

Full-Atom MPNN Based Redesign of Plant Dehydrogenase Enables Thermostability Enhancement Without Loss of Stereoselectivity

Bruno Di Geronimo,^{1#} Jasmin Zuson,^{2#} Ana Udženija,² Andrea M. Chánique,^{2,‡} Robert Kourist^{2,3*}
and Shina Caroline Lynn Kamerlin^{1,4,5*}

1. School of Chemistry and Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive NW, Atlanta, GA 30332-0400, USA
2. Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, Graz 8010, Austria
3. BioTechMed-Graz, Mozartgasse 12, 8010 Graz, Austria
4. School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Drive, Atlanta GA 30332-0400, USA
5. Department of Chemistry, Lund University, Box 124, 22100 Lund, Sweden

These authors contributed equally.

‡ Author current address: Kura Biotech, Av. Gramado 1410 interior, Parcela 20, Puerto Varas 5550000, Chile.

Corresponding author email addresses: kourist@tugraz.at, skamerlin3@gatech.edu

Summary

Protein stabilization is a “Holy Grail” of biocatalysis, and stability design is an area of intense research interest. While it is increasingly feasible to effectively increase enzyme thermostability, optimization without compromising activity or selectivity remains a significant challenge. Here, we use full-atom protein sequence design with sidechain conditioning (FAMPNN) to engineer thermostable variants of the borneol dehydrogenase from *Salvia rosmarinus* (SrBDH1), an enzyme from a family where unselective enzymes dominate, and selectivity is determined by dynamical considerations. By combining FAMPNN design with residue conservation analysis and avoiding active site residues, we were able to computationally design SrBDH1 variants with up to 10 °C enhanced thermostability and strongly increased half-life time at elevated temperature, while retaining selectivity towards (+)-borneol. This design framework, integrating *de novo* and physics-based protein design tools, demonstrates that stability can be enhanced without disrupting functionally relevant dynamics, providing a route to engineer robust and selective biocatalysts.

Keywords: protein stability | stereoselectivity | ProteinMPNN | borneol dehydrogenase | enzyme design | enzyme engineering

Introduction

Protein stability engineering is considered a “Holy Grail” of biocatalysis, as it is essential for allowing enzymes to operate under harsh industrial conditions, such as high temperature and organic solvents.^{1,2} While an abundance of experimental and computational approaches exist to engineer stable proteins more broadly,¹⁻³ a main challenge to be overcome is the function-stability tradeoff.³⁻¹⁵ Specifically, most natural proteins are only marginally stable,^{3,16-18} and mutations that improve protein function (for instance catalytic activity or ligand binding) often do so at the cost of protein stability. Conversely, attempts to improve stability can compromise catalytic function or enzyme selectivity.³⁻¹⁵

A number of strategies have been developed to address this issue. Among them, directed evolution is a powerful tool to find “compensatory” mutations – changes elsewhere in the protein scaffold that fix folding issues created by new functional mutations.¹⁹⁻²⁵ Studies utilizing classic error-prone PCR (epPCR) report increased thermal stability after screening a minimum of 19,000 clones in one round of directed evolution.²⁶ More complex combined approaches, aim to minimize the screening effort of the generated libraries by design, but still screen 7 to 8 thousand clones per round.²⁷ However, achieving this is often only possible over multiple rounds of laboratory evolution, which carries both monetary and human time costs, and is likely to hit optimization plateaus.²⁸⁻³² Another alternative is the use of ancestral sequence reconstruction (ASR),³³⁻³⁵ which tends to predict proteins with substantively higher thermostability,³⁶⁻⁴⁰ although the origins of this (reproduction of early earth conditions vs. algorithmic bias) remain controversial.^{36,41,42} However, while ASR is an effective tool to quickly obtain enhanced stability, the predicted variants can have solubility and expression challenges, or other issues such as for instance lower capacity to bind an internal cofactor.

More recently, there has been substantive interest in using pure phylogenetic,^{39,43,44} hybrid phylogenetic/physics-based^{44,45} and machine learning-based approaches such as RFDiffusion,^{46,47} ProteinMPNN,⁴⁸ ThermoMPNN,⁴⁹ Evo-Velocity,⁵⁰ and ESM-2/ESM-Fold,⁵¹ among other examples, to predict stable and functional protein scaffolds *de novo*. Recently, we used ASR to generate the ancestors of a heterogeneous family of plant borneol dehydrogenases (BDHs) that are associated to the oxidation of (+)-1-borneol to camphor.⁵² While most of these short-chain reductases accept both enantiomers of 1-borneol,⁵³⁻⁵⁵ the enzymes from rosemary and sage have outstanding enantioselectivity in the kinetic resolution of this diborane type monoterpenol.⁵⁶ Not surprisingly, given function-stability tradeoffs,³⁻¹⁵ most ASR characterized ancestors were no longer enantioselective, with the exception of the direct ancestors of the enzymes from rosemary and sage.⁵² Similar to previously characterized ASR-generated proteins,^{57,58} shifts in selectivity across the transition from predicted ancestral to extant enzymes could be associated to perturbations in conformational dynamics.

Although ASR has been to some extent a “gold-standard” for easy acquisition of thermostability,³⁶⁻³⁹ it is clearly not without functional cost. Similarly to ASR, *de novo* designed proteins have also been exhibited to be highly thermostable.⁵⁹⁻⁶³ In the current work, we explore whether ML-based approaches can be used to obtain thermostability *without* compromising activity and stereoselectivity in *de novo* designed enzymes, using FAMPNN as a model approach, based on prior successes of this approach in enhancing protein thermostability.⁶⁴ In addition, the last two decades have shown that focus of saturation mutagenesis into the active site is a powerful method to alter the selectivity of an enzyme.⁶⁵⁻⁶⁹ In many cases, variation of first shell amino acids is sufficient for such an improvement, and it is not necessary to address the periphery.⁶⁶ In this work, we hypothesized that excluding active site amino acids from the redesign strategy will

protect the active site, and retain selectivity. To test this hypothesis, we generated a set of 8 BDH variants by redesign of peripheral regions of the BDH from *Salvia rosmarinus* (SrBDH1^{54,70}), while conserving residues that are supposed to participate in the catalytic machinery and the active-site hydrophobic pockets.

Using this approach, coupled with experimental validation of computational designs, we demonstrate that we were able to obtain a variant with a 10°C higher unfolding temperature, and with minimal trade-off in activity (~65% of the specific activity retained) and, crucially, while retaining selectivity, unlike in the ASR-based variants predicted in prior work.⁵² This is significant given that maintaining selectivity in SrBDH1 appears to be particularly challenging.⁵² Finally, computational characterization of both the computationally redesigned sequences and the ancestral proteins from our prior ASR study⁵² indicated that thermostability and stereoselectivity are governed by distinct and partially decoupled structural features, with selectivity primarily emerging from the dynamic stabilization of catalytically competent conformations rather than static active-site geometry. While BDH is itself an industrially important enzyme for the synthesis of borane-type monoterpenoids,^{52-54,71,72} our study also highlights the power of strategic exploitation of ML-based approaches for protein stability design without compromising enzyme selectivity, which has major applications for biocatalysis and provides a possible avenue also for the stabilization of engineered enzyme variants having artificially improved stereoselectivity.

Results and Discussion

Sequence Generation and Thermostability Predictions

ConSurf analysis^{73,74} was performed on the borneol dehydrogenase of *Salvia rosmarinus* (SrBDH1, PDB ID: 6ZYZ⁵⁴), providing us with residue conservation cutoffs,⁷⁵ normalized from 1 to 9, from more variable to highly conserved positions (**Figures 1A** and **S1**). We then applied two

different thresholds (residue conservation cutoffs of ≤ 4 and ≤ 5), and then performed full-atom protein sequence design with sidechain conditioning using FAMPNN (Full-Atom MPNN), an inverse folding model that co-designs sequence and sidechains.⁶⁴ Taking sidechains into account is important, as while other more specialized options such as ThermoMPNN⁴⁹ also exist, our goal is not just to enhance thermostability, but to do so without compromising selectivity, which in turn is determined by sidechain conformations and dynamics.⁷⁶⁻⁸⁰ A further key advantage of FAMPNN⁶⁴ is that it outputs *de novo* SrBDH1 designs as fully resolved all-atom Protein Data Bank (PDB)⁸¹ structures, ensuring intrinsic structural self-consistency, and eliminating the need for post-hoc side-chain repacking. This also makes it easy to evaluate and rank FAMPNN⁶⁴ output by feeding it into PyRosetta⁸² to compute $\Delta\Delta G$ values.

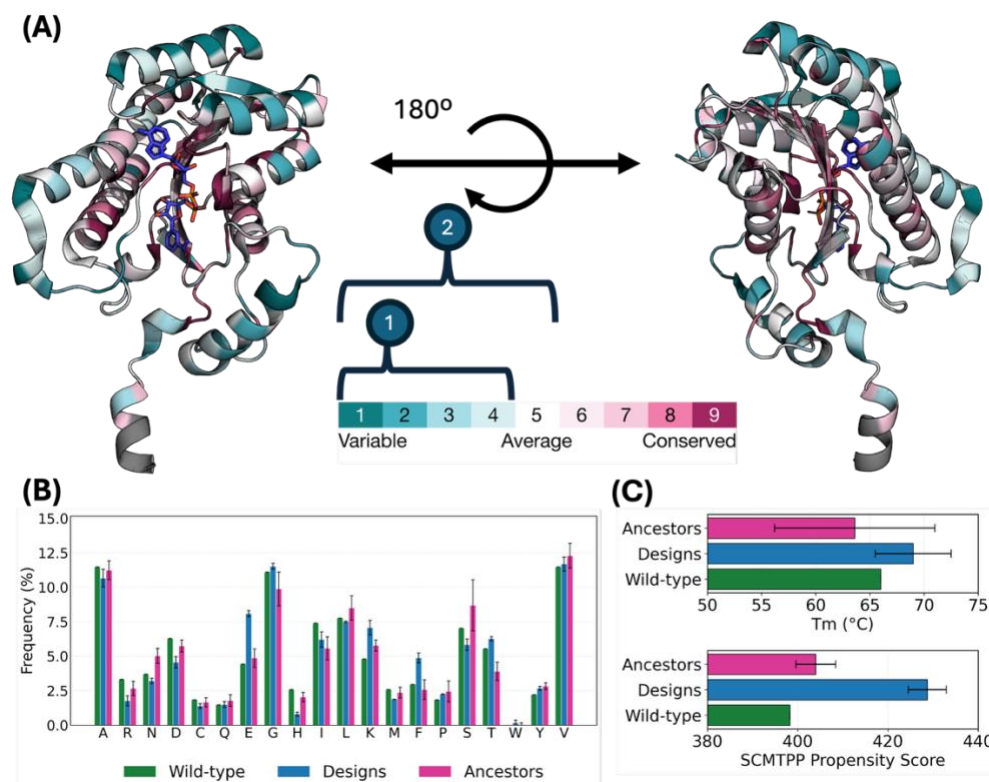


Figure 1. (A) Borneol dehydrogenase (BDH) residue conservation, performed using ConSurf,^{73,74} with the BDH from *Salvia rosmarinus* (SrBDH1)⁵⁴ used as a reference. Chain A from the SrBDH1 structure (PDB

ID: 6ZYZ⁵⁴) is shown in cartoon representation, and colored by a ConSurf^{73,74} conservation score (1–9; cyan = variable → magenta = conserved). NAD⁺ is shown as blue sticks. Two orientations of this enzyme (left/right) are shown to highlight a conserved cluster of residues around the active-site pocket, adjacent to the NAD⁺ cofactor. **(B)** Amino acid composition of wild-type *SrBDH1* (green), designed sequences (blue) and ancestral sequences (magenta), performed using the ExPASy webserver.⁸³ **(C)** Experimental thermostability mean values, together with SCMTTP⁸⁴ (“Scoring Card Method Thermophilic Protein Predictor”)-derived propensity score predictions and standard deviations, for the wild-type system (green), designed *SrBDH1* variants (blue), and ancestral sequences (magenta).

Prior work has indicated a number of sequence and structural factors being important indicators of thermostability. These include: (1) hydrophobic content of the sequence,⁸⁵⁻⁸⁸ which can be measured using the Kyte-Doolittle scale,⁸⁹ either as % hydrophobic content of the sequence, or through Grand Average of Hydropathicity scores (GRAVY index, where the GRAVY score is simply the sum hydropathy values of all amino acids divided by the total number of amino acids). (2) Related to this, statistical analysis suggests that globular proteins from thermophilic organisms have a significantly higher aliphatic index (a measure of the volume occupied by their aliphatic side chains) than their mesophilic counterparts.⁹⁰ (3) Analysis of sequence composition has suggested that thermostable proteins show an increased frequency of Arg and Tyr side chains, and a reduced frequency of Cys and Ser side chains, compared to their mesophilic counterparts⁹¹ (in the case of hyperthermostable archaeal protein tyrosine phosphatases, as also observed a higher frequency of Glu and Leu residues⁹²). (4) Finally, from a structural perspective, thermostable proteins have both a larger fraction of their residues in an α -helical conformation and avoid Pro to a larger extent in those α -helices,⁹¹ they have a higher amount of salt bridges and hydrogen bonds compared to mesophilic counterparts,⁹¹ and maintain a higher percentage of their residues engaged in hydrophobic contacts (in line with expected increased hydrophobic content).⁹²

For sequence-based putative thermostability analysis, we compared two sets of sequences to that of wild-type *SrBDH1*: (1) FAMPNN designs⁶⁴ selecting the top 4 sequences based on Rosetta Energy Unit (REU) values⁹³ from PyRosetta,⁸² with two design batches, *i.e.*, designs at residue positions with ConSurf^{73,74} residue conservation cutoffs of ≤ 4 and ≤ 5 , and (2) a series of ASR predictions from our prior work.⁵² The sequences being considered are provided in the **Supplementary Information**, and the corresponding analysis is provided in **Figure 1**, with the corresponding raw data provided in the **Supplementary Table**). Amino acid composition, instability indices, aliphatic indices, and GRAVY scores were all calculated using the ProtParam tool on the ExPASy webserver.⁸³ Protein thermostability was predicted from sequence based on propensity scores calculated using the SCMTTP webserver, a sequence-based thermophilic protein predictor (TPP) that uses a scoring card method (SCM).⁸⁴

Our sequence-based analysis indicates, most importantly, that both FAMPNN⁶⁴ and ASR sequences have increased SCMTTP propensity scores compared to wild-type *SrBDH1*, with the FAMPNN⁶⁴ designs having on average even higher propensity scores than the ASR sequences, suggestive of the highest thermostability (398.3 for wild-type *SrBDH1*, 404.0 ± 4.4 for the ASR sequences, and 428.7 ± 4.2 for the FAMPNN⁶⁴ designs). In addition, the predicted sequences all also show lower GRAVY scores than the experimental designs, slightly lower instability indices, and similar aliphatic indices. Finally, in terms of sequence composition, curiously, we see a reduced frequency of Arg side chains and a very modest increase in frequency of Tyr side chains in the FAMPNN⁶⁴ designs compared to wild-type, as well as a reduced frequency of Cys and Ser side chains compared to both wild-type and the ASR constructs, hinting at greater thermostability.⁹¹ Interestingly, we also see an increased frequency of Glu, and Phe residues in the designed sequences, compared to either wild-type or ASR sequences. Finally, despite their

highlighted importance in prior work,^{85-88,90} we do not see meaningful trends in either aliphatic indices or GRAVY scores, that would distinguish between wild-type *SrBDH1* and the FAMPNN⁶⁴ or ASR sequences.⁵²

Following from this, we structurally and dynamically profiled wild-type *SrBDH1* and the *de novo* designed *SrBDH1* constructs, as well as three ancestral sequences from our prior work⁵² for reference, based on molecular dynamics simulations performed as described in the **Materials and Methods**. The three selected ancestral sequences, N32, N5 and N6 have T_m values ranging from 61-77 °C, see **Section “Experimental Characterization and Validation”**). Structural stability was evaluated by monitoring non-covalent interactions including hydrogen bonds, salt bridges, and hydrophobic interactions (**Figure 2**). Additionally, solvent accessible surface area (SASA) and root mean square fluctuations (RMSF) of backbone C α -atoms were calculated (**Figures S2 to S4**, and **Table S1**).

Analysis of non-covalent interactions during our simulations (**Figure 2**) reveals clear differences in interaction propensity between wild-type *SrBDH1*, the new designed variants, and the ancestral *SrBDH1* variants from our prior work.⁵² The ancestral enzymes simulated display a higher number of hydrogen bonds compared to both the wild-type enzyme and designed variants, suggesting increased internal stabilization through polar interactions. In contrast, the designed variants exhibit an increased number of salt bridges, consistent with enhanced electrostatic stabilization, which has been suggested to be associated with increased protein thermostability.⁹⁴⁻
⁹⁷ Moreover, hydrophobic contacts are more prevalent in the designed systems compared to both wild-type and ancestral variants, reflecting a redistribution of stabilizing interactions within the protein scaffold, likely driven by the increased abundance of phenylalanine residues in the designs.

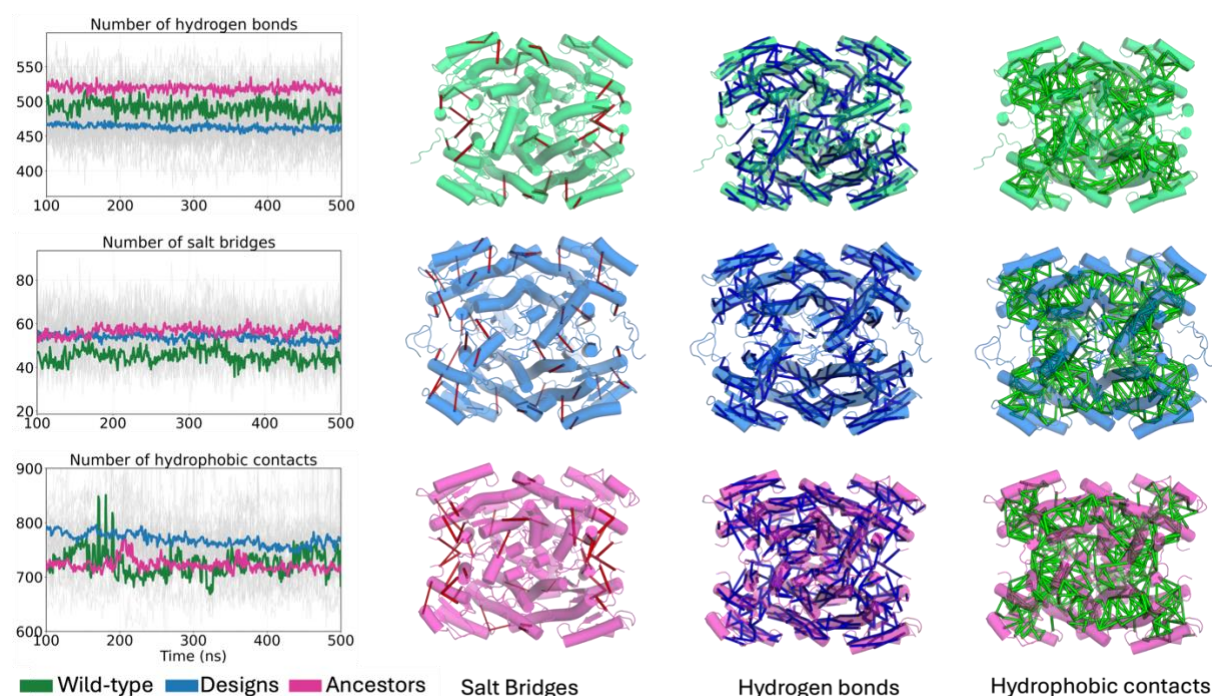


Figure 2. Computational analysis of non-covalent interaction networks associated with *SrBDH1* stability across wild-type, designed variants, and ancestral systems. Data are derived from the last 400 ns of 3 x 500 ns conventional MD simulations. Hydrogen bonds were identified using CPPTRAJ,⁹⁸ (and salt bridges and hydrophobic contacts were quantified using MDAnalysis,^{99,100} with mean values for each system shown in color. The cartoon representations highlight the distribution of hydrogen bonds, salt bridges, and hydrophobic interactions mapped onto representative protein conformations, illustrating differences in interaction networks between wild-type (green), designed variants (blue), and ancestral proteins (magenta).

Although these differences may appear moderate at the global level, they arise from specific residue substitutions that reorganize interaction networks across the protein scaffold. For example, comparison between the wild-type and design 01-080 (**Figure S5**) reveals the introduction of new salt bridges through the A27R, R31E, Q46R, and L49E as well as N60D, V135K, V135R, and S180K substitutions. In addition, hydrophobic interactions are enhanced by substitutions such as H85W, L122F, H158F and V176F (**Figures S5**), contributing to a more tightly packed and

stabilized protein scaffold. Monitoring of the solvent accessible surface area (SASA) values ($\text{\AA}^2$) across variants, including both total and hydrophobic contributions, shows a consistent decrease in surface exposure relative to the wild-type enzyme for both designs and ancestral proteins (**Figures S2** and **Table S1**). This, in turn, is likely to contribute to enhanced compactness and structural robustness of the designs, potentially reinforced by an increased abundance of charged residues and the formation of stabilizing salt-bridges at the protein surface. Indeed, analysis of non-covalent interactions (**Figure S2B**) confirms that most FAMPNN designs⁶⁴ show a reduction in total hydrogen bonds but an increase in salt bridge interactions over the course of the last 400 ns of simulation time (3 trajectories per system), with particularly pronounced increases observed for designs 02-037, 02-389, and 02-911, which reach mean values of approximately 62 salt bridges per system. In contrast, designs 01-037 and 01-173 have non-covalent interaction totals that are comparable to wild-type *SrBDH1*. Further, while the total number of hydrogen bonds is generally reduced in the designs relative to the wild type, variants 01-080, 01-099, and 02-037 retain relatively high hydrogen bond counts throughout the simulations.

Following from this, we also dynamically profiled wild-type *SrBDH1* designed variants, and ancestors considering C_{α} -atom RMSF across our MD simulations (**Figure 3**). We note that the relationship between protein thermostability and flexibility is complex,¹⁰¹ as are the structural features that correlated with protein thermal stability.^{84,91} The general view of protein flexibility is that protein thermostability is linked to increased scaffold rigidity.¹⁰¹⁻¹⁰⁷ However, several studies have demonstrated that this is not necessarily the case, and that flexibility and thermal stability can be decoupled.^{101,108-111} In fact, in prior work it has been shown that significantly increased thermostability of ancestral β -lactamases identified through ASR is coupled with increased conformational flexibility,^{58,112} which in turn controls the evolution of substrate specificity.^{58,113}

This said, comparative RMSF analysis across all four protein chains reveals a pronounced increase in flexibility in the region spanning residues 190-225 in all designs (**Figures 3, S3 and S4**).

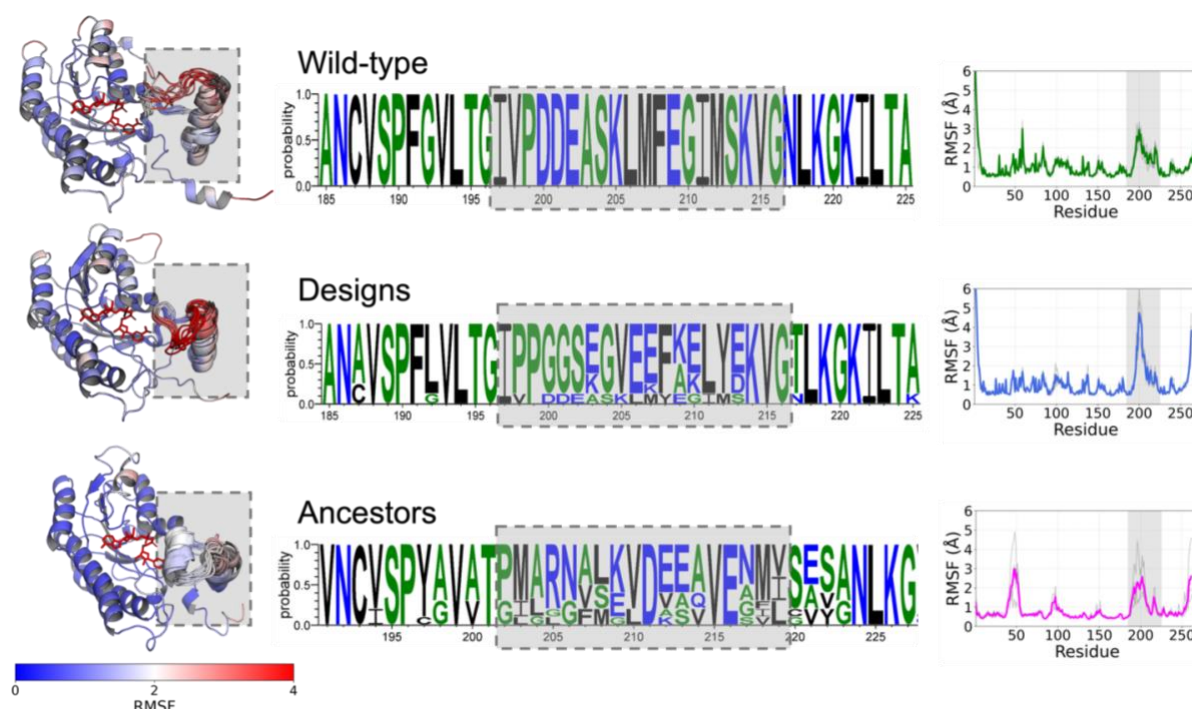


Figure 3. Cartoon representations show ensembles of sampled conformations for residues 185–225, extracted at 50 ns intervals from the last 400 ns of 3 x 500 ns MD simulations of each system, highlighting increased mobility in the wild-type, design 01-080, and ancestor N5. RMSF values are mapped onto the structures using a color scale from blue (low flexibility) to red (high flexibility), corresponding to values between 0 and 4 Å. Detailed RMSF profiles are also shown for wild-type (green), designed variants (blue), and ancestral proteins (magenta). Shown here is also the sequence variability within this mobile region (residues 185-225), calculated using multiple sequence alignment generated from MUSCLE¹¹⁴ for the ancestral sequences and illustrated as a WebLogo¹¹⁵ comparing wild-type *SrBDH1*, the FAMPNN⁶⁴ designs, and the ancestors. C_{α} -atom root-mean-square fluctuation (RMSF, Å) analysis was performed to compare the scaffold flexibility of wild-type *SrBDH1*, the FAMPNN⁶⁴ designs, and the ancestral variants from our prior work. The sequence diversification in the highlighted region correlates with altered local flexibility, where designed variants show increased mobility compared to the wild-type and ancestral systems, that may underlie their reduced catalytic activity.

Sequence conservation analysis (**Figure 3**) indicates that residues 200–210 represent one of the most variable segments among the FAMPNN⁶⁴ designs, and that in particular, there is enrichment in glycine residues (which are known to promote backbone flexibility) at positions 200, 201, and 204 providing a higher RMSF in this region (**Figure S3** and **S4**). In contrast, this enrichment is not present at all on the ancestors, and some of them present lower RMSF than the wild-type. This localized increase in conformational dynamics may compromise the precise positioning of structural elements required for efficient catalysis and could therefore impact the enzymatic activity of these variables. The agnostic identification of this loop from our simulations is particularly interesting, given that this loop/helix region has been suggested to play an important role in substrates recognition in other short-chain dehydrogenases (see *e.g.*, refs. ¹¹⁶⁻¹¹⁹). Further, as a point of comparison, we note that in prior work engineering *de novo* Kemp eliminase activity into ASR predicted ancestral β -lactamases, we analogously observed displacement of a key helix as being essential for introduction of the *de novo* activity.¹¹² Further, in these systems, both the *de novo* activity and the significantly increased thermostability of ASR predictions was associated with increased helix and overall scaffold flexibility.¹¹²

Experimental Characterization and Validation

Eight redesigns of *SrBDH1* were chosen for biochemical characterization based on their predicted change in Gibbs free energy of folding ($\Delta\Delta G$, **Table S1**). The subset was derived from two design strategies based on residue conservation cutoffs; Strategy 01 (cutoff <4) and Strategy 02 (cutoff <5). The less conservative approach used in Strategy 01 includes redesigns 01-037, 01-080, 01-099 and 01-173, while Strategy 02 includes redesigns 02-037, 02-389, 02-911 and 02-939. With the exception of design 01-173 (**Figure S6**), all redesigns exhibited soluble expression in *Escherichia coli* BL21 (DE3) and were purified (**Figure S7**). In previous work, most BDH from

different plants and their reconstructed ancestors could be produced in soluble in *E. coli*. However, a few were found exclusively in the insoluble fraction.^{52,55} While the two redesigns 01-037 and 02-939 show 38% and 64% of the protein yield observed in *SrBDH1* of 98 mg/L, all other designs display expression yields between 85-90% of *SrBDH1* (**Table S2**). This compares favorably to the protein yield of the thermostabilized ancestor N5 of 18%.⁵²

The catalytic performance of the redesigned variants was evaluated in the oxidation of the pure enantiomers (+)- and (-)-borneol, and racemic isoborneol. We observe that Strategy 01 redesigns consistently have about 10-fold higher activity than Strategy 02 redesigns towards all substrates investigated (**Table 1**). Further, Strategy 02 redesigns generally display an up to 10-fold reduction in oxidative rate towards (+)-borneol compared to Strategy 01 redesigns. This suggests that the more conservative redesign approach destabilized the overall catalytic function, leading to a loss in activity. Meanwhile, redesigns of the less conservative Strategy 01 retained 40-65% of the wild-type activity towards (+)-borneol. Notably, and similarly to previously reported ancestors of BDHs,⁵² all redesigned variants show stereopreference for the oxidation of (+)-isoborneol.

The investigation of the enantioselectivity (E-value) in the kinetic resolution of *rac*-borneol confirmed that all redesigns retained exceptional enantioselectivity ($E > 200$). This is a difference to the reconstructed ancestors, where the thermostabilized ancestor N5 has one of the highest (-)-borneol conversion rates, whereas other ancestors with conserved selectivity did not have higher stability than the extant enzyme. The redesign 02-911 had slightly lower selectivity ($E = 189$), which is still sufficiently high for synthetic application. While all redesigns retained the very high E-value, a moderate variation of the relative activities between borneol and isoborneol was observed. 01-080 displays the highest activity toward (+)-borneol, while showing a significant

reduction in activity toward *rac*-isoborneol, whereas 01-099 has even lower activity towards (+)-borneol, but converts *rac*-isoborneol faster than 01-080.

Table 1. Oxidative rates for (+)-borneol, (-)-borneol, and *rac*-isoborneol given in [mU/mg] catalyzed by the purified redesigns.^a

Enzyme	(+)-borneol (mU mg ⁻¹)	(-)-borneol (mU mg ⁻¹)	<i>rac</i> -isoborneol (mU mg ⁻¹)	E-value
01-037	116.9 ± 13.0	3.6 ± 0.4	44.5 ± 5.3	>200
01-080	195.1 ± 23.0	2.3 ± 0.3	20.3 ± 3.4	>200
01-099	189.4 ± 3.8	1.5 ± 0.1	68.8 ± 7.2	>200
02-037	11.6 ± 0.7	0.1 ± 0.1	3.9 ± 2.3	>200
02-389	21.7 ± 2.6	0.5 ± 0.1	9.8 ± 2.7	>200
02-911	10.9 ± 1.3	0.2 ± 0.0	5.2 ± 0.9	189
02-939	22.3 ± 0.8	0.6 ± 0.0	6.5 ± 1.1	>200
<i>SrBDH1</i> ^a	296.9 ± 5.3	1.3 ± 0.1	68.3 ± 16.0	>200

^a E-values of *rac*-borneol are calculated as enantiomeric excess after 24h. Reaction conditions: 2 μM enzyme, 20% DMSO, 25°C, 300 rpm, 9 mM NAD⁺, 9 mM of substrate, 0.1 M Tris-HCl pH 9.0. Measurements were carried out in triplicate and quantified *via* GC analysis. ^b Data from Ref. ⁵².

In the determination of unfolding temperatures (T_m) (**Table 2**), nearly all redesigns demonstrated enhanced thermostability compared to *SrBDH1*. Strategy 01 redesigns generally show a trend towards increased thermal profiles, specifically 01-080 showed the highest stabilization with a T_m of +10°C in comparison to *SrBDH1*. Only one Strategy 02 redesign, 02-911, shows a decrease in T_m of -1°C, retaining a similar stability as the wild-type. The investigated ancestors⁵² either show lower T_m in the case of N7 to N74 (55-61°C), similar T_m in N6 (66°C), or a 5-11°C increase in T_m compared to *SrBDH1* in the case of N4 and N5.

Table 2. Unfolding temperatures (T_m) of the purified redesigned variants, wild-type *SrBDH1* and ancestral BDHs, as well as corresponding thermal stability predictions.^a

Enzyme	Unfolding Temperature [°C]	SCMTPP Propensity Score	ThermoProt HMM Score	TmProt Score
<i>SrBDH1</i>	66	398.26	0.0779	48.1
<i>Designed Variants</i>				
01-037	67	429.72	0.137	47.7
01-080	76	435.15	0.121	47.4
01-099	70	430.91	0.091	49.4
02-037	69	429.66	0.136	50.8
02-389	68	422.97	0.141	50.8
02-911	65	428.54	0.163	49.9
02-939	68	423.69	0.076	49.6
<i>Ancestral Variants</i>				
N4	71	404.64	0.051	47.3
N5	77	409.88	0.068	47.3
N6	66	406.86	0.052	46.8
N7	55	394.61	0.035	47.4
N30	56	404.48	0.040	48.9
N32	61	405.07	0.126	47.3
N39	62	403.30	0.103	46.6
N74	61	403.55	0.116	47.1

^a Unfolding temperatures for the designed variants were measured *via* differential scanning fluorimetry. Unfolding temperatures for the ancestral variants from ref. ⁵² were measured *via* CD spectroscopy. Shown here for comparison are also propensity scores calculated using the SCMTPP webserver,⁸⁴ ThermoProt Hidden Markov Model (HMM) scores,⁵² and predicted melting temperatures, T_m , using TmProt.¹²⁰

Comparison of the designed variants to the ancestral and wild-type variants (**Table 2**) indicates that the redesigned variants show similar T_m distributions to the thermostable ancestral variants. Within the subset of ancestral proteins, N5 is the most stable ancestor with a T_m of 77°C. However, N5 displays an E-value of 13 for *rac*-borneol, while the redesign 01-080 retains an E-value of >200. Circular dichroism (CD) spectroscopy was used to cross-validate the DSF data collected for the redesign 01-080 and the wild-type *SrBDH1*. CD analysis estimated higher denaturation points for both *SrBDH1* (71°C) and 01-080 (81°C), while retaining the same difference ΔT_m of +10°C in

design 01-080. Additionally, cooling of the denatured sample back to room temperature, following the initial temperature scan revealed that 01-080 has the capacity for refolding in contrast to *SrBDH1* (**Figure S8**).

The high catalytic activity, thermal stability, and outstanding stereoselectivity of 01-080 thus position it as the best performing variant of either the ancestral or redesigned variants. As such further characterization of the variant was conducted. The apparent Michaelis-Menten constant (K_{app}) was determined. The redesign displayed a K_{app} of 13.7 ± 5.3 mM for (+)-borneol, representing a 6.8-fold increase compared to the wild-type K_M of 2.02 ± 0.18 mM as reported by Chánique *et. al.*, in 2021 (**Figure S9**).⁵⁴ Due to the solubility limits of the substrate borneol (4mM in water, increased to 10 mM in aqueous solution with 20% DMSO addition) and thus missing saturation of the enzyme, only the apparent Michaelis-Menten constant could be determined. As wild-type SDRs are known to exist as either di- or tetramers,¹²¹ size-exclusion chromatography (SEC) was performed to validate the quaternary structure of 01-080 in comparison to *SrBDH1*. The analysis confirmed that the redesigned 01-080 maintains the native tetrameric structure of *SrBDH1* without forming higher-order aggregates (**Figure S10**).

Following from this, computational prediction of thermal stability from sequence remains a challenging problem. To this end, we have compared a range of currently available tools for thermostability prediction from sequence, to assess whether they are able to correctly reproduce experimental trends. Impressively, the scoring-card method SCMTTP propensity scores used in the design process (**Figure 1** and **Table 1**) give a significant large positive relationship between experimental and predicted values (Spearman $\rho = 0.6514$, two-tailed P -value = 0.0063, see **Supplementary Table**), with propensity scores of 398.26 for the wild-type *SrBDH1*, 428.7 ± 4.2 for the designed variants, and 404.0 ± 4.4 for the ancestral variants. This qualitatively follows

experimental data. In contrast, when comparing to more recent machine learning based approaches (e.g., the Hidden Markov Model (HMM) scores from ThermoProt⁵² or the unfolding temperatures from TmProt,¹²⁰ implemented into EnzymeMiner 2.0¹²²), in both cases we do not get significant correlation between predictions and experimental data, and both tools consistently predict all sequences to be moderately stable or not thermostable. For comparison, we also tested whether an older tool, ThermoPred,¹²³ which uses a support vector machine based method in order to predict whether a protein sequence corresponds to a thermophilic protein or not, could correctly identify thermophilic proteins in our set, and this tool gave much higher thermophilic probability to the designed sequences compared to the other sequences (see the **Supplementary Table**). The inferior performance of the more modern machine-learning based tools on the designed variants is likely due to dependence on training set, and poor performance when faced with novel sequences. Further, the training sets will be dominated by data on natural and extant sequences, whereas here we are dealing with artificial tetrameric scaffolds. We note that, obviously, two ML-based methods is not exhaustive, and this data was just shown for testing and illustrative comparison.

Finally, while the T_m -value is a good indicator for protein stability, determination of activity losses during incubation under process-relevant conditions is decisive for the catalytic performance. In view of the poor solubility of dibornane-type monoterpenols, conducting the biotransformation at elevated temperatures is desirable. The half-life time ($\tau_{1/2}$) of 01-080 and *Sr*BDH1 was determined at 50°C, 65°C and 70°C (**Figure S11**). At 70°C, both variants rapidly denatured, with catalytic activity completely lost within 5 min of incubation. In contrast, at 50°C the enzymes were both remarkably stable, complicating reliable determination of the $\tau_{1/2}$ -value. Within a 3-day window of incubation at 50°C, 70% of the activity of 01-080 and 48% in *Sr*BDH1 remained. On day 4, precipitation of both enzymes lead to the termination of the experiment. The

most pronounced difference was observed at 65°C. Design 01-080 displayed a $\tau_{1/2}$ of 4.23 minutes, and retained residual activity of 14% after 30 minutes of incubation, whereas the wild-type *SrBDH1* was completely deactivated within the first 5 minutes. The higher residual activity of 01-080 at day 3 of incubation at 50°C, paired with the retention of activity at 65°C highlights a successful stabilization of the redesign in comparison to the wild-type. Furthermore, activity assays conducted at 45°C and 55°C showed a constant activity ratio between the two enzymes (**Table S3**), confirming that while *SrBDH1* is functional at elevated temperatures, 01-080 provides a wider operational window for high-temperature biocatalysis.

Modelling Stereoselectivity in the Most Stable Design (Design 01-080)

Following the strategy established in our previous work,⁵² we assessed putative stereoselectivity toward (+)- and (-)-borneol using short TorchANI-Amber ML/MM simulations,¹²⁴⁻¹²⁶ in which borneol was treated at the machine learning level while the protein environment was described classically. Substrate stability in a catalytically active conformation in the active site was assessed by monitoring two key catalytic distances, d_1 , which corresponds to the hydride-transfer geometry between the C₄ atom of the NAD⁺ nicotinamide ring and the oxidizable carbon of borneol, and d_2 , which describes the hydride transfer between the borneol hydroxyl group and the catalytic tyrosine (**Figure 4**). In our prior work, these distances provided clear discrimination between the different borneol enantiomers in simulations of wild-type and ancestral *SrBDH1* variants, with good agreement with experimental selectivity trends.⁵²

For each variant/enantiomer pair, 10 ns ML/MM simulations were sufficient to stabilize productive binding modes and resolve dynamic differences between (+)- and (-)-borneol. Analysis of the resulting d_1/d_2 distance distributions and corresponding kernel density plots (**Figure 4**) reveals that our most stable design, 01-080, selectively maintains catalytically competent

geometries for (+)-borneol, while failing to do so for the (–)-borneol (see also **Table S3**). In contrast, while ancestor N6 still shows high activity with (+)-borneol, it is the most active ASR variant on (–)-borneol.⁵² These results indicate a clear preference for (+)-borneol in these designs and suggest that in these *de novo* designs stereoselectivity arises from differential dynamic stabilization of reactive conformations rather than from static binding geometries alone.

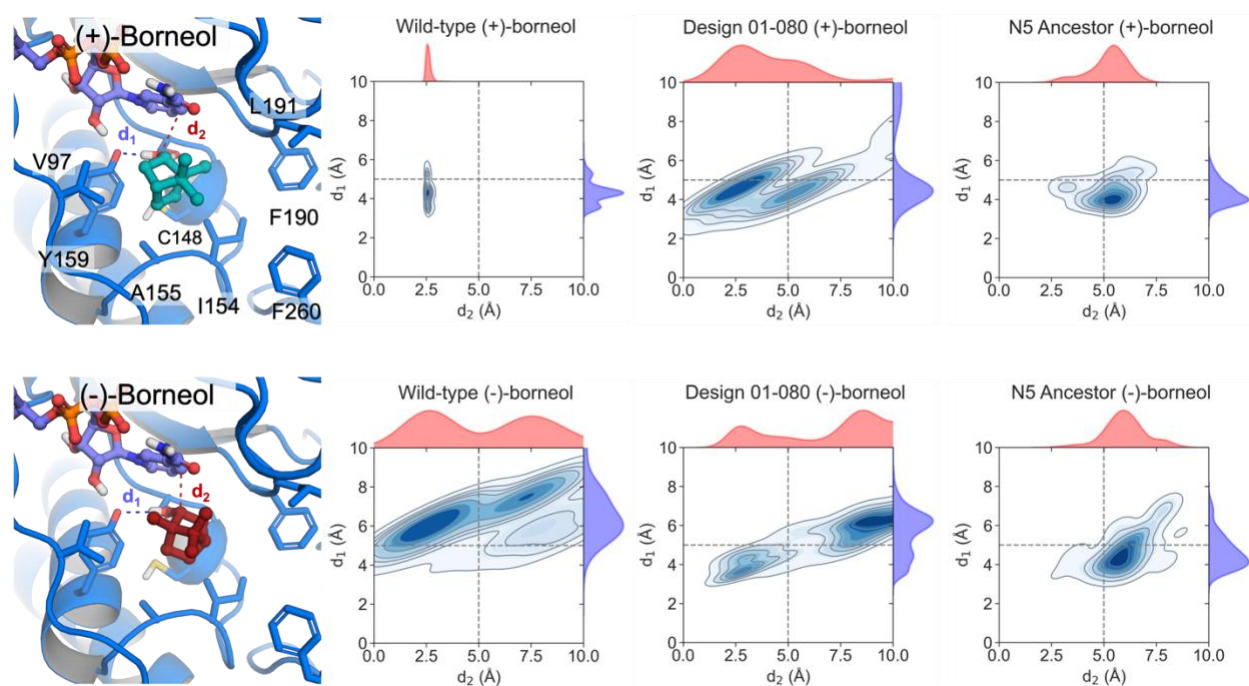


Figure 4. Kernel density estimates (KDEs) of the distribution of key catalytic distances between (+)- and (–)-borneol and wild-type and *de novo* designed *SrBDH1*, during 3 x 10 ns TorchANI-Amber¹²⁶ ML/MM simulations. Distance d_1 (blue) corresponds to the interaction between the catalytic tyrosine and the substrate, while d_2 (red) represents the hydride-transfer donor-acceptor distance. Numbers refer to FAMPNN designs;⁶⁴ for the corresponding sequences see the section “**Wild-type, Designed 01-080 and Reconstructed Sequence N5**” in the **Supporting Information**.

Conclusions

Protein stabilization is an important (but challenging) “Holy Grail” problem for biocatalysis,^{1,2} in particular if one wants to achieve this without compromising activity or selectivity.³⁻¹⁵ There has recently been an explosion of approaches for protein stability design (*e.g.*, refs. ^{39,43-48,50,51,127}, among others). In the current study, we performed full-atom protein sequence design with sidechain conditioning using FAMPNN,⁶⁴ to engineer thermostable variants of the borneol dehydrogenase from *Salvia rosmarinus* (*SrBDH1*^{54,70}). FAMPNN has previously been suggested to be a powerful tool for design of thermostable and active enzymes.⁶⁴

Multi-objective optimization (in this case enhancing thermostability while also retaining enantioselectivity) is challenging in general, and particularly so in the BDH enzyme family, for a number of reasons. (1) This family includes both selective and unselective enzymes towards their natural chiral substrate (1-borneol);⁵² however, unselective enzymes dominate, as stereoselective (+)-1-borneol oxidation does not confer an obvious evolutionary advantage to most plants – with the exception of perhaps those, where the (–)-enantiomer is found as constituent of the essential oils.^{71,128-131} (2) Further, from a structural perspective, the composition of the hydrophobic active site pockets of unselective and selective BDH is very similar,⁵² and point mutations in the highly selective *SrBDH1* active site did not reduce its selectivity towards (+)-borneol.⁵⁴ Building on this, detailed molecular simulations indicate that BDH selectivity is driven by dynamical, not structural considerations,⁵² and thus sequence or structure based approaches alone would struggle to predict or rationalize this selectivity. Finally, (3) characterization of ancestral BDHs obtained from ASR indicate that selectivity is lost in ancestors of the highly selective *SrBDH1*.⁵² While it would be possible to recreate selectivity with first-shell mutations, as discussed, for instance, ref. ⁶⁶, it would be necessary to do so for each unselective variant. It can be argued that similar considerations

apply for biocatalysts in which higher selectivity or another catalytic property was introduced by enzyme engineering. It is highly unlikely that reconstructed ancestors of these sequences retain these improved properties, and the re-introduction into the ancestors by mutagenesis could be at best tedious and might not even work (see also refs. ^{132,133} for examples of ASR impacting selectivity and other enzyme characteristics).

By avoiding the *SrBDH1* active site residues and guiding our FAMPNN⁶⁴ design protocol with residue conservation filtering using ConSurf,^{73,74} we were able to generate *SrBDH1* variants with up to 10 °C enhanced thermostability, while retaining selectivity towards (+)-borneol, unlike the ancestral enzymes identified in our prior work.⁵² Detailed sequence and dynamical analysis indicated once again the importance of conformational dynamics expressed through both scaffold flexibility and active site plasticity in retaining this selectivity, and, encouragingly, use of a scoring card method (SCMTPP)⁸⁴ allowed us to predict thermal stability from sequence in good correlation with experimental measurements (see the **Supplementary Table**). This is significant, because as there is an explosion of ML-based tools to predict enzyme thermal stability (e.g., refs. ^{84,123,134,135}) kinetic parameters (e.g., refs. ¹³⁶⁻¹⁴⁰) and even specificity and selectivity (e.g., refs. ¹⁴¹⁻¹⁴³, note these references are a subset of current studies as the field moves so fast), it will become possible to increasingly automate the design process computationally, minimizing the need for experimental involvement apart from validating top-ranking computational hits, making complete computational design¹⁴⁴ a reality. This provides a transformative leap forward for biocatalysis.

Materials and Methods

ConSurf analysis. Residue conservation scores were calculated using the ConSurf webserver^{73,74} following the standard workflow, including homolog retrieval using the HMMER webserver,¹⁴⁵

multiple sequence alignment with MAFFT,^{146,147} and Bayesian estimation of evolutionary conservation (for methodological details, see ^{73,74}).

Chai-1 models. Initial enzyme-ligand complexes were built with Chai-1,¹⁴⁸ starting from the *SrBDH1* FASTA sequence (see the **Supporting Information** for sequence), together with the SMILES representations of NAD⁺ and borneol. Protonation states were assigned for pH 7 using the H++ server,^{149,150} with the catalytic tyrosine constrained to its deprotonated form. Hydrogen atoms for the oxidized NAD⁺ cofactor was introduced with Open Babel.¹⁵¹

MD simulation setup. Starting coordinates were obtained from Chai-1,¹⁴⁸ as described above. Protein residues were described using the AMBER ff19SB force field.¹⁵² Force-field parameters for NAD⁺, borneol and the deprotonated tyrosine residue were derived by restrained electrostatic potential (RESP) fitting¹⁵³ at the HF/6-31G(d) level of theory, using Antechamber,¹⁵⁴ based on gas-phase geometries optimized at the B3LYP/6-31G(d) level of theory using Gaussian 16 Rev. B.01.¹⁵⁵ All other non-standard parameters were obtained using the General AMBER Force Field (GAFF2).¹⁵⁶ All systems were assembled using the tleap module of AmberTools25,¹⁵⁷ and subjected to an initial vacuum energy minimization using the *sander* module of AMBER24.¹⁵⁷ This was followed by solvation in a truncated-octahedral box of OPC water molecules,¹⁵⁸ extending at least 13 Å from the protein in all axes.

Each solvated system was then minimized in three successive stages, relaxing first hydrogen atoms, then the solvent, and finally the complete system. Heating was performed from 100 to 300 K over 50 ps of simulation time under NVT conditions, using a Langevin thermostat (collision frequency 1.0 ps⁻¹),¹⁵⁹ while restraining all solute atoms with harmonic potentials of 40 kcal mol⁻¹ Å⁻² to preserve the initial geometry during equilibration. Then, the systems were switched to a constant pressure (NPT) scheme, using a Berendsen barostat, at 300 K.¹⁶⁰ Long range interactions

were treated using the Particle Mesh Ewald (PME) method,¹⁶¹ with an 10 Å cut-off for non-bonded interactions. The SHAKE algorithm¹⁶² was employed to restrain the hydrogen atoms in water molecules, allowing for a 2 fs step size for all simulations. Equilibration was performed in three replicates per system, using different initial velocities (assigning different random seeds).

Equilibrated structures were then used as starting points for two different sets of simulations: 3 x 500 ns production MD simulations, and 3 x 10 ns hybrid machine learning/molecular mechanics (ML/MM) simulations (one production MD or ML/MM trajectory from each of three equilibration replicates per system). ML/MM simulations were performed using the TorchANI-AMBER¹²⁴⁻¹²⁶ interface implemented in AmberTools25.¹⁵⁷ In these simulations, the borneol ligand was treated as the ML region, and described using a neural-network potential trained on high-level quantum chemical data, enabling an explicit description of polarization and short-range electronic effects at near QM accuracy, while the remainder of the system was modelled at the MM level, using the previously defined parameters for our MD simulations, as in a standard QM/MM calculation. The ML/MM coupling was implemented using electrostatic embedding with a 12Å ML-MM interaction cutoff, and the ANI architecture^{1,163,164} was exploited using an ANI-MBIS-Q model.¹⁶⁵ Substrate stability was monitored throughout the simulations by tracking two key catalytic distances, namely the interaction between the catalytic tyrosine and the substrate (d_1), and the hydride-transfer donor-acceptor distance (d_2) (see **Figure 4**). Prior work demonstrated this protocol can adequately discriminate between borneol enantiomers in ML/MM simulations of a range of BDHs.⁵² Both conventional MD and ML/MM trajectories were analyzed using CPPTRAJ⁹⁸ and MDAnalysis.^{99,100} Data processing, statistical analysis, and visualization were carried out using Matplotlib.⁹²

A Zenodo package containing all starting structures, representative input files, representative frames from our simulations in PDB format, simulation checkpoint files and any non-standard parameters can be found at DOI: 10.5281/zenodo.19267529.

Chemicals. All chemicals were purchased from Sigma Aldrich (Vienna, Austria) or Roth (Karlsruhe, Germany) and used without further purification unless otherwise stated.

Cloning of the redesigned SrBDH1 variants. Genes of the redesigned SrBDH1 variants (01-037, 01-080, 01-099, 01-173, 02-037, 02-389, 02-911, and 02-939) carrying an N-terminal 6x His-tag and complementary sequences to the vector pET28a(+) were obtained as synthetic genes (Twist Bioscience, USA). pET28a(+) (final conc. 50 ng/μL) was linearized by restriction digest with NcoI HF (1 μL) and XhoI (1 μL) (New England Biolabs, Germany) in Cut Smart buffer (5μL, final reaction volume 50μL) at 37°C for 15 minutes. Enzymes were deactivated at 80°C for 20 minutes. The synthetic genes were assembled and cloned into the vector pET28a(+) via Gibson Assembly following the original protocol.¹⁶⁶ The reaction was carried out at 50°C, 300 rpm for 45 minutes. Chemo-competent *Escherichia coli* TOP10 cells were transformed with 5 μL of the assembly mix. Isolated plasmids (Wizard Plus SV Minipreps DNA Purification System, Promega, USA) were sequenced by Sanger sequencing (Microsynth AG, Switzerland) and subsequently used for transformation of *E. coli* BL21 (DE3) for expression of the redesigns.

Investigation of enzyme expression. Expression studies were carried out to investigate the optimal expression conditions of the redesigned variants in *E. coli* BL21 (DE3). Main cultures of *E. coli* BL21 (DE3) pET28a(+) carrying the respective redesigned variant were prepared by inoculating 100 mL of LB medium containing kanamycin (40 mg/L) in baffled shaking flasks, with overnight cultures to a final OD₆₀₀ of 0.2. After the inoculation, the main cultures were incubated at 37°C and 130 rpm until an OD₆₀₀ of 0.8-1 was reached. At OD₆₀₀ of 0.8-1 the protein expression was

induced by addition of 1 mM IPTG. Following the induction, the culture was split into 25 ml aliquots and transferred to sterile 50 ml falcon tubes to assess different expression temperatures. Protein expression proceeded overnight at 18, 28 or 37°C and 130 rpm. Before induction, 4 h and 20 h after induction, samples corresponding to an OD₆₀₀ of 1. The samples were centrifuged at 13.000 rpm, 4°C for 10 min (Centrifuge 5415 R, F-45-24-11, Eppendorf, Germany) and the supernatant discarded. The cells were lysed by addition of 100 µl BugBuster Master Mix (Merck Millipore, USA) followed by incubation at 25°C, 350 rpm for 15 min. Samples were subsequently centrifuged at 13,000 rpm, 4°C for 10 min (Centrifuge 5415 R, F-45-24-11, Eppendorf, Germany). The supernatant was collected and the pellet was resuspended in 100 µL of 8M urea for further analysis on SDS-PAGE.

Cultivation and expression. Main cultures of *E. coli* BL21 (DE3) pET28a(+) carrying the respective redesigned variant were prepared by inoculating 250 mL of LB medium containing kanamycin (40 mg/L) in baffled shaking flasks, with overnight cultures to a final OD₆₀₀ of 0.1. After the inoculation, the main cultures were incubated at 37°C and 130 rpm until an OD₆₀₀ of 0.8-1 was reached. At OD₆₀₀ of 0.8-1 the protein expression was induced by addition of 1mM IPTG. Protein expression proceeded overnight at 28°C, 130 rpm.

Cell harvest and lysis. Cells were harvested by centrifugation at 4,000 rpm, 4°C for 15 min (Avanti® J-20 XP and JXN-26, rotor JA-10, Beckman coulter, USA). After the centrifugation, the supernatant was discarded and the cell pellet was resuspended in 20 mL of buffer A (100 mM KPi, 500 mM NaCl, 10% glycerol v/v, pH 8.0). The suspension was centrifuged at 4,000 rpm and 4°C for 15 min (Centrifuge 5810R, A-4-62, Eppendorf, Germany). The supernatant was discarded, and the pellets were then either stored at -20°C for later use or resuspended in 20 mL of buffer A for cell lysis. Cell lysis was facilitated by sonication at 4°C (output control: 6, duty cycle 70%, 6 min;

Sonifier S-250A, Branson Ultrasonics™, USA). The lysed samples were centrifuged at 13,200 rpm, 4°C for 15 min (Centrifuge 5415 R, F-45-24-11, Eppendorf, Germany) to remove cell debris. The supernatant (cell free extract) was filtered through a 0.2 µm syringe filter (Filtropur S, PES, 0.2 µm, Sarstedt, Germany) and used for protein purification.

Protein purification. The proteins were purified by affinity chromatography using High Affinity Ni-Charged Resin (GenScript, China) in gravity columns. Columns contained 5 mL of the sepharose resin. The columns were washed with three volumes of ddH₂O, and adjusted with two volumes of buffer A, before chilling on ice under rotary shaking at 150 rpm (Titramax 1000, Heidolph Instruments, Germany) for 15 min. Next, the filtered cell free extract was incubated with the sepharose under rotary shaking on ice, at 150 rpm for 15 min. The flowthrough was collected, and the columns were washed with nine column volumes of buffer B (100 mM KPi, 500 mM NaCl, 30 mM imidazole, 10% glycerol v/v, pH 8.0). Proteins were eluted by application of three column volumes of buffer C (100 mM KPi, 500 mM NaCl, 500 mM imidazole, 10% glycerol v/v, pH 8.0). Gravity columns were washed with three column volumes of buffer C, and ddH₂O and then stored in 20% EtOH at 4°C. The elution fractions were desalted by dialysis overnight with Spectra/Por® 1 Standard RC Dry Dialysis kits (MWCO 6-8 kDa, Spectrum Labs, SAD) in buffer A at 4°C under rotary shaking at 150 rpm (Titramax 1000, Heidolph Instruments, Germany). Desalted protein samples were aliquoted and stored at -20°C.

Protein quantification and SDS-PAGE analysis. Protein concentration was determined spectrophotometrically at 562 nm in a 96-well format, following the Pierce™ BCA Protein Assay Kit protocol (Thermo Fisher Scientific, US) in the Plate Reader Synergy (BioTek Instruments, US). Measurements were performed in triplicates.

SDS-PAGE was conducted to analyze the purified fractions. The samples were mixed with NuPage Sample Reducing Agent (10x), NuPage LDS Sample Buffer (4x) (Thermo Fischer Scientific, US) and ddH₂O to a final concentration of 1 mg/ml. Samples were heated at 90°C for 10 min. Precast SDS gels (NuPAGE 4-12% Bis-Tris Gel, Thermo Fisher Scientific, US) were inserted into a XCell SureLock Mini-Cell (Thermo Fisher Scientific, US), filled with 1x NuPAGE MES SDS running buffer (Thermo Fisher Scientific, US). 5 µl of the protein standard (PageRuler Prestained, Thermo Fisher Scientific, US) and 10 µl of the samples (1 mg/mL) were loaded onto the gel. The gel was developed under a current of 200 V, and 110 mA for 35 min (PowerEase®500, Invitrogen, US). Afterward, the gels were stained by submerging them in (de-)staining solution (4.9% v/v phosphoric acid, 6.5% v/v EtOH, 0.011% w/v Coomassie Blue G-250 and 1% w/v β-cyclodextrin) for 40 min at room temperature under slow shaking on a tabletop horizontal shaker. Further destaining was facilitated by ddH₂O at room temperature and continued shaking for 1 h.

Size-exclusion chromatography (SEC). SEC was carried out on an Äkta pure system (Cytiva, USA) equipped with a Superdex 200 Increase 10/300 GL (Cytiva, USA). The system was equilibrated to buffer A. Protein Standard Mix 15-600 kDA (Sigma Aldrich, USA) was dissolved according to the manufacturers protocol and applied to the column for calibration, before recording the wildtype *SrBDH1* and 01-080. Signals were recorded at 220, 280 and 340 nm.

Differential scanning fluorimetry. Determination of the unfolding temperatures of redesigned variants and *SrBDH1* at reaction conditions was conducted with the ThermoFluor protocol.¹⁶⁷ A 200x solution of SYPRO™ Orange (Thermo Fisher Scientific, US) was prepared by diluting a 5000x stock solution with reaction buffer (0.1 mM Tris-HCl, pH 9.0). Enzyme dilutions of 5, 10, and 15 µM were prepared by the addition of reaction buffer. 45 µl of diluted sample solution or buffer (blank) was combined with 5 µl of 200x SYPRO™ Orange solution. Denaturation of the

proteins was determined by an increase in fluorescence upon protein denaturation over increased temperature (7500 Real-time PCR System, Applied Biosystems, US). After an initial holding temperature of 30 sec at 25°C, a linear gradient up to 95°C with a 0.3% increase was applied. Samples were measured in triplicate. The resulting unfolding curves were plotted as fluorescent signal over temperature after blank subtraction. The unfolding temperature (T_m) was determined as the inflection point (maximum of the first derivative). To minimize noise and improve peak detection, raw data were smoothed using a Savitzky-Golay filter prior to derivative calculation.

Circular dichroism (CD) spectroscopy. CD spectroscopy of the ancestral variants was carried out using a Jasco J-1500 CD Spectrometer (USA). Samples were prepared at a concentration of 0.1 mg/ml of enzyme in 10 mM KP_i , 50 mM Na_2SO_4 at pH 8.0. The signal was recorded at 194 or 217 nm during a temperature ramp of 1°C/min from 20°C to 95°C. Unfolding temperatures were calculated using Spectral Manager. CD spectra of the wild-type *SrBDH1* and the redesigned variant 01-080 were recorded using a Chirascan™ Circular Dichroism Spectrometer (Applied Photophysics Ltd., UK) at 25°C, between 180-260 nm with a step size of 1 nm, at a protein concentration of 2 mg/mL in buffer A. Spectral scans were carried out in triplicate. Thermal denaturation curves of the wild-type *SrBDH1* and 01-080 were recorded using a Chirascan™ Circular Dichroism Spectrometer (Applied Photophysics Ltd., UK), between 190 to 260 nm with a step size of 1 nm. A temperature gradient from 25 °C to 100 °C was applied in ramp mode. Protein refolding was probed by cooling the sample back to 25 °C. The melting temperature (T_m) was determined from the thermal shift of the denaturation curves at 220 nm. T_m was determined as the inflection point (maximum of the first derivative) of the thermal shift. To minimize noise and improve peak detection, raw data were smoothed using a Savitzky-Golay filter prior to derivative calculation.

Determination of the specific activity. Determination of the oxidation rates towards (+)-borneol, (–)-borneol or *r*-isoborneol was performed with purified enzymes. Reactions containing 0.2, or 9 mM of the respective alcohol (dissolved in DMSO), 9 mM NAD⁺, 20% total DMSO (v/v) in 0.1 M Tris-HCl (pH 9.0) were started by the addition of 2 μM purified enzyme (final concentration). The reactions were carried out at 25°C, 300 rpm. Over time, multiple samples were taken and stopped by 1:1 addition of 1 M NaOH. Substrate and product were extracted via a two-step extraction process with 1:2 MTBE, the organic layers were collected and dried over Na₂SO₄ and product formation was quantified via gas chromatography. Experiments were carried out in triplicate. The oxidation rates of the wild-type *SrBDH1* and the 01-080 variant at elevated temperatures were recorded similarly at 45 °C and 50 °C incubation temperature and treated as described before.

Determination of k_{app} . Kinetic parameters of 01-080 were obtained by measuring initial velocities in the presence of varying concentrations of (+)-borneol (0.5–9 mM) at a constant 9 mM of NAD⁺ under the conditions given above. Reactions were started by the addition of 20 μM purified enzyme. Kinetic parameters were obtained by fitting the obtained data to the Michaelis-Menten equation with SigmaPlot 14.0 (Systat Software). Experiments were carried out in triplicates.

*Determination of *E*-value.* Determination of the stereoselectivity expressed as the enantiomeric ratio (*E*-value) was carried out by racemic resolution. Reactions containing 9 mM *r*-borneol (dissolved in DMSO), 9 mM NAD⁺, 20% total DMSO (v/v) in 0.1 M Tris-HCl (pH 9.0) were started by the addition of 2 μM purified enzyme (final concentration). Reactions were carried out at 25°C, 300 rpm and stopped after 24 h, and treated as described above.

Quantitative GC analysis. Achiral GC analysis of borneol and camphor was carried out on a Shimadzu GC-2030 GC-FID system equipped with a ZB-5 column (30 m length, 0.32 mm inner

diameter, 0.25 μ M film thickness, Shimadzu, JP). Samples were injected in split mode (9:1) with an injection temperature of 250°C and FID temperature of 320°C. The instrument was programmed to an initial oven temperature of 50°C. After 5 min of holding time, a linear increase of 40°C/min up to 300°C was set, followed by 5 min holding. N₂ was set to a constant flow of 30 ml/min. Chiral analysis was carried out on a Shimadzu GC-2030 GC-FID system equipped with a Hydrodex β TBDM column (25 m length, 0.25 mm inner diameter, Macherey-Nagel, Germany). Samples were injected in split mode (50:1) at an injection temperature of 250°C and FID temperature of 320°C. The instrument was programmed to an initial oven temperature of 90°C. After 5 min of holding time, a linear increase of 2°C/min up to 150°C was set, followed by a linear increase of 45°C/min up to 200°C and (2 min holding). N₂ was set to a constant flow of 24 ml/min.

Half-life time determination. Half-life time of 01-080 and SrBDH1 were determined via spectrophotometric measurements. During the measurement, the increase of NADH was monitored at 340 nm within 30 min. Reactions at t_0 were performed at RT, with 20 μ M purified enzyme, 2 mM (+)-borneol (dissolved in DMSO, 1% v/v DMSO in 0.1 M Tris-HCl, pH 9.0). Reactions were started with the addition of 2 mM NAD⁺ (final conc.). At later timepoints, reaction mixes containing the purified enzyme, DMSO and buffer were incubated at 50, 65 or 70°C. After incubation, the samples were cooled to room temperature and (+)-borneol was added. The reactions were started and monitored as described before.

Acknowledgments

SCLK is the Georgia Research Alliance – Vasser Wooley Chair in Molecular Design at Georgia Tech. We thank Georgia Tech and the Georgia Research Alliance for support of this project. We further acknowledge the National Academic Infrastructure for Supercomputing in Sweden (NAISS) for awarding this project access to the LUMI supercomputer, owned by the EuroHPC

Joint Undertaking, hosted by CSC (Finland) and the LUMI consortium through grant agreement no. NAISS 2024/8-16 and NAISS 2025/8-11. This work used the Hive cluster, which is supported by the National Science Foundation under grant number 1828187. This research was supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA. B.D.G. would like to thank Dr. Ignacio Pickering and Prof. Dario Estrin for their support on the TorchANI-Amber ML/MM simulations. J.Z. would like to thank Anna Schürfer for her help with the CD measurements.

Declaration of Interests

S.C.L.K. is an informal consultant for Dayhoff Labs, a technology company that builds foundational AI models to solve problems at the chemistry/biology interface.

Supporting Information

Additional computational and experimental analysis, all SrBDH1 sequences of relevance to this work, and detailed sequence-based analysis of physicochemical properties are provided as Supporting Information. A Zenodo data package with necessary data to reproduce our simulations is provided at DOI: 10.5281/zenodo.19267529.

References

1. Smith, J.S., Zubatyuk, R., Nebgen, B., Lubbers, N., Barros, K., Roitberg, A.E., Isayev, O., and Tretiak, S. (2020). The ANI-1ccx and ANI-1x Data Sets, Coupled Cluster and Density Functional Theory Properties for Molecules. *Sci. Data* 7, 134.
2. Wu, S., Snajdrova, R., Moore, J.C., Baldenius, K., and Bornscheuer, U.T. (2020). Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angew. Chem. Int. Ed.* 60, 88–119.
3. Goldenzweig, A., and Fleishman, S.J. (2018). Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* 87, 105–129.
4. Shoichet, B.K., Baase, W.A., Kuroki, R., and Matthews, B.W. (1995). A Relationship Between Protein Stability and Protein Function. *Proc. Natl. Acad. Sci. USA* 92, 452–456.
5. Beadle, B.M., and Shoichet, B.K. (2002). Structural Bases of Stability-Function Tradeoffs in Enzymes. *J. Mol. Biol.* 321, 285–296.
6. Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. USA* 103, 5869–5874.
7. Bloom, J.D., Arnold, F.H., and Wilke, C.O. (2007). Breaking Proteins with Mutations: Threads and Thresholds. *Mol. Sys. Biol.* 3, 76.
8. Tokuriki, N., Francois, S., Schymkowitz, J., Serrano, L., and Tawfik, D.S. (2007). The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J. Mol. Biol.* 369, 1318–1332.
9. Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How Protein Stability and New Functions Trade-Off. *PLoS Comput. Biol.* 4, e1000002.

10. Traxlmayr, M.W., Lobner, E., Antes, B., Kainer, M., Wiederkum, S., Hasenhindl, C., Stadlmayr, G., Rueker, F., Woisetschlaeger, M., Moulder, K., and Obinger, C. (2013). Directed Evolution of Her2/neu-binding IgG1-Fc for Improved Stability and Resistance to Aggregation by Using Yeast Surface Display. *Prot. Eng. Des. Sel.* 26, 255–265.
11. Miller, S.R. (2017). An Appraisal of the Enzyme Stability-Activity Trade-Off. *Evolution* 71, 1876–1887.
12. Studer, S., Hansen, D.A., Pianowski, Z.L., Mittl, P.R., Debon, A., Guffy, S.L., Der, B.S., Kuhlman, B., and Hilvert, D. (2018). Evolution of a Highly Active and Enantiospecific Metalloenzyme From Short Peptides. *Science* 362, 1285–1288.
13. Teufl, M., Zajc, C.U., and Traxlmayr, M.W. (2022). Engineering Strategies to Overcome the Stability-Function Trade-Off in Proteins. *ACS Synth. Biol.* 11, 1030–1039.
14. King, B.R., Sumida, K.H., Caruso, J.L., Baker, D., and Zalatan, J. (2025). Computational Stabilization of a Non-heme Iron Enzyme Enables Efficient Evolution of New Function. *Angew. Chem. Int. Ed.* 64, e202414705.
15. Ding, K., Li, Z., Tu, T., Luo, J., and Luo, Y. (2026). Deconvolving Mutation Effects on Protein Stability and Function with Disentangled Protein Language Models. *bioRxiv* 2026, 10.64898/62026.64802.64803.703560.
16. Dill, K.A. (1990). Dominant Forces in Protein Folding. *Biochemistry* 29, 7133–7155.
17. Taverna, D.M., and Goldstein, R.A. (2002). Why Are Proteins Marginally Stable? *Proteins* 46, 105–109.
18. Wilson, A.E., Kosater, W.M., and Liberles, D.A. (2020). Evolutionary Processes and Biophysical Mechanisms: Revisiting Why Evolved Proteins Are Marginally Stable. *J. Mol. Evol.* 88, 415–417.

19. Drummond, D.A., Iverson, B.L., Georgiou, G., and Arnold, F.H. (2005). Why High-error-rate Random Mutagenesis Libraries are Enriched in Functional and Improved Proteins. *J. Mol. Biol.* *350*, 806–816.
20. Camps, M., Herman, A., Loh, E., and Loeb, L.A. (2007). Genetic Constraints on Protein Evolution. *Crit. Rev. Biochem. Mol. Biol.* *42*, 313–326.
21. Brown, K.M., Costanzo, M.S., Xu, W., Roy, S., Lozovsky, E.R., and Hartl, D.L. (2010). Compensatory Mutations Restore Fitness During the Evolution of Dihydrofolate Reductase. *Mol. Biol. Evol.* *27*, 2682–2690.
22. Studer, R.A., Christin, P.-A., Williams, M.A., and Orengo, C.A. (2014). Stability-Activity Tradeoffs Constrain the Adaptive Evolution of RubisCO. *Proc. Natl. Acad. Sci. USA* *111*, 2223–2228.
23. Starr, T.N., and Thornton, J.W. (2016). Epistasis in Protein Evolution. *Prot. Sci.* *25*, 1204–1218.
24. Stimple, S.D., Smith, M.D., and Tessier, P.M. (2019). Directed Evolution Methods for Overcoming Trade-Offs Between Protein Activity and Stability. *AIChE J.* *66*, e16814.
25. Jayaraman, V., Toledo-Patiño, S., Noda-Garíá, L., and Laurino, P. (2022). Mechanisms of Protein Evolution. *Prot. Sci.* *31*, e4362.
26. Xing, H., Wang , P., Yan, X., Yang, Y., Li, X., Liu, R., and Zhou, Z. (2023). Thermostability Enhancement of *Escherichia coli* Phytase By Error-Prone Polymerase Chain Reaction (epPCR) and Site-Directed Mutagenesis. *Front. Bioeng. Biotechnol.* *11*, 1167530.

27. Acevedo, J.P., Reetz, M.T., Asenjo, J.A., and Parra, L.P. (2017). One-Step Combined Focused epPCR and Saturation Mutagenesis for Thermostability Evolution of a New Cold-Active Xylanase. *Enzyme Microb. Technol.* *100*, 60–70.
28. Romero, P.A., and Arnold, F.H. (2009). Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* *10*, 866–876.
29. Tokuriki, N., and Tawfik, D.S. (2009). Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* *19*, 596–604.
30. Goldsmith, M., and Tawfik, D.S. (2012). Directed Enzyme Evolution: Beyond the Low-Hanging Fruit. *Curr. Opin. Struct. Biol.* *22*, 406–412.
31. Tokuriki, N., Jackson, C.J., Afriat-Jurnou, L., Wyganowski, K.T., Tang, R., and Tawfik, D.S. (2012). Diminishing Returns and Tradeoffs Constrain the Laboratory Optimization of an Enzyme. *Nat. Commun.* *3*, 1257.
32. Sailer, Z.R., and Harms, M.J. (2017). High-Order Epistasis Shapes Evolutionary Trajectories. *PLoS Comput. Biol.* *13*, e1005541.
33. Ashkenazi, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., and Pupko, T. (2012). FastML: A Web Server for Probabilistic Reconstruction of Ancestral Sequences. *Nucleic Acids Res.* *40*, W580–W584.
34. Gumulya, Y., and Gillam, E.M.J. (2017). Exploring the Past and the Future of Protein Evolution with Ancestral Sequence Reconstruction: The ‘Retro’ Approach to Protein Engineering. *Biochem J.* *474*, 1–19.
35. Prakinee, K., Phaisan, S., Kongjaroon, S., and Chaiyen, P. (2024). Ancestral Sequence Reconstruction for Designing Biocatalysts and Investigating Their Functional Mechanisms. *JACS Au* *4*, 4571–4591.

36. Williams, P.D., Pollock, D.D., Blackburne, B.P., and Goldstein, R.A. (2006). Assessing the Accuracy of Ancestral Protein Reconstruction Methods. *PLoS Comput. Biol.* 2, e69.
37. Gaucher, E.A., Govindarajan, S., and Ganesh, O.K. (2008). Paleotemperature Trend for Precambrian Life Inferred from Resurrected Proteins. *Nature* 451, 704–707.
38. Perez-Jimenez, R., Li, J., Kosuri, P., Sanchez-Romero, I., Wiita, A.P., Rodriguez-Larrea, D., Chueca, A., Holmgren, A., Miranda-Vizuet, A., Becker, K., et al. (2009). Diversity of Chemical Mechanisms in Thioredoxin Catalysis Revealed by Single-Molecule Force Spectroscopy. *Nat. Struct. Mol. Biol.* 16, 890–896.
39. Risso, V.A., Gavira, J.A., Mejia-Carmona, D.F., Guacher, E., and Sanchez-Ruiz, J.M. (2013). Hyperstability and Substrate Promiscuity in Laboratory Resurrections of Precambrian β -Lactamases. *Biochem. J.* 429, 243–249.
40. Myrtollari, K., Calderini, E., Kracher, D., Schöngäßner, T., Galušić, S., Slavica, A., Taden, A., Mokos, D., Schrüfer, A., Wirnsberger, G., et al. (2024). Stability Increase of Phenolic Acid Decarboxylase by a Combination of Protein and Solvent Engineering Unlocks Applications at Elevated Temperatures. *ACS Sustainable Chem. Eng.* 12, 3575–3584.
41. Bershtein, S., Goldin, K., and Tawfik, D.S. (2008). Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* 379, 1029–1044.
42. Akanuma, S., Nakajima, Y., Yokobori, S.-I., Kimura, M., Nemoto, N., Mase, T., Miyazono, K.-I., Tanokura, M., and Yamagishi, A. (2013). Experimental Evidence for the Thermophilicity of Ancient Life. *Proc. Natl. Acad. Sci. USA* 110, 11067–11072.
43. Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P.G.M., and Wyss, M. (2002). The Consensus Concept for Thermostability Engineering of Proteins: Further Proof of Concept. *Protein Eng.* 15, 403–411.

44. Jochens, H., Aerts, D., and Bornscheuer, U.T. (2010). Thermostabilization of an Esterase By Alignment-Guided Focused Directed Evolution. *Protein Eng. Des. Sel.* 23, 903–909.
45. Wood, B.M., Dzamba, M., Fu, X., Gao, M., Shuaibi, M., Barroso-Luque, L., Abdelmaqsoud, K., Gharakhanyan, V., Kitchin, J.R., Levine, D.S., et al. (2025). UMA: A Family of Universal Models for Atoms. *arXiv:2506.23971*.
46. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahren, W., Borst, A.J., Ragotte, R.J., Milles, L.K., et al. (2023). *De Novo* Design of Protein Structure and Function With RFDiffusion. *Nature* 620, 1089–1100.
47. Ahern, W., yim, J., Tischer, D., Salike, S., Woodbury, S.M., Kim, D., Kalvet, I., Kipnis, Y., Coventry, B., Altae-Tran, H.R., et al. (2025). Atom Level Enzyme Active Site Scaffolding Using RFDiffusion2. *bioRxiv*, DOI: 10.1101/2025.1104.1109.648075.
48. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., De Haas, R.J., Bethel, N., et al. (2022). Robust Deep Learning-Baese Protein Sequence Design Using ProteinMPNN. *Science* 378, 49–56.
49. Dieckhaus, H., Brocidiaco, M., Randolph, N.Z., and Kuhlman, B. (2024). Transfer Learning to Leverage Larger Datasets for Improved Prediction of Protein Stability Changes. *Proc. Natl. Acad. Sci. USA* 121, e2314853121.
50. Hie, B.L., Yang, K.K., and Kim, P.S. (2022). Evolutionary Velocity With Protein Language Models Predicts Evolutionary Dynamics of Diverse Proteins. *Cell Syst.* 13, P274–285.E276.
51. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-Scale Prediction of Atomic-Level Protein Structure With a Language Model. *Science* 379, 1123–1130.

52. Zuson, J., Helmer, C.P.O., Di Geronimo, B., Chánique, A.M., Kavčiaková, K., Teijeiro-Juiz, R., Brickel, S., Drienovská, I., Kracher, D., Kamerlin, S.C.L., et al. (2025). Deciphering the Evolutionary Origin of the Enantioselectivity of Short-Chain Dehydrogenases from Plants Toward I-Borneol. *bioRxiv*, DOI: 10.1101/2025.1107.1117.664155.
53. Drienovská, I., Kolanović, D., Chánique, A., Sieber, V., Hofer, M., and Kourist, R. (2020). Molecular Cloning and Functional Characterization of a Two Highly Stereoselective Borneol Dehydrogenase from *Salvia officianalis* L. *Phytochemistry* 172, 112227.
54. Chánique, A.M., Dimos, N., Drienovská, I., Calderini, E., Pantín, M.P., Helmer, C.P.O., Hofer, M., Sieber, V., Parra, L.P., Loll, B., and Kourist, R. (2021). A Structural View on the Stereospecificity of Plant Borneol-Type Dehydrogenases. *ChemCatChem* 13, 2262–2277.
55. Dimos, N., Helmer, C.P.O., Chánique, A.M., Wahl, M.C., Kourist, R., Hilal, T., and Loll, B. (2022). CryoEM Analysis of Small Plant Biocatalysts at Sub-2Å Resolution. *Acta Crystallogr. D Struct. Biol.* 78, 113–123.
56. Calderini, E., Drienovská, I., Myrtollari, K., Pressnig, M., Sieber, V., Schwab, H., Hofer, M., and Kourist, R. (2021). Simple Plug-In Synthetic Step for the Synthesis of (-)-Camphor from Renewable Starting Materials. *ChemBioChem* 22, 2951–2956.
57. Campitelli, P., Modi, T., Kumar, S., and Banu Ozkan, S. (2020). The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* 49, 267–288.

58. Zou, T., Risso, V.A., Gavira, J.A., Sanchez-Ruiz, J.M., and Ozkan, S.B. (2015). Evolution of Conformational Dynamics Determines the Conversion of a Promiscuous Generalist into a Specialist Enzyme. *Mol. Biol. Evol.* 32, 132–143.
59. Cao, L., Coventry, B., Goresnik, I., Huang, B., Sheffler, W., Park, J.S., Jude, K.M., Markovic, I., Kadam, R.U., Versheuren, K.H.G., et al. (2022). Design of Protein-Binding Proteins from the Target Structure Alone. *Nature* 605, 551–560.
60. Vázquez Torres, S., Leung, J.Y., Venkatesh, P., Lutz, I.D., Hink, F., Huynh, H.-H., Becker, J., Yeh, A.H.-W., Uergens, D., Bennet, N.R., et al. (2024). *De Novo* Design of High-Affinity Binders of Bioactive Helical Peptides. *Nature* 626, 435–442.
61. Vázquez Torres, S., Valle, M.B., Mackessy, S.P., Menzies, S.K., Casewell, N.R., Ahmadi, S., Burlet, N.J., Muratsphahić, E., Sappington, I., Overath, M.D., et al. (2025). *De Novo* Designed Proteins Neutralize Lethal Snake Venom Toxins. *Nature* 639, 225–231.
62. Zheng, B., Lu, Z., Wang, S., Liu, L., Ao, M., Zhou, Y., Tang, G., Wang, R., Liu, Y., Zhang, H., et al. (2026). Computational Design of Superstable Proteins Through Maximized Hydrogen Bonding. *Nat. Chem.* 18, 364–373.
63. Lampinen, V., Burastero, O., Guazzelli, I.P., Vögele, F., Pinheiro, F., Nowak, J.S., Alai, M.M.G., and Kjaergaard, M. (2026). Global Analysis of Thermal and Chemical Denaturation Using CheMelt: Thermodynamic Dissection of Highly Thermostable *De Novo* Designed Proteins. *bioRxiv*, DOI: 10.64898/2026.64804.64807.716910.
64. Shuai, R.W., Widatalla, T., Hang, P.-S., and Hie, B.L. (2025). Sidechain Conditioning and Modeling for Full-Atom Protein Sequence Design With FAMPNN. *bioRxiv*, DOI: 10.1101/2025.1102.1113.637498.

65. Reetz, M.T., and Jaeger, K.E. (2000). Enantioselective Enzymes for Organic Synthesis Created by Directed Evolution. *Chemistry* 6, 407–412.
66. Morley, K.L., and Kazlauskas, R.J. (2005). Improving Enzyme Properties: When Are Close Mutations Better? *Trends Biotechnol.* 23, 231–237.
67. Reetz, M.T., and Carballeira, J.D. (2007). Iterative Saturation Mutagenesis (ISM) for Directed Evolution of Functional Enzymes. *Nat. Protoc.* 2, 891–903.
68. Lutz, S. (2010). Beyond Directed Evolution - Semi-Rational Protein Design. *Curr. Opin. Biotechnol.* 21, 734–743.
69. Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.-G., Opperman, D.J., Reetz, M.T., and Acevedo, J.P. (2012). Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* 2, 83–92.
70. Croteau, R., Hooper, C.L., and Felton, M. (1978). Biosynthesis of Monoterpenes: Partial Purification and Characterization of a Bicyclic Monoterpenol Dehydrogenase From Sage (*Salvia officinalis*). *Arch. Biochem. Biophys.* 188, 182–193.
71. Sarker, L.S., Galata, M., Demissie, Z.A., and Mahmoud, S.S. (2012). Molecular Cloning and Functional Characterization of Borneol Dehydrogenase From the Glandular Trichomes of *Lavandula x Intermedia*. *Arch. Biochem. Biophys.* 528, 163–170.
72. Tsang, H.-L., Huang, J.-L., Lin, Y.-H., Huang, K.-F., Lu, P.-L., Lin, G.-H., Khine, A.A., Hu, A., and Chen, H.-P. (2016). Borneol Dehydrogenase from *Pseudomonas* sp. Strain TCU-HL1 Catalyzes the Oxidation of (+)-Borneol and Its Isomers to Camphor. *Appl. Environ. Microbiol.* 82, 6378–6385.

73. Armon, A., Graur, D., and Ben-Tal, N. (2001). ConSurf: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface Mapping of Phylogenetic Information. *J. Mol. Biol.* *307*, 447–463.
74. Ben Chorin, A., Masrati, G., Kessel, A., Narunsky, A., Sprinzak, J., Lahav, S., Ashkenazy, H., and Ben-Tal, N. (2020). ConSurf-DB: An Accessible Repository for the Evolutionary Conservation Patterns of the Majority of PDB Proteins. *Protein Sci.* *29*, 258–267.
75. Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* *19*, 163–164.
76. James, L.C., and Tawfik, D.S. (2003). Conformational Diversity and Protein Evolution - A 60 Year Old Hypothesis Revisited. *Trends Biochem. Sci.* *28*, P361–P368.
77. Ollikainen, N., de Jong, R., and Kortemme, T. (2015). Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-design of Protein-Ligand Specificity. *PLoS Biol.* *11*, e1004335.
78. Maria-Solano, M.A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J., and Osuna, S. (2018). Role of Conformational Dynamics in the Evolution of Novel Enzyme Function. *Chem. Commun.* *54*, 6622–6634.
79. Petrovic, D., Risso, V.A., Kamerlin, S.C.L., and Sanchez-Ruiz, J.M. (2018). Conformational Dynamics and Enzyme Evolution. *J. R. Soc. Interface* *15*, 20180330.
80. Agarwal, P.K., Bernard, D.N., Bafna, K., and Doucet, N. (2020). Enzyme Dynamics: Looking Beyond a Single Structure. *ChemCatChem* *12*, 4704–4720.
81. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* *28*, 235–242.

82. Chaudhury, S., Lyskov, S., and Gray, J.J. (2010). PyRosetta: A Script-Based Interface for Implementing Molecular Modeling Algorithms Using Rosetta. *Bioinformatics* 26, 689–691.
83. Gasteiger, E., Hoogland, C., Gattiker, A., Davaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server (Humana Press).
84. Charoenkwan, P., Chotpatiwetchkul, W., Lee, V.S., Nantasenamat, C., and Shoombuatong, W. (2021). A Novel Sequence-Based Predictor for Identifying and Characterizing Thermophilic Proteins Using Estimated Propensity Scores of Dipeptides. *Sci. Rep.* 11, 23782.
85. Haney, P.L., Stees, M., and Konisky, J.J. (1999). Analysis of Thermal Stabilizing Interactions in Mesophilic and Thermophilic Adenylate Kinases from the Genus *Methanococcus*. *Biol. Chem.* 274, 28453–28458.
86. Haney, P.L., Badger, J.H., Buldak, G.L., Reich, C.L., Wose, C.R., and Olsen, G.J. (1999). Thermal Adaptation Analyzed by Comparison of Protein Sequences from Mesophilic and Extremely Thermophilic *Methanococcus* Species. *Proc. Natl. Acad. Sci. USA* 96, 3578–3583.
87. Melchionna, S., Sinibaldi, R., and Briganti, G. (2006). Explanation of the Stability of Thermophilic Proteins Based on Unique Micromorphology. *Biophys. J.* 90, 4204–4212.
88. Gromiha, M.M., Pathak, M.C., Saraboji, K., Ortlund, E.A., and Gaucher, E.A. (2013). Hydrophobic Environment is a Key Factor for the Stability of Thermophilic Proteins. *Proteins* 81, 715–721.

89. Kyte, J., and Doolittle, R.F. (1982). A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* *157*, 105–132.
90. Ikai, A. (1980). Thermostability and Alphatic Index of Globular Proteins. *J. Biochem.* *88*, 1895–1898.
91. Kumar, S., Tsai, C.-J., and Nussinov, R. (2000). Factors Enhancing Protein Thermostability. *Prot. Eng. Des. Sel.* *13*, 179–191.
92. Yehorova, D., Alansson, N., Shen, R., Denson, J.M., Robinson, M., Risso, V.A., Molina, N.R., Loria, J.P., Gaucher, E.A., Sanchez-Ruiz, J.M., et al. (2026). Conformational Dynamics and Catalytic Backups in a Hyper-Thermostable Engineered Archaeal Protein Tyrosine Phosphatase *JACS Au* *6*, 59–81.
93. Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O’Meara, M.J., DiMaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* *13*, 3031–3048.
94. Kumar, S., and Nussinov, R. (2002). Relationship Between Ion Pair Geometries and Electrostatic Strength in Proteins. *Biophys. J.* *83*, 1595–1612.
95. Lee, C.-W., Wang, H.-J., Hwang, J.-K., and Tseng, C.-P. (2014). Protein Thermal Stability Enhancement by Designing Salt Bridges: A Combined Computational and Experimental Study. *PLoS One* *9*, e112751.
96. Romero-Romero, S., Costas, M., Silva Manzano, D.-A., Kordes, S., Rojas-Ortega, E., Tapia, C., Guerra, Y., Shanmugaratnam, S., Rodríguez-Romero, A., Baker, D., et al. (2021). The Stability Landscape of *De Novo* TIM Barrels Explored by a Modular Design Approach. *J. Mol. Biol.* *433*, 167153.

97. Cho, Y., Dauparas, J., Tsuboyama, K., Rocklin, G.J., and Ovchinnikov, S. (2025). Stable *De Novo* Protein Design via Joint Conformational Landscape and Sequence Optimization. *Nat. Commun.* *17*, 8.
98. Roe, D.R., and Cheatham, T.E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory. Comput.* *9*, 3084–3095.
99. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comp. Chem.* *32*, 2319–2327.
100. Gowers, R., Linke, M., Barnoud, J., Reddy, T., Melo, M., Seyler, S., Domanski, J., Dotson, D., Buchoux, S., Kenney, I., and Beckstein, O. (2016). MDAAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. 15th Python Sci. Conf.*
101. Rathi, P.C., Fulton, A., Jaeger, K.-E., and Göhlke, H. (2016). Application of Rigidity Theory to the Thermostabilization of Lipase A from *Bacillus subtilis*. *PLoS Comput. Biol.* *12*, e1004754.
102. Vihinen, M. (1987). Relationship of Protein Flexibility to Thermostability. *Prot. Eng.* *1*, 477–480.
103. Macedo-Ribeiro, S., Darimont, B., Sterner, R., and Huber, R. (1996). Small Structural Changes Account for the High Thermostability of 1[4Fe-4S] Ferredoxin from the Hyperthermophilic Bacterium *Thermotoga maritima*. *Structure* *4*, 1291–1301.
104. Závodsky, P., Kardos, J., Svingor, A., and Petsko, G.A. (1998). Adjustment of Conformational Flexibility is a Key Event in the Thermal Adaptation of Proteins. *Proc. Natl. Acad. Sci. USA* *95*, 7406–7411.

105. Jaenicke, R., and Böhm, G. (1998). The Stability of Proteins in Extreme Environments. *Curr. Opin. Struct. Biol.* 8, 738–748.
106. Robinson-Rechavi, M., Alibés, A., and Godzik, A. (2006). Contribution of Electrostatic Interactions, Compactness and Quaternary Structure to Protein Thermostability: Lessons From Structural Genomics of *Thermotoga maritima*. *J. Mol. Biol.* 356, 547–557.
107. Sun, Z., Liu, Q., Qu, G., Feng, Y., and Reetz, M.T. (2019). Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* 2019, 1626–1665.
108. Jaenicke, R. (2000). Do Ultrastable Proteins From Hyperthermophiles Have High or Low Conformational Rigidity? *Proc. Natl. Acad. Sci. USA* 97, 2962–2964.
109. Karshikoff, A., Nilsson, L., and Ladenstein, R. (2015). Rigidity *versus* flexibility: The Dilemma of Understanding Protein Thermal Stability. *FEBS J.* 282, 3899–3917.
110. Hernandez, G., Jenney Jr., F.E., Adams, M.W., and LeMaster, D.M. (2000). Millisecond Time Scale Conformational Flexibility in a Hyperthermophilic Protein at Ambient Temperature. *Proc. Natl. Acad. Sci. USA* 97, 3166–3170.
111. Vittozzi, Y., Petri, L., Chiurazzi, M.J., Liséon, P., Grech Hernández, C., Majee, A., Shankar, N., Burow, M., and Wenkel, S. (2025). Antagonistic and Synergistic Roles of Tomato AFP3 Isoforms in Hormonal Regulation and Development. *bioRxiv*, 10.1101/2025.1106.1118.660365.
112. Risso, V.A., Martinez-Rodriguez, S.M., Candel, A.M., Krüger, D.M., Pantoja-Uceda, D., Ortega-Muñoz, M., Santoyo-Gonzalez, F., Gaucher, E.A., Kamerlin, S.C.L., Bruix, M., et al. (2017). *De Novo* Active Sites for Resurrected Precambrian Enzymes. *Nat. Commun.* 8, 16113.

113. Yehorova, D., Crean, R.M., Kasson, P.M., and Kamerlin, S.C.L. (2024). Friends and Relatives: Insight Into Conformational Regulation from Orthologues and Evolutionary Lineages Using KIF and KIN. *Faraday Discuss.* 252, 341–353.
114. Edgar, R.C. (2004). MUSCLE: A Multiple Sequence Alignment With High Accuracy and High Throughput. *Nucleic Acids Res.* 32, 1792–1797.
115. Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. *Genome Res.* 14, 1188–1190.
116. Tanaka, N., Nonaka, T., Tanabe, T., Yoshimoto, T., Tsuru, D., and Mitsui, Y. (1996). Crystal Structures of the Binary and Ternary Complexes of 7 α -Hydroxysteroid Dehydrogenase from *Escherichia coli*. *Biochemistry* 35, 7715–7730.
117. Andersson, A., Jordan, D., Schneider, G., and Lindqvist, Y. (1997). A Flexible Lid Controls Access to the Active Site in 1,3,8-Trihydroxynaphthalene Reductase. *FEBS Lett.* 400, 173–176.
118. Smilda, T., Reinders, P., and Beintema, J.J. (1998). Modeling Studies of Conformational Changes in the Substrate-Binding Loop of *Drosophila* Alcohol Dehydrogenase. *Biochem. Genet.* 36, 37–49.
119. Chen, Y.-L., Chou, Y.-H., Lee, H.-J., Wang, T.-P., and Hwang, C.-C. (2026). Hydrophobic Interactions of a Substrate With L193 on the Flexible Substrate-Binding Loop Contribute to 3 α -Hydroxysteroid Dehydrogenase/Carbonyl Reductase Catalytic Efficiency. *Protein Sci.* 35, e70555.
120. Pailozian, K., Kohout, P., Damborsky, J., and Mazurenko, S. (2026). TmProt 1.0.
121. Kallberg, Y., Oppermann, U., Jörnvall, H., and Persson, B. (2022). Short-Chain Dehydrogenases/Reductases (SDRs). *Eur. J. Biochem.* 269, 4409–4417.

122. Hon, J., Borko, S., Stourac, J., Prokop, Z., Zendulka, J., Bednar, D., Martinek, T., and Damborsky, J. (2020). EnzymeMiner: Automated Mining of Soluble Enzymes With Diverse Structures, Catalytic Properties and Stabilities. *Nucleic Acids Res.* *48*, W104–W109.
123. Lin, H., and Chen, W. (2011). Prediction of Thermophilic Proteins Using Feature Selection Technique. *J. Microbiol. Methods* *84*, 67–70.
124. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J.S., and Roitberg, A.E. (2020). TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neutral Network Potentials. *J. Chem. Inf. Mod.* *60*, 3408–3415.
125. Pickering, I., Xue, J., Huddleston, K., Terrel, N., and Roitberg, A.E. (2025). TorchANI 2.0: An Extensible, High-Performance Library for the Design, Training, and Use of NN-IPs. *J. Chem. Inf. Model.* *65*, 11656–11671.
126. Pickering, I., Semelak, J.A., Xue, J., and Roitberg, A.E. (2025). TorchANI-Amber: Bridging Neural Network Potentials and Classical Biomolecular Simulations. *J. Phys. Chem. B* *129*, 11927–11938.
127. Wlodawer, A., Dauter, Z., Rubach, P., Minor, W., Loch, J.I., Brzezinski, D., Gilski, M., and Jaskolski, M. (2024). Waterless Structures in the Protein Data Bank. *IUCrJ* *11*, 966–976.
128. Polichuk, D.R., Zhang, Y., Reed, D.W., Schmidt, J.F., and Covello, P.S. (2010). A Glandular Trichome-Specific Monoterpene Alcohol Dehydrogenase from *Artemisia annua*. *Phytochemistry* *71*, 1264–1269.

129. Tian, N., Tang, Y., Xxiong, S., Tian, D., Chen, Y., Wu, D., Liu, Z., and Liu, S. (2015). Molecular Cloning and Functional Identification of a Novel Borneol Dehydrogenase from *Artemisia annua* L. *Ind. Crop. Prod.* *77*, 190–195.
130. Ma, R., Su, P., Jin, B., Guo, J., Tian, M., Mao, L., Tang, J., Chen, T., Lai, C., Zeng, W., et al. (2021). Molecular Cloning and Functional Identification of a High-Efficiency Borneol Dehydrogenase from *Cinnamomum camphora* (L.) Presl. *Plant Phys. Biochem.* *158*, 363–371.
131. Hu, X., Deng, H., Bai, Y., Fan, T.-P., Zheng, X., and Cai, Y. (2022). Heterologous Expression and Characterization of a Borneol Dehydrogenase from *Arabidopsis lyrata* and its Application for the Enzymatic Resolution of *rac*-Camphor. *Mol. Catal.* *530*, 112572.
132. Livada, J., Vargas, A.M., Martinez, C.A., and Lewis, R.D. (2023). Ancestral Sequence Reconstruction Enhances Gene Mining Efforts for Industrial Ene Reductases by Expanding Enzyme Panels with Thermostable Catalysts. *ACS Catal.* *13*, 2576–2585.
133. Randall, R.N., Radford, C.E., Roof, K.A., Natarajan, D.K., and Gaucher, E.A. (2016). An Experimental Phylogeny to Benchmark Ancestral Sequence Reconstruction. *Nat. Comm.* *7*, 12847.
134. Muellers, S.N., Allen, K.N., and Whitty, A. (2023). MEnTaT: A Machine-Learning Approach for the Identification of Mutations to Increase Protein Stability. *Proc. Natl. Acad. Sci. USA* *120*, e2309884120.
135. Rodella, C., Lazaridi, S., and Lemmin, T. (2024). TemBERTure: Advancing Protein Thermostability Prediction With Deep Learning and Attention Mechanisms. *Bioinformat. Adv.* *4*, vbae103.

136. Yu, H., Deng, H., He, J., Keasling, J.D., and Luo, X. (2023). UniKP: A Unified Framework for the Prediction of Enzyme Kinetic Parameters. *Nat. Commun.* *14*, 8211.
137. Gollub, M.G., Backes, T., Kaltenbach, H.-M., and Stelling, J. (2024). ENKIE: A Package for Predicting Enzyme Kinetic Parameter Values and Their Uncertainties. *Bioinformatics* *40*, btae652.
138. Malli, A., Vasyutyn, D., and Kim, J.R. (2025). Advances in Machine Learning Models for Predicting Enzyme Kinetic Parameters. *J. Chem. Inf. Mod.* *66*, 42–60.
139. Boorla, V.H., and Maranas, C.D. (2025). CatPred: A Comprehensive Framework for Deep Learning *In Vitro* Enzyme Kinetic Parameters. *Nat. Commun.* *16*, 2072.
140. Qiu, S., Saeed, H., Leonard, W., Li, F., and Yang, A. (2026). Machine Learning for Enzyme Catalytic Activity: Current Progress and Future Horizons. *Brief. Bioinformat.* *27*, bbag002.
141. Kroll, A., Ranjan, S., Engqvist, M.K.M., and Lercher, M.J. (2023). A General Model to Predict Small Molecule Substrates of Enzymes Based on Machine and Deep Learning. *Nat. Commun.* *14*, 2787.
142. Li, Z.-L., Pei, S., Chen, Z., Huang, T.-Y., Wang, X.-D., Shen, L., Chen, X., Wang, Q.-Q., Wang, D.-X., and Ao, Y.-F. (2024). Machine Learning-Assisted Amidase-Catalytic Enantioselectivity Prediction and Rational Design of Variants for Improving Enantioselectivity. *Nat. Commun.* *15*, 8778.
143. Cui, H., Su, Y., Dean, T.J., Yu, T., Zhang, Z., Peng, J., Shukla, D., and Zhao, H. (2025). Enzyme Specificity Prediction Using Cross-Attention Graph Neural Networks. *Nature* *647*, 639–647.

144. Listov, D., Vos, E., Hoffka, G., Hoch, S.Y., Berg, A., Hamer-Rogotner, S., Dym, O., Kamerlin, S.C.L., and Fleishman, S.J. (2025). High-Efficiency Kemp Eliminases by Complete Computational Design. *Nature* *643*, 1421–1427.
145. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER Web Server: 2018 Update. *Nucleic Acids Res.* *46*, W200–W204.
146. Katoh, K., Kuma, K.-I., Toh, H., and Miyata, T. (2005). MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment. *Nucleic Acids Res.* *33*, 511–518.
147. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* *30*, 772–780.
148. Chai Discovery team, Boitreaud, J., Dent, J., McPartlon, M., Meier, J., Reis, V., Rogozhonikov, A., and Wu, K. (2024). Chai-1: Decoding the Molecular Interactions of Life. *bioRxiv*, DOI: 10.1101/2024.1110.1110.615955.
149. Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., and Onufriev, A. (2005). H++: A Server for Estimating pK_as and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* *33*, W368–W371.
150. Anandakrishnan, R., Aguilar, B., and Onufriev, A.V. (2012). H++ 3.0: Automatic pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* *40*, W537–W541.
151. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: An Open Chemical Toolbox. *J. Cheminform.* *3*, 33.
152. Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguet, L., Huan, H., Migués, A.N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020). ff19SB: Amino-Acid-Specific

- Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* *16*, 528–552.
153. Woods, R.J., and Chappelle, R. (2000). Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates. *J. Mol. Struct.* *1-3*, 149–156.
 154. Wang, J., Wang, W., Kollman, P.A., and Case, D.A. (2006). Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* *25*, 247–260.
 155. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., et al. (2016). Gaussian 16, Revision A.03 (Gaussian Inc.).
 156. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and Testing of a General Amber Force Field. *J. Comput. Chem.* *25*, 1157–1174.
 157. Case, D.A., Aktulga, H.M., Belfon, K., Ben-Shalom, I.Y., Berryman, J.T., Brozell, S.R., Carvahol, F.S., Cerutti, D.S., Cheatham III, T.E., Cisneros, G.A., et al. (2025). Amber 2025 (University of California).
 158. Izadi, S., Anandakrishnan, R., and Onufriev, A. (2014). Building Water Molecules: A Different Approach. *5*, 3863–3871.
 159. Davidchack, R.L., Handel, R., and Tretyakov, M.V. (2009). Langevin Thermostat for Rigid Body Dynamics. *J. Chem. Phys.* *130*, 234101.
 160. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* *81*, 3684–3690.

161. Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* *98*, 10089–10092.
162. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* *23*, 327–341.
163. Smith, J.S., Isayev, O., and Roitberg, A.E. (2017). ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* *8*, 3192–3203.
164. Smith, J.S., Isayev, O., and Roitberg, A.E. (2017). ANI-1, A Data set of 20 Million Calculated Off-Equilibrium Conformations for Organic Molecules. *Sci. Data* *4*, 170193.
165. Semelak, J.A., Pickering, I., Huddleston, K., Olmos, J., Grassano, J.S., Clemente, C.M., Drusin, S.I., Marti, M., Gonzalez Lebrero, M.C., Roitberg, A.E., and Estrin, D.A. (2025). Advancing Multiscale Molecular Modeling with Machine Learning-Derived Electrostatics. *J. Chem. Theory Comput.* *21*, 5194–5207.
166. Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchinson III, C.A., and Smith, H.O. (2009). Enzymatic Assembly of DNA Molecules Up to Several Hundred Kilobases. *Nat. Methods* *6*, 343–345.
167. Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protocol.* *2*, 2212–2221.

Graphical Abstract

