modern systems can still be challenging because several factors such as codon usage, cellular protein folding machinery, and environmental contexts should be taken into account when facing soluble expression issues.^{125,126} It is also likely that the substrate spectrum and catalytic promiscuity could be altered during evolution. Ancestral enzymes might have different functions compared to their modern counterparts. Therefore, the integration of computational predictions into the ASR process to refine ancestral sequences is encouraged. Furthermore, employing high-throughput screening techniques could allow exploration of the functional landscape of reconstructed enzymes.

■ FUTURE PERSPECTIVES

We have shown that the ASR approach has become a powerful tool to advance the biocatalysis field for delivering enzymes with exquisite properties and activity as well as unraveling key molecular features governing function and selectivity. As obtaining robust enzymes is always required for biotechnological and synthetic applications, it is undoubtedly clear that reconstructing the ancient enzymes can serve as an essential tool in providing desirable stable biocatalysts. Many experimentally characterized reconstructed ancestral enzymes display notably higher stability and evolvability, thereby promoting the trend of harnessing ASR to construct thermostable scaffolds. ASR has also been proven to be a valuable tool for engineering activity, substrate, and function promiscuity. It has been hypothesized and observed that ancestral enzymes are “generalists” with promiscuous substrates and functions; yet, they are not efficient. On the noteworthy basis that the ancestral enzymes are susceptible to adapting to different fitness landscapes and protein stability

generally promotes evolvability,¹²⁷ the ancestral enzymes represent ideal scaffolds to be coupled with complementary protein engineering tools, (*i.e.*, rational-based design) to create more efficient enzymes. A growing interest has been seen in repurposing protein scaffolds identified through ASR as potential starting points for the generation of new enzyme activities. It should be noted that the field of *de novo* protein design has made remarkable progress over the past decade through advancements in bioinformatics as well as artificial intelligence tools.¹²⁸ Nevertheless, its success largely depends on understanding the relationships between protein sequence, structure, and function. We expect that the combinatorial approach employing ASR and other enzyme engineering tools will be broadly used to provide a wide range of novel and robust biocatalysts.

It is clearly shown that ASR has become a valuable tool to provide insights into the changes within protein sequences and structures across evolution. More specifically, this approach represents a hallmark to solving diverse fundamental questions of biocatalysts. ASR provides robust scaffolds for crystallization and other experiments that otherwise could not be achieved with modern enzymes. ASR allows us to identify amino acid residues that are crucial for protein functions. Identifying these residues might not be possible by only comparing existing proteins. Intrigued by the many studies reviewed here, we foresee that future work will employ ASR to reveal the mechanisms that cause functional diversity in specific protein families.

The ASR process may still be challenging in the downstream steps because it relies on accurate gene synthesis and experimental characterizations. It is now more possible to apply computational methods to analyze the inferred sequences prior to experimental validation. Current state-of-the-art bioinformatics and artificial intelligence tools such as AlphaFold¹²⁹ and RoseTTAFold³⁶ have been emerging as practical tools to analyze protein structures and predict ligand and protein interactions close to experimental quality. Additionally, there are numerous tools to predict physical characteristics of proteins such as thermostability^{130,131} and solubility¹³¹ and, most recently, a platform to predict protein functions.¹³² These computational tools can be employed for *in silico* refinement of the ancestral candidates before validating experimental studies and thus should increase the success rate of obtaining optimized enzymes. It has been remarked that enzyme engineering is the third wave that revolutionizes biocatalysis.⁴ We believe that the ASR approach will have an impact as an emerging technology to drive the design and evolution of biocatalysts.

■ AUTHOR INFORMATION

Corresponding Author

Pimchai Chaiyen – School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Wangchan Valley, Rayong 21210, Thailand; orcid.org/0000-0002-8533-1604; Email: pimchai.chaiyen@vistec.ac.th

Authors

Kridsakorn Prakinee – School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Wangchan Valley, Rayong 21210, Thailand; orcid.org/0000-0003-2421-1518

Suppalak Phaisan – School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Wangchan Valley, Rayong 21210, Thailand

Sirus Kongjaroon – School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Wangchan Valley, Rayong 21210, Thailand; orcid.org/0009-0007-0351-2326

Complete contact information is available at:
<https://pubs.acs.org/10.1021/jacsau.4c00653>

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge support from Vidyasirimedhi Institute of Science and Technology (VISTEC), NSRF via the Program Management Unit for Human Resources & Institutional Development (PMU-B) Research and Innovation Grant B05F640089, Thailand Science Research and Innovation (TSRI) Grant FRB670026/0457, and KasikornBank Public Company Limited. Figures were created using PyMOL and BioRender.

REFERENCES

- (1) Wu, S.; Snajdrova, R.; Moore, J. C.; Baldenius, K.; Bornscheuer, U. T. Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem., Int. Ed.* **2021**, *60*, 88–119.
- (2) Yi, D.; Bayer, T.; Badenhorst, C. P. S.; Wu, S.; Doerr, M.; Höhne, M.; Bornscheuer, U. T. Recent trends in biocatalysis. *Chem. Soc. Rev.* **2021**, *50*, 8003–8049.
- (3) France, S. P.; Lewis, R. D.; Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au.* **2023**, *3*, 715–735.
- (4) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the third wave of biocatalysis. *Nature.* **2012**, *485*, 185–194.
- (5) Phintha, A.; Chaiyen, P. Rational and mechanistic approaches for improving biocatalyst performance. *Chem. Catal.* **2022**, *2*, 2614–2643.
- (6) Pongpamorn, P.; Watthaisong, P.; Pimviriyakul, P.; Jaruwat, A.; Lawan, N.; Chitnumsub, P.; Chaiyen, P. Identification of a Hotspot Residue for Improving the Thermostability of a Flavin-Dependent Monooxygenase. *ChemBioChem.* **2019**, *20*, 3020–3031.
- (7) Prakinee, K.; Phintha, A.; Visitsatthawong, S.; Lawan, N.; Sucharitakul, J.; Kantiwiriyanitch, C.; Damborsky, J.; Chitnumsub, P.; van Pée, K.-H.; Chaiyen, P. Mechanism-guided tunnel engineering to increase the efficiency of a flavin-dependent halogenase. *Nat. Catal.* **2022**, *5*, 534–544.
- (8) Kongjaroon, S.; Lawan, N.; Trisrivirat, D.; Chaiyen, P. Enhancement of tryptophan 2-monooxygenase thermostability by semi-rational enzyme engineering: a strategic design to minimize experimental investigation. *RSC Chem. Biol.* **2024**, *5*, 989.
- (9) Yang, J.; Li, F.-Z.; Arnold, F. H. Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering. *ACS Cent. Sci.* **2024**, *10*, 226–241.
- (10) Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **2004**, *5*, 366–375.
- (11) Gaucher, E. A.; Govindarajan, S.; Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature.* **2008**, *451*, 704–707.
- (12) Nicoll, C. R.; Massari, M.; Fraaije, M. W.; Mascotti, M. L.; Mattevi, A. Impact of ancestral sequence reconstruction on mechanistic and structural enzymology. *Curr. Opin. Struct. Biol.* **2023**, *82*, 102669.
- (13) Pandya, C.; Farelli, J. D.; Dunaway-Mariano, D.; Allen, K. N. Enzyme Promiscuity: Engine of Evolutionary Innovation. *J. Biol. Chem.* **2014**, *289*, 30229–30236.
- (14) Akanuma, S.; Nakajima, Y.; Yokobori, S.-i.; Kimura, M.; Nemoto, N.; Mase, T.; Miyazono, K.-i.; Tanokura, M.; Yamagishi, A. Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 11067–11072.
- (15) Spotlight on protein structure design. *Nat. Biotechnol.* **2024**, *42*, 157–157.
- (16) Hochberg, G. K. A.; Thornton, J. W. Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* **2017**, *46*, 247–269.
- (17) Siddiq, M. A.; Hochberg, G. K. A.; Thornton, J. W. Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* **2017**, *47*, 113–122.
- (18) Joy, J. B.; Liang, R. H.; McCloskey, R. M.; Nguyen, T.; Poon, A. F. Y. Ancestral Reconstruction. *PLoS Comput. Biol.* **2016**, *12*, No. e1004763.
- (19) Thomson, R. E. S.; Carrera-Pacheco, S. E.; Gillam, E. M. J. Engineering functional thermostable proteins using ancestral sequence reconstruction. *J. Biol. Chem.* **2022**, *298*, 102435.
- (20) Pinto, G. P.; Corbella, M.; Demkiv, A. O.; Kamerlin, S. C. L. Exploiting enzyme evolution for computational protein design. *Trends Biochem. Sci.* **2022**, *47*, 375–389.
- (21) Mascotti, M. L. Resurrecting Enzymes by Ancestral Sequence Reconstruction. *Methods Mol. Biol.* **2022**, 2397, 111–136.
- (22) Spence, M. A.; Kaczmarek, J. A.; Saunders, J. W.; Jackson, C. J. Ancestral sequence reconstruction for protein engineers. *Curr. Opin. Struct. Biol.* **2021**, *69*, 131–141.
- (23) Fitch, W. M. Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Syst. Zool.* **1971**, *20*, 406–416.
- (24) Hartigan, J. A. Minimum Mutation Fits to a Given Tree. *Biometrics.* **1973**, *29*, 53–65.
- (25) Pupko, T.; Pe, I.; Shamir, R.; Graur, D. A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. *Mol. Biol. Evol.* **2000**, *17*, 890–896.
- (26) Yang, Z.; Kumar, S.; Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics.* **1995**, *141*, 1641–50.
- (27) Zhang, J.; Nei, M. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* **1997**, *44*, S139–S146.
- (28) Huelsenbeck, J. P.; Bollback, J. P. Empirical and Hierarchical Bayesian Estimation of Ancestral States. *Syst. Biol.* **2001**, *50*, 351–366.
- (29) Hanson-Smith, V.; Johnson, A. PhyloBot: A Web Portal for Automated Phylogenetics, Ancestral Sequence Reconstruction, and Exploration of Mutational Trajectories. *PLoS Comput. Biol.* **2016**, *12*, No. e1004976.
- (30) Musil, M.; Khan, R. T.; Beier, A.; Stourac, J.; Konegger, H.; Damborsky, J.; Bednar, D. FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. *Brief. Bioinform.* **2021**, *22*, bbaa337.
- (31) Hon, J.; Marusiak, M.; Martinek, T.; Kunka, A.; Zendulka, J.; Bednar, D.; Damborsky, J. SoluProt: prediction of soluble protein expression in *Escherichia coli*. *Bioinformatics.* **2021**, *37*, 23–28.
- (32) Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276–W279.
- (33) Ferruz, N.; Schmidt, S.; Höcker, B. ProteinTools: a toolkit to analyze protein structures. *Nucleic Acids Res.* **2021**, *49*, W559–W566.
- (34) Stourac, J.; Vavra, O.; Kokkonen, P.; Filipovic, J.; Pinto, G.; Brezovsky, J.; Damborsky, J.; Bednar, D. Caver Web 1.0: identification

- of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Res.* **2019**, *47*, W414–W422.
- (35) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. **2023**, *379*, 1123–1130.
- (36) Krishna, R.; Wang, J.; Ahern, W.; Sturmfels, P.; Venkatesh, P.; Kalvet, I.; Lee, G. R.; Morey-Burrows, F. S.; Anishchenko, I.; Humphreys, I. R.; et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science*. **2024**, *384*, No. eadl2528.
- (37) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. **2024**, *630*, 493–500.
- (38) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
- (39) The Uniprot Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531.
- (40) Hon, J.; Borko, S.; Stourac, J.; Prokop, Z.; Zendulka, J.; Bednar, D.; Martinek, T.; Damborsky, J. EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities. *Nucleic Acids Res.* **2020**, *48*, W104–W109.
- (41) Oberg, N.; Zallot, R.; Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *J. Mol. Biol.* **2023**, *435*, 168018.
- (42) Kanehisa, M.; Goto, S.; Kawashima, S.; Nakaya, A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* **2002**, *30*, 42–46.
- (43) Sievers, F.; Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **2018**, *27*, 135–145.
- (44) Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066.
- (45) Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–7.
- (46) Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobittg, M.; Montanyola, A.; Chang, J. M.; Taly, J. F.; Notredame, C. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **2011**, *39*, W13–7.
- (47) Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One*. **2010**, *5*, No. e9490.
- (48) Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **2014**, *30*, 1312–3.
- (49) Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–21.
- (50) Huelsenbeck, J. P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. **2001**, *17*, 754–755.
- (51) Drummond, A. J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **2007**, *7*, 214.
- (52) Foley, G.; Mora, A.; Ross, C. M.; Bottoms, S.; Sützl, L.; Lamprecht, M. L.; Zaugg, J.; Essebler, A.; Balderson, B.; Newell, R.; et al. Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). *PLoS Comput. Biol.* **2022**, *18*, No. e1010633.
- (53) Cai, W.; Pei, J.; Grishin, N. V. Reconstruction of ancestral protein sequences and its applications. *BMC Evol. Biol.* **2004**, *4*, 33.
- (54) Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591.
- (55) Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027.
- (56) Ashkenazy, H.; Penn, O.; Doron-Faigenboim, A.; Cohen, O.; Cannarozzi, G.; Zomer, O.; Pupko, T. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **2012**, *40*, W580–W584.
- (57) Arenas, M.; Bastolla, U. ProtASR2: Ancestral reconstruction of protein sequences accounting for folding stability. *Methods Ecol. Evol.* **2020**, *11*, 248–257.
- (58) Miyazaki, J.; Nakaya, S.; Suzuki, T.; Tamakoshi, M.; Oshima, T.; Yamagishi, A. Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis. *J. Biochem.* **2001**, *129*, 777–82.
- (59) Gaucher, E. A.; Thomson, J. M.; Burgan, M. F.; Benner, S. A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature*. **2003**, *425*, 285–8.
- (60) Shimizu, H.; Yokobori, S.; Ohkuri, T.; Yokogawa, T.; Nishikawa, K.; Yamagishi, A. Extremely thermophilic translation system in the common ancestor commonote: ancestral mutants of Glycyl-tRNA synthetase from the extreme thermophile *Thermus thermophilus*. *J. Mol. Biol.* **2007**, *369*, 1060–9.
- (61) Stetter, K. O. Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. B* **2006**, *361*, 1837–1843.
- (62) Hobbs, J. K.; Shepherd, C.; Saul, D. J.; Demetras, N. J.; Haaning, S.; Monk, C. R.; Daniel, R. M.; Arcus, V. L. On the Origin and Evolution of Thermophily: Reconstruction of Functional Precambrian Enzymes from Ancestors of *Bacillus*. *Mol. Biol. Evol.* **2012**, *29*, 825–835.
- (63) Hueting, D. A.; Vanga, S. R.; Syrén, P.-O. Thermoadaptation in an Ancestral Diterpene Cyclase by Altered Loop Stability. *J. Phys. Chem. B* **2022**, *126*, 3809–3821.
- (64) Livada, J.; Vargas, A. M.; Martinez, C. A.; Lewis, R. D. Ancestral Sequence Reconstruction Enhances Gene Mining Efforts for Industrial Ene Reductases by Expanding Enzyme Panels with Thermostable Catalysts. *ACS Catal.* **2023**, *13*, 2576–2585.
- (65) Jones, B. S.; Ross, C. M.; Foley, G.; Pozhydaieva, N.; Sharratt, J. W.; Kress, N.; Seibt, L. S.; Thomson, R. E. S.; Gumulya, Y.; Hayes, M. A.; et al. Engineering Biocatalysts for the C-H Activation of Fatty Acids by Ancestral Sequence Reconstruction. *Angew. Chem., Int. Ed.* **2024**, *63*, No. e202314869.
- (66) Nguyen, V.; Wilson, C.; Hoemberger, M.; Stiller, J. B.; Agafonov, R. V.; Kutter, S.; English, J.; Theobald, D. L.; Kern, D. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science*. **2017**, *355*, 289–294.
- (67) Wang, Z.-K.; Feng, D.-T.; Su, C.; Li, H.; Rao, Z.-M.; Rao, Y.-J.; Lu, Z.-M.; Shi, J.-S.; Xu, Z.-H.; Gong, J.-S. Designing ASSMD Strategy for Exploring and Engineering Extreme Thermophilic Ancestral Nitrilase for Nitriles Biocatalysis. *ACS Catal.* **2024**, *14*, 13825–13838.
- (68) Gumulya, Y.; Baek, J.-M.; Wun, S.-J.; Thomson, R. E. S.; Harris, K. L.; Hunter, D. J. B.; Behrendorff, J. B. Y. H.; Kulig, J.; Zheng, S.; Wu, X.; et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **2018**, *1*, 878–888.
- (69) Stepankova, V.; Bidmanova, S.; Koudelakova, T.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Strategies for Stabilization of Enzymes in Organic Solvents. *ACS Catal.* **2013**, *3*, 2823–2836.
- (70) Barruetabeña, N.; Alonso-Lerma, B.; Galera-Prat, A.; Joudé, N.; Barandiaran, L.; Aldazabal, L.; Arbulu, M.; Alcalde, M.; De Sancho, D.; Gavira, J. A.; et al. Resurrection of efficient Precambrian endoglucanases for lignocellulosic biomass hydrolysis. *Commun. Chem.* **2019**, *2*, 76.
- (71) Thomas, A.; Cutlan, R.; Finnigan, W.; van der Giezen, M.; Harmer, N. Highly thermostable carboxylic acid reductases generated by ancestral sequence reconstruction. *Commun. Biol.* **2019**, *2*, 429.
- (72) Chisholm, L. O.; Orlandi, K. N.; Phillips, S. R.; Shavlik, M. J.; Harms, M. J. Ancestral Reconstruction and the Evolution of Protein Energy Landscapes. *Annu. Rev. Biophys.* **2024**, *53*, 127–146.
- (73) Scossa, F.; Fernie, A. R. Ancestral sequence reconstruction - An underused approach to understand the evolution of gene function in plants? *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1579–1594.

- (74) Bart, A. G.; Harris, K. L.; Gillam, E. M. J.; Scott, E. E. Structure of an ancestral mammalian family 1B1 cytochrome P450 with increased thermostability. *J. Biol. Chem.* **2020**, *295*, 5640–5653.
- (75) Nakano, S.; Kozuka, K.; Minamino, Y.; Karasuda, H.; Hasebe, F.; Ito, S. Ancestral L-amino acid oxidases for deracemization and stereoinversion of amino acids. *Commun. Chem.* **2020**, *3*, 181.
- (76) Hendrikse, N. M.; Holmberg Larsson, A.; Svensson Gelius, S.; Kuprin, S.; Nordling, E.; Syrén, P.-O. Exploring the therapeutic potential of modern and ancestral phenylalanine/tyrosine ammonia-lyases as supplementary treatment of hereditary tyrosinemia. *Sci. Rep.* **2020**, *10*, 1315.
- (77) Chen, X.; Dou, Z.; Luo, T.; Sun, Z.; Ma, H.; Xu, G.; Ni, Y. Directed reconstruction of a novel ancestral alcohol dehydrogenase featuring shifted pH-profile, enhanced thermostability and expanded substrate spectrum. *Bioresour. Technol.* **2022**, *363*, 127886.
- (78) Myrtollari, K.; Calderini, E.; Kracher, D.; Schöngafner, T.; Galušić, S.; Slavica, A.; Taden, A.; Mokos, D.; Schrüfer, A.; Wirnsberger, G.; et al. Stability Increase of Phenolic Acid Decarboxylase by a Combination of Protein and Solvent Engineering Unlocks Applications at Elevated Temperatures. *ACS Sustain. Chem. Eng.* **2024**, *12*, 3575–3584.
- (79) Joho, Y.; Vongsouthi, V.; Spence, M. A.; Ton, J.; Gomez, C.; Tan, L. L.; Kaczmarek, J. A.; Caputo, A. T.; Royan, S.; Jackson, C. J.; et al. Ancestral Sequence Reconstruction Identifies Structural Changes Underlying the Evolution of Ideonella sakaiensis PETase and Variants with Improved Stability and Activity. *Biochemistry.* **2023**, *62*, 437–450.
- (80) Koblan, L. W.; Doman, J. L.; Wilson, C.; Levy, J. M.; Tay, T.; Newby, G. A.; Maianti, J. P.; Raguram, A.; Liu, D. R. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **2018**, *36*, 843–846.
- (81) Hendrikse, N. M.; Charpentier, G.; Nordling, E.; Syrén, P.-O. Ancestral diterpene cyclases show increased thermostability and substrate acceptance. *FEBS J.* **2018**, *285*, 4660–4673.
- (82) Schriever, K.; Saenz-Mendez, P.; Rudraraju, R. S.; Hendrikse, N. M.; Hudson, E. P.; Biundo, A.; Schnell, R.; Syrén, P.-O. Engineering of Ancestors as a Tool to Elucidate Structure, Mechanism, and Specificity of Extant Terpene Cyclase. *J. Am. Chem. Soc.* **2021**, *143*, 3794–3807.
- (83) Trisrivirat, D.; Sutthaphirom, C.; Pimviriyakul, P.; Chaiyen, P. Dual Activities of Oxidation and Oxidative Decarboxylation by Flavoenzymes. *ChemBioChem.* **2022**, *23*, No. e202100666.
- (84) Leese, C.; Fotheringham, I.; Escalettes, F.; Speight, R.; Grogan, G. Cloning, expression, characterisation and mutational analysis of L-aspartate oxidase from *Pseudomonas putida*. *J. Mol. Catal. B: Enzym.* **2013**, *85–86*, 17–22.
- (85) Trisrivirat, D.; Lawan, N.; Chenprakhon, P.; Matsui, D.; Asano, Y.; Chaiyen, P. Mechanistic insights into the dual activities of the single active site of L-lysine oxidase/monooxygenase from *Pseudomonas* sp. AIU 813. *J. Biol. Chem.* **2020**, *295*, 11246–11261.
- (86) Nakano, S.; Minamino, Y.; Hasebe, F.; Ito, S. Deracemization and Stereoinversion to Aromatic α -Amino Acid Derivatives with Ancestral L-Amino Acid Oxidase. *ACS Catal.* **2019**, *9*, 10152–10158.
- (87) Nakano, S.; Niwa, M.; Asano, Y.; Ito, S. Following the Evolutionary Track of a Highly Specific L-Arginine Oxidase by Reconstruction and Biochemical Analysis of Ancestral and Native Enzymes. *Appl. Environ. Microbiol.* **2019**, *85*, No. e00459-19.
- (88) Ishida, C.; Miyata, R.; Hasebe, F.; Miyata, A.; Kumazawa, S.; Ito, S.; Nakano, S. Reconstruction of Hyper-Thermostable Ancestral L-Amino Acid Oxidase to Perform Deracemization to D-Amino Acids. *ChemCatChem.* **2021**, *13*, 5228–5235.
- (89) Tomoiagă, R. B.; Ursu, M.; Boros, K.; Nagy, L. C.; Bencze, L. C. Ancestral L-amino acid oxidase: From substrate scope exploration to phenylalanine ammonia-lyase assay. *J. Biotechnol.* **2023**, *377*, 43–52.
- (90) Jones, B. J.; Evans, R. L., 3rd; Mylrea, N. J.; Chaudhury, D.; Luo, C.; Guan, B.; Pierce, C. T.; Gordon, W. R.; Wilmot, C. M.; Kazlauskas, R. J. Larger active site in an ancestral hydroxynitrile lyase increases catalytically promiscuous esterase activity. *PLoS One.* **2020**, *15*, No. e0235341.
- (91) Chaloupkova, R.; Liskova, V.; Toul, M.; Markova, K.; Sebestova, E.; Hernychova, L.; Marek, M.; Pinto, G. P.; Pluskal, D.; Waterman, J.; et al. Light-Emitting Dehalogenases: Reconstruction of Multifunctional Biocatalysts. *ACS Catal.* **2019**, *9*, 4810–4823.
- (92) Jaroensuk, J.; Chuaboon, L.; Chaiyen, P. Biochemical reactions in vitro ATP production and their applications. *Mol. Catal.* **2023**, *537*, 112937.
- (93) Bar-On, Y. M.; Milo, R. The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 4738–4743.
- (94) Erb, T. J.; Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO₂ fixing enzyme. *Curr. Opin. Biotechnol.* **2018**, *49*, 100–107.
- (95) Pang, J.-J.; Shin, J.-S.; Li, S.-Y. The Catalytic Role of RuBisCO for in situ CO₂ Recycling in *Escherichia coli*. *Front. Bioeng. Biotechnol.* **2020**, *8*, 543807.
- (96) Poudel, S.; Pike, D. H.; Raanan, H.; Mancini, J. A.; Nanda, V.; Rickaby, R. E. M.; Falkowski, P. G. Biophysical analysis of the structural evolution of substrate specificity in RuBisCO. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 30451–30457.
- (97) Karkehabadi, S.; Peddi, S. R.; Anwaruzzaman, M.; Taylor, T. C.; Cederlund, A.; Genkov, T.; Andersson, L.; Spreitzer, R. J. Chimeric Small Subunits Influence Catalysis without Causing Global Conformational Changes in the Crystal Structure of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase. *Biochemistry.* **2005**, *44*, 9851–9861.
- (98) Schulz, L.; Guo, Z.; Zarzycki, J.; Steinchen, W.; Schuller, J. M.; Heimerl, T.; Prinz, S.; Mueller-Cajar, O.; Erb, T. J.; Hochberg, G. K. A. Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science.* **2022**, *378*, 155–160.
- (99) Kuchimaru, T.; Iwano, S.; Kiyama, M.; Mitsumata, S.; Kadonosono, T.; Niwa, H.; Maki, S.; Kizaka-Kondoh, S. A luciferin analogue generating near-infrared bioluminescence achieves highly sensitive deep-tissue imaging. *Nat. Commun.* **2016**, *7*, 11856.
- (100) Wathaisong, P.; Kamutira, P.; Kesornpun, C.; Pongsupasa, V.; Phonbuppha, J.; Tinikul, R.; Maenpuen, S.; Wongnate, T.; Nishihara, R.; Ohmiya, Y.; et al. Luciferin Synthesis and Pesticide Detection by Luminescence Enzymatic Cascades. *Angew. Chem., Int. Ed.* **2022**, *61*, No. e202116908.
- (101) Seliger, H. H.; Lall, A. B.; Lloyd, J. E.; Biggley, W. H. THE COLORS OF FIREFLY BIOLUMINESCENCE—I. OPTIMIZATION MODEL. *Photochem. Photobiol.* **1982**, *36*, 673–680.
- (102) Bessho-Uehara, M.; Konishi, K.; Oba, Y. Biochemical characteristics and gene expression profiles of two paralogous luciferases from the Japanese firefly *Pyrocoelia atripennis* (Coleoptera, Lampyridae, Lampyrinae): insight into the evolution of firefly luciferase genes. *Photochem. Photobiol. Sci.* **2017**, *16*, 1301–1310.
- (103) Fallon, T. R.; Lower, S. E.; Chang, C.-H.; Bessho-Uehara, M.; Martin, G. J.; Bewick, A. J.; Behringer, M.; Debat, H. J.; Wong, I.; Day, J. C.; et al. Firefly genomes illuminate parallel origins of bioluminescence in beetles. *eLife* **2018**, *7*, No. e36495.
- (104) Oba, Y.; Konishi, K.; Yano, D.; Shibata, H.; Kato, D.; Shirai, T. Resurrecting the ancient glow of the fireflies. *Sci. Adv.* **2020**, *6*, No. eabc5705.
- (105) Huijbers, M. M. E.; Montersino, S.; Westphal, A. H.; Tischler, D.; van Berkel, W. J. H. Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **2014**, *544*, 2–17.
- (106) Phintha, A.; Chaiyen, P. Unifying and versatile features of flavin-dependent monooxygenases: Diverse catalysis by a common C4a-(hydro)peroxyflavin. *J. Biol. Chem.* **2023**, *299*, 105413.
- (107) Chiang, C.-H.; Wymore, T.; Rodríguez Benítez, A.; Hussain, A.; Smith, J. L.; Brooks, C. L.; Narayan, A. R. H. Deciphering the evolution of flavin-dependent monooxygenase stereoselectivity using ancestral sequence reconstruction. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120*, No. e2218248120.
- (108) Eggerichs, D.; Weindorf, N.; Mascotti, M. L.; Welzel, N.; Fraaije, M. W.; Tischler, D. Vanillyl alcohol oxidase from *Diplodia*

corticola: Residues Ala420 and Glu466 allow for efficient catalysis of syringyl derivatives. *J. Biol. Chem.* **2023**, *299*, 104898.

(109) Sugiura, S.; Nakano, S.; Niwa, M.; Hasebe, F.; Matsui, D.; Ito, S. Catalytic mechanism of ancestral L-lysine oxidase assigned by sequence data mining. *J. Biol. Chem.* **2021**, *297*, 101043.

(110) Bailleul, G.; Yang, G.; Nicoll, C. R.; Mattevi, A.; Fraaije, M. W.; Mascotti, M. L. Evolution of enzyme functionality in the flavin-containing monooxygenases. *Nat. Commun.* **2023**, *14*, 1042.

(111) Schenkmyerova, A.; Toul, M.; Pluskal, D.; Baatallah, R.; Gagnot, G.; Pinto, G. P.; Santana, V. T.; Stuchla, M.; Neugebauer, P.; Chaiyen, P.; et al. Catalytic mechanism for Renilla-type luciferases. *Nat. Catal.* **2023**, *6*, 23–38.

(112) Gao, T.; Damborsky, J.; Janin, Y. L.; Marek, M. Deciphering Enzyme Mechanisms with Engineered Ancestors and Substrate Analogues. *ChemCatChem.* **2023**, *15*, No. e202300745.

(113) Schenkmyerova, A.; Pinto, G. P.; Toul, M.; Marek, M.; Hernychova, L.; Planas-Iglesias, J.; Daniel Liskova, V.; Pluskal, D.; Vasina, M.; Emond, S.; et al. Engineering the protein dynamics of an ancestral luciferase. *Nat. Commun.* **2021**, *12*, 3616.

(114) Wilding, M.; Peat, T. S.; Kalyaanamoorthy, S.; Newman, J.; Scott, C.; Jermini, L. S. Reverse engineering: transaminase biocatalyst development using ancestral sequence reconstruction. *Green Chem.* **2017**, *19*, 5375–5380.

(115) Kawamura, Y.; Ishida, C.; Miyata, R.; Miyata, A.; Hayashi, S.; Fujinami, D.; Ito, S.; Nakano, S. Structural and functional analysis of hyper-thermostable ancestral L-amino acid oxidase that can convert Trp derivatives to D-forms by chemoenzymatic reaction. *Commun. Chem.* **2023**, *6*, 200.

(116) Galanie, S.; Entwistle, D.; Lalonde, J. Engineering biosynthetic enzymes for industrial natural product synthesis. *Nat. Prod. Rep.* **2020**, *37*, 1122–1143.

(117) O'Connell, A.; Barry, A.; Burke, A. J.; Hutton, A. E.; Bell, E. L.; Green, A. P.; O'Reilly, E. Biocatalysis: landmark discoveries and applications in chemical synthesis. *Chem. Soc. Rev.* **2024**, *53*, 2828–2850.

(118) Goudey, B.; Geard, N.; Verspoor, K.; Zobel, J. Propagation, detection and correction of errors using the sequence database network. *Brief. Bioinform.* **2022**, *23*, bbac416.

(119) Williams, P. D.; Pollock, D. D.; Blackburne, B. P.; Goldstein, R. A. Assessing the Accuracy of Ancestral Protein Reconstruction Methods. *PLoS Comput. Biol.* **2006**, *2*, No. e69.

(120) Abadi, S.; Azouri, D.; Pupko, T.; Mayrose, I. Model selection may not be a mandatory step for phylogeny reconstruction. *Nat. Commun.* **2019**, *10*, 934.

(121) Hanson-Smith, V.; Kolaczowski, B.; Thornton, J. W. Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty. *Mol. Biol. Evol.* **2010**, *27*, 1988–1999.

(122) Vialle, R. A.; Tamuri, A. U.; Goldman, N. Alignment Modulates Ancestral Sequence Reconstruction Accuracy. *Mol. Biol. Evol.* **2018**, *35*, 1783–1797.

(123) Aadland, K.; Kolaczowski, B. Alignment-Integrated Reconstruction of Ancestral Sequences Improves Accuracy. *Genome Biol. Evol.* **2020**, *12*, 1549–1565.

(124) Foley, G.; Mora, A.; Ross, C. M.; Bottoms, S.; Sützl, L.; Lamprecht, M. L.; Zaugg, J.; Essebier, A.; Balderson, B.; Newell, R.; et al. Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). *PLoS Comput. Biol.* **2022**, *18*, No. e1010633.

(125) Liu, Y.; Yang, Q.; Zhao, F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annu. Rev. Biochem.* **2021**, *90*, 375–401.

(126) Rothman, J. E.; Schekman, R. Molecular Mechanism of Protein Folding in the Cell. *Cell.* **2011**, *146*, 851–854.

(127) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5869–5874.

(128) Listov, D.; Goverde, C. A.; Correia, B. E.; Fleishman, S. J. Opportunities and challenges in design and optimization of protein function. *Nat. Rev. Mol. Cell Biol.* **2024**, *25*, 639–653.

(129) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* **2021**, *596*, 583–589.

(130) Modarres, H. P.; Mofrad, M. R.; Sanati-Nezhad, A. Protein thermostability engineering. *RSC Adv.* **2016**, *6*, 115252–115270.

(131) Planas-Iglesias, J.; Marques, S. M.; Pinto, G. P.; Musil, M.; Stourac, J.; Damborsky, J.; Bednar, D. Computational design of enzymes for biotechnological applications. *Biotechnol. Adv.* **2021**, *47*, 107696.

(132) Yu, T.; Cui, H.; Li, J. C.; Luo, Y.; Jiang, G.; Zhao, H. Enzyme function prediction using contrastive learning. *Science.* **2023**, *379*, 1358–1363.