Computational stabilization of a non-heme iron enzyme enables efficient evolution of new function

Authors

Brianne R. King, Kiera H. Sumida, Jessica L. Caruso, David Baker, and Jesse G. Zalatan*

Corresponding Author

Jesse G. Zalatan − Department of Chemistry, University of Washington, Seattle, Washington 98195, United States; Email: zalatan@uw.edu

Authors

Brianne R. King − Department of Chemistry, University of Washington, Seattle, Washington 98195, United States

Kiera H. Sumida − Department of Chemistry and Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States

Jessica L. Caruso − Department of Chemistry, University of Washington, Seattle, Washington 98195, United States

David Baker − Institute for Protein Design, Department of Biochemistry, and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, United States

Abstract

Directed evolution has emerged as a powerful tool for engineering new biocatalysts. However, introducing new catalytic residues can be destabilizing, and it is generally beneficial to start with a stable enzyme parent. Here we show that the deep learning tool ProteinMPNN can be used to redesign an Fe(II)/αKG superfamily enzyme for greater stability, solubility, and expression while retaining both native activity and an industrially-relevant non-native function. We performed site-saturation mutagenesis with both the wild type and stabilized design variant and screened for activity increases in a non-native C-H hydroxylation reaction. We observed substantially larger increases in non-native activity for variants obtained from the stabilized scaffold compared to those from the wild-type enzyme. Deep learning tools like ProteinMPNN are user-friendly and widely-accessible, and relatively straightforward structural criteria were sufficient to obtain stabilized variants while preserving catalytic function. Our work suggests that stabilization by computational sequence redesign could be routinely implemented as a first step in directed evolution campaigns for novel biocatalysts.

Introduction

Directed evolution is a powerful method to generate enzymes for new chemical transformations.[1,2] However, catalytic functional groups often have destabilizing effects on protein structure, and altering active site groups for new reactions can lead to unstable, non-functional proteins.[3–11] Initiating a directed evolution campaign from a stabilized variant can be an effective way to overcome this problem.[12,13] Typically, a starting point for directed evolution is obtained by screening a library of candidate enzymes for a desired promiscuous activity. If a thermostable homolog with similar catalytic properties can be identified, it can then be used as a starting point for evolution of the desired function.[14,15] Alternatively, there are a variety of strategies to produce stable variants using directed evolution,[16,17] ancestral reconstruction,[14] protein recombination,[18,19] or computational engineering.[20–22] These methods are often time and resource intensive, highlighting the need for simple and accessible alternatives.

An important class of enzymes where stabilization would be useful is the non-heme iron(II) α-ketoglutarate-dependent oxygenase (Fe(II)/αKG) superfamily. These enzymes have emerged as a rich source of potential new biocatalysts. Fe(II)/αKG enzymes can perform asymmetric C-H oxyfunctionalization reactions on small molecule substrates, which could allow expedient and sustainable diversification of simple building blocks to a range of complex polyfunctional compounds.[23–26] The advantages offered by this enzyme family include a high degree of chemical flexibility in the iron-containing active site due to multiple open coordination sites, the utilization of benign molecular oxygen as an oxidant, and use of the inexpensive and readily available co-factor αKG. However, Fe(II)/αKGs can be relatively unstable,[27,28] which may limit their practical applications in organic synthesis.

The recent use of an Fe(II)/αKG in an industrial-scale drug biosynthesis pathway highlights both the potential advantages and drawbacks of this family for biocatalysis. An engineered Fe(II)/αKG was used to catalyze an enantioselective C-H hydroxylation to produce a key intermediate for the anti-cancer drug belzutifan.[28] The reaction could be performed on kilogram-scale and bypassed five steps of the pre-existing chemical synthesis route. Notably, this effort required an extensive, large-scale directed evolution campaign. Furthermore, early rounds of screening yielded stabilizing mutations before significant improvements in turnover could be obtained in later rounds. These findings highlight the potential utility of Fe(II)/αKG for practical, industrial-scale green chemistry but also the importance of enzyme stability in the evolution of new function.

Recent applications of deep learning to protein design have provided new and relatively straightforward methods to stabilize protein scaffolds,[22,29,30] and there is broad interest in applying these approaches to directed evolution.[31] Here we demonstrate that the deep learning tool ProteinMPNN[29,30] enables more efficient optimization of a synthetically-relevant, non-native C-H hydroxylation reaction in an Fe(II)/αKG family member. A critical step was restricting the redesign from modifying active site and adjacent residues, which would otherwise be readily mutated to stabilize the enzyme. With a stabilized starting point for site-saturation mutagenesis, we observed substantially larger increases in non-native activity compared to the same mutations in the wild-type parent enzyme. We suggest that this designed stabilization approach should be routinely used in future directed evolution campaigns with the Fe(II)/αKG superfamily and will likely be effective in a broad range of other enzyme families.

Results and Discussion

Fe(II)/αKGs with promiscuous activity for C-H hydroxylation of free carboxylate substrates

Within the Fe(II)/αKG enzyme superfamily, free amino acid hydroxylases are attractive candidates for engineering new reactions.[23–26] Because Fe(II)/αKG amino acid hydroxylases already have catalytic machinery to interact with amine and carboxylate functional groups in amino acids, we hypothesized that they might have promiscuous activity for substrates containing only an amine or only a carboxylate. These molecules are important feedstocks for early-stage oxyfunctionalization reactions in multi-step syntheses. Selectivity for these reactions has been historically difficult to achieve with traditional transition metal catalysis, and a biocatalytic process could offer improved regio- and stereoselectivity.[32,33]

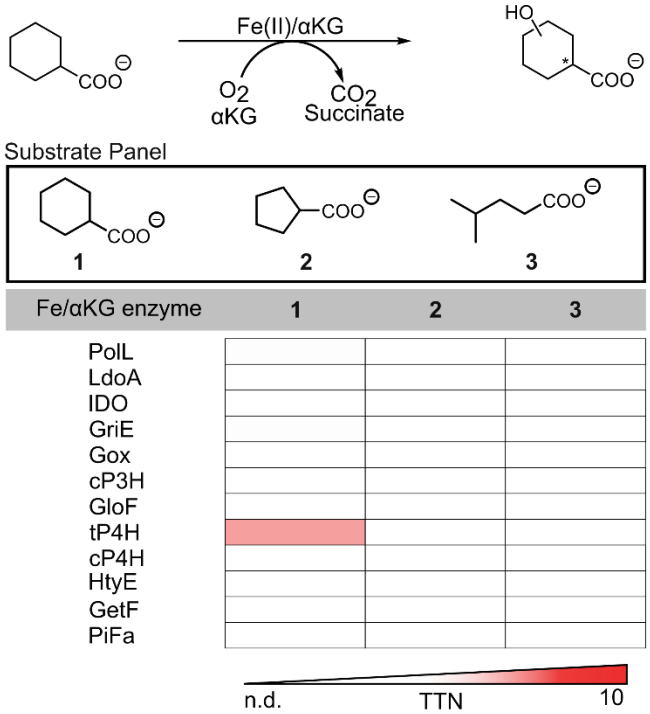

Figure 1. Initial whole-cell reaction screen data with a panel of Fe(II)/αKG amino acid hydroxylases and free acid substrate analogues. Reactions were performed in whole cell from 50 mL expression where whole cell volume was 1/20th the expression volume. Reactions were carried out in MOPS (pH 7.0, 50 mM) with 20 mM substrate, 60 mM αKG (as disodium salt), 1 mM ferrous ammonium sulfate, and 1 mM L-ascorbic acid.

We initially screened a panel of 12 Fe(II)/αKG amino acid hydroxylases for the ability to hydroxylate free carboxylates (Figure 1). We chose a set of candidate carboxylate substrates (**1-3**) that are structurally analogous to the native amino acid substrates pipecolic acid, proline, and leucine. We used whole-cell biocatalysis and liquid chromatography-mass spectrometry (LCMS) to detect products. We confirmed that the native amino acid reaction products are detectable with all 12 members of the enzyme panel (Table S7). We then screened for promiscuous activity with carboxylates, and observed that one enzyme, tP4H,[34] has detectable activity with substrate **1** (Figure 1). The total turnover number (TTN) with this substrate was ~5 after 24 hr incubation with 10 μM enzyme, ~130-fold lower than the TTN for the corresponding native amino acid substrate. The reaction of tP4H with substrate **1** gives the *trans* product with an *d.r.* of 4:1. tP4H produces exclusively *trans* product with its native substrate pipecolic acid,[34] suggesting that the free carboxylate substrate **1** and the native substrate are positioned similarly in the enzyme active site with respect to the iron center. To confirm that tP4H and not a contaminating enzyme was responsible for the observed non-native activity, we mutated active site residues that are involved in Fe(II), substrate, or αKG binding. Because tP4H does not have an experimental structure, we identified these active site residues using an Alphafold2[35] model (Figure S6) and comparisons to structures of the highly homologous Fe(II)/αKG enzyme GriE.[36,37] In all cases active site mutations produced activity decreases for the non-native substrate **1** (Figure 2). We also observed increased product yield with increasing wild-type tP4H concentration (Figure 2). Together these results confirm that tP4H is responsible for the non-native reaction.

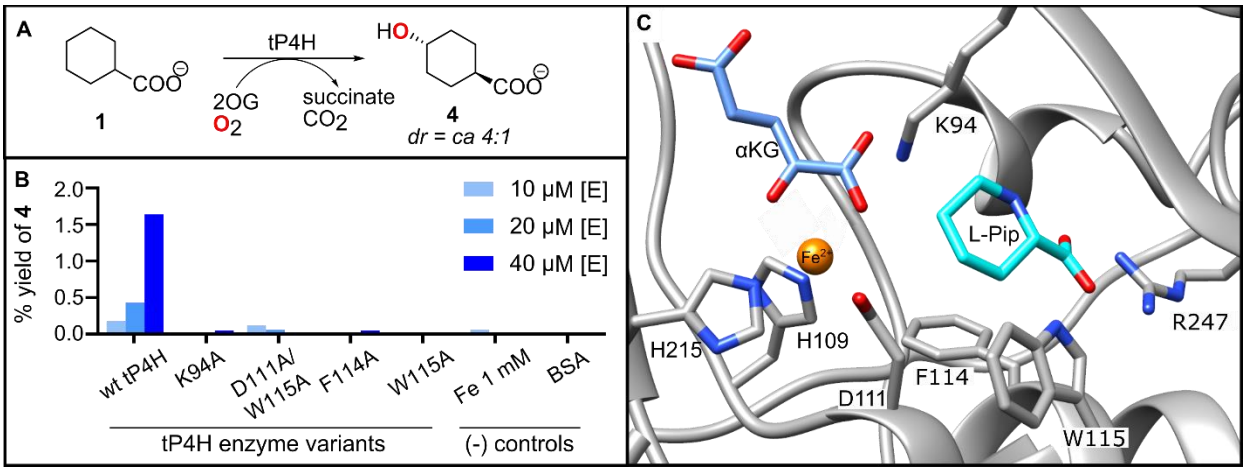

Figure 2. Validation of tP4H activity with free acid **1.** (A) Reaction of tP4H with substrate **1** to form *trans*-4-hydroxycyclohexane carboxylic acid **4**. (B) Yield of **4** after reaction of **1** with tP4H variants, as well as negative control reaction with Fe(II) and bovine serum albumin (BSA). The Fe 1 mM control was run in the absence of added enzyme. Purified enzyme concentration was varied between 10-40 μM with 20 mM **1**, 40 mM αKG, 1 mM ferrous ammonium sulfate, and 1 mM L-ascorbic acid in MES buffer (50 mM, pH 6.8). Reactions were carried out for 24 hours at 25 °C and quantified with analytical LC-MS. (C) Structural model of the tP4H showing key active site residues. Fe(II), αKG, and L-Pipecolic acid were modeled in using Chimera.[38]

Directed evolution of carboxylate C-H hydroxylation from the wild-type tP4H parent enzyme

To improve carboxylate hydroxylase activity in tP4H, we conducted a directed evolution campaign by varying first- and second-shell substrate binding residues identified in the active site from the Alphafold2 structural model. We defined the first-shell as any residues that contact the amino acid substrate, based on comparisons to the structure of ligand-bound GriE.[37] We defined the second shell as any residues that make contacts with first shell residues. We used the 22c-trick method for single site-saturation mutagenesis at each target position.[39] For the first screening round, we chose three tP4H active site residues based on their potential role in substrate specificity: H58, F114, and L174. Based on our tP4H structural model (Figure S6B), H58 likely contacts the amine of native amino acid substrates and is presumably not needed or detrimental for carboxylate substrates that lack an amine. F114 likely provides a substrate hydrophobic contact, and L174 is part of a loop that could affect substrate binding. We screened whole cell biocatalysis reactions in

96 well plates for improved TTN and 80% *trans* selectivity in reactions with substrate **1**. Based on production of hydroxylated product **4**, the top 5% of mutants were chosen for validation with purified enzymes. We obtained several variants with modest activity improvements, and the best performer was mutant H58L with a TTN of 7 (Figure 3). In a second round starting from H58L, we rescreened mutants at F114 and L174 and screened an additional 14 first and second shell residues (Table S4). We identified the improved variant H58L/W170Q with a TTN of 15. In a third round starting from H58L/W170Q, we screened 9 residues that showed activity increases in previous rounds and identified the improved variant H58L/W170Q/E118K with a TTN of 31 (Figure 3/Table S4). Overall, after three rounds of directed evolution for improved carboxylate hydroxylase activity we obtained an 6-fold improvement in TTN and maintained >80% selectivity for the *trans* reaction product.

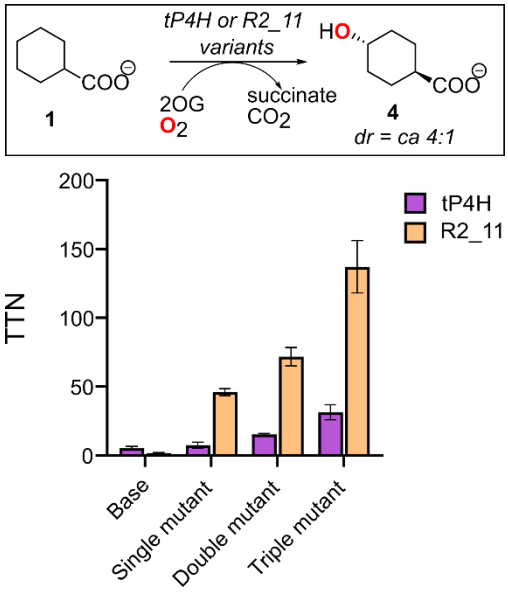

Figure 3. Directed evolution of tP4H and stabilized variant R2_11. Single, double, and triple mutants for the tP4H trajectory are H58L, H58L/W170Q, H58L/W170Q/E118K respectively. Single, double, and triple mutants for the R2_11 trajectory are H58F, H58F/L174G, and H58F/L174G/V57H, respectively. Reactions were carried out for 24 hr at 25 °C using purified enzyme (10-20 μM) in MES buffer (50 mM, pH 6.8), with 20 mM cyclohexane carboxylic acid **1**, 40 mM αKG, 1 mM ferrous ammonium sulfate, and 1 mM ascorbic acid. Concentration of **4** in quenched reaction samples was quantified by analytical LC-MS analysis.

Stabilization of tP4H with ProteinMPNN

Throughout the directed evolution process described above, attempts to express and purify screen hits for validation often produced insoluble enzyme. We also found that the parent wild-type tP4H enzyme loses activity with time (Figure 4). These observations are consistent with prior reports on tP4H behavior.[34] The marginal stability of tP4H and its selected variants suggests that stability might be a limiting factor in our directed evolution campaign. It is possible for enzyme stability to improve during directed evolution, whether by selecting more stable variants during each round or by chance. For example, the evolved Fe(II)/αKG PsEFE had slight stability improvements compared to wild-type, despite the researchers not selecting for improved stability.[40] In another case, the Fe(II)/αKG UbP4H was successfully screened for improved stability after initial screening rounds destabilized the enzyme.[27] However, the benefits of starting with a highly stable enzyme for directed evolution are well-established.[12–14] A stabilized protein scaffold could potentially increase the population of active, properly-folded protein or provide access to other mutants that otherwise do not fold.

We used the deep learning tool ProteinMPNN[29,30] to obtain stabilized variants of tP4H. We first used ProteinMPNN to optimize the entire tP4H sequence. Unsurprisingly, we found that the predicted sequences eliminated key active site residues, which is likely to disrupt enzymatic activity (Figure S8). This behavior is consistent with the well-established propensity for catalytic active site residues to be destabilizing.[3–10] To preserve catalytic function, we fixed the active site residues for all subsequent designs. We defined the active site as any residues that contact the amino acid substrate, Fe(II), or the αKG cofactor, based on our Alphafold2 model (Figure S6) and the structure of ligand-bound GriE.[37] Because other residues throughout the protein could also be important, we also tested four additional strategies using either sequence conservation or distance

metrics. To identify important conserved residues, we constructed a multiple sequence alignment (MSA) and selected tP4H residues conserved in at least 35%, 70%, or 95% of sequences (Methods and Table S4). Alternatively, we fixed any residues with side chains within a 10 Å sphere from the substrate binding pocket. Using these five starting points (fix active site only, active site + 35%/70%/95% conservation, active site + 10 Å sphere), we generated 48 ProteinMPNN sequences per method and selected 4 each (20 total) for activity screens. Selection was based on calculated top ranked Cα-RMSD values matched to the input tP4H structure. We obtained only one variant that had any detectable activity, with catalytic efficiency ~35-fold lower than wild-type tP4H (Figure S7B). This variant was designed from the sequence where >35% conserved residues are fixed, which constrains more residues than the >70% or >95% cutoffs. This result suggests that even weakly conserved residues may need to be fixed to maintain activity. Further analysis of the variant with detectable activity revealed a ~3.5-fold increase in the K_M for the αKG cofactor (Figure S7A). We identified two residues in proximity to αKG, L228 and V230, that were mutated in the redesigned sequence. These sequence changes may have contributed to improved stability at the expense of cofactor binding and positioning, leading to the decrease in activity. Notably, L228 and V230 were fixed in the designs generated from fixed active site + 10 Å sphere, but none of these designs had detectable activity. Taken together, these findings suggest that additional criteria will be needed to identify critical functional residues that should be fixed prior to sequence redesign.

To generate stabilized variants that maintain catalytic activity, we performed another round of ProteinMPNN sequence redesign with three new strategies to fix important residues. In each case, we fixed the active site as defined above plus residues L228 and V230. For the first approach, we fixed all residues at tP4H positions conserved in 35% of the MSA. We chose this cutoff because

it was the only one from our initial set that produced a stabilized variant with any detectable activity, and we expected that fixing L228 or V230 could further improve these designs. For the second and third approaches, we identified highly conserved positions regardless of whether the wild-type tP4H residue is the most highly conserved amino acid. These strategies were based on previous work suggesting that more stringent constraints are necessary to maintain activity in ProteinMPNN redesign.[30] Each position was ranked based on the % conservation of the most frequent amino acid present in the MSA, and positions in the top 50% or 70% were fixed as the wild-type tP4H residue. Using these three strategies, we selected 32 designs each of 48 generated for a total of 96 sequences (Supplementary spreadsheet – ProteinMPNN sequences_metrics). Of these designs, 69 expressed detectable quantities of protein by SDS-PAGE and 11 had detectable activity above background for the native substrate. For the active enzymes we proceeded to measure thermostability and kinetic parameters for the native pipecolic acid substrate and the promiscuous carboxylate substrate **1**. The variant with the highest k_{cat} for pipecolic acid was R2_11 (Table S7), with a k_{cat} of 0.16 s^{-1} compared to 0.17 s^{-1} for wild-type. R2_11 was designed from the method where the top 70% ranked conserved residues were fixed. R2_11 has modestly slower (~3-fold) non-native carboxylate hydroxylase activity compared to wild-type tP4H and exhibits a 20 °C increase in thermostability by temperature dependent circular dichroism (Fig 4). When activity is measured as a function of time, R2_11 maintains activity over a timescale of days, which is a substantial improvement compared to wild-type tP4H (Fig 4).

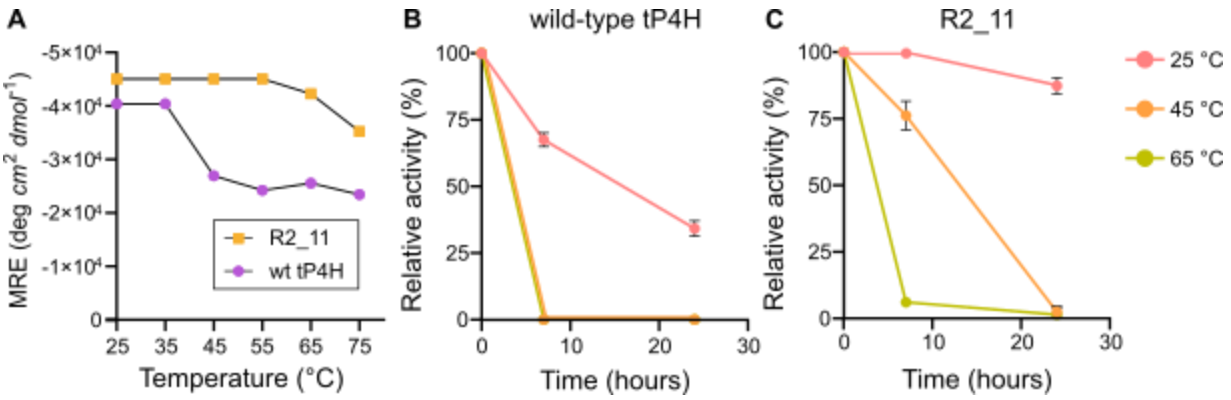

Figure 4. Stability and activity of wild-type tP4H and ProteinMPNN design R2_11. (A) Temperature-dependent circular dichroism (CD) of wild-type tP4H and R2_11. (B) Activity-stability analysis of wild-type tP4H. (C) Activity-stability analysis of R2_11. Relative activity was determined using PBP assay described in Supplementary Information.

Directed evolution of carboxylate C-H hydroxylation from ProteinMPNN-stabilized tP4H

To optimize carboxylate hydroxylase activity in ProteinMPNN-stabilized tP4H, we conducted directed evolution using a similar strategy to our approach with wild-type tP4H, with minor modifications. In the first round of site saturation mutagenesis, we started with a larger pool of 19 first- and second-shell residues including the three sites from prior round one (H58, F114, and L174) and the 16 additional sites from prior round 2. As before, we screened whole cell biocatalysis reactions for improved TTN and >80% *trans* selectivity with carboxylic acid substrate **1**. The top hit was H58F, which is the same position but a different mutant than the previous round one winner, H58L. R2_11_H58F displayed a 27-fold increase in TTN relative to the parent R2_11 (Figure 3). This effect is substantially bigger than the <2-fold improvement obtained with H58L relative to wild-type tP4H. Notably, the 27-fold increase in a single round from stabilized R2_11 was already larger than the total 6-fold improvement from three rounds of directed evolution from wild-type tP4H. Given the strong performance of the H58F mutant, we also evaluated its effect in the wild-type tP4H background and observed a small, <2-fold increase in TTN, similar to the effect of H58L on wild-type tP4H (Figure S9B). Thus, the strong, 27-fold improvement with the H58F mutant depends on the context of the stabilized R2_11 backbone.

In a second round of screening from R2_11_H58F, we selected the 18 residues that were screened in prior round two from wild-type tP4H. We retained this large pool of residues to ensure a direct comparison to the tP4H directed evolution workflow. This round identified improved variant L174G. R2_11_H58F/L174G has a TTN of 72. This TTN is a 1.6-fold improvement from parent R2_11_H58F and outstrips any variant obtained from the wild-type tP4H backbone (Figure 3). In a third round of screening from R2_11_H58F/L174G, we selected 9 residues that were screened in prior round 3 from wild-type tP4H. This round identified the improved variant V57H with a TTN of 138, a 1.7-fold improvement from the previous round.

Overall, the ProteinMPNN-stabilized tP4H directed evolution campaign produced an 80-fold improvement in TTN from the base R2_11 redesign, compared to a modest 6-fold improvement in the wild-type tP4H evolutionary trajectory. Although the R2_11 parent starts ~3-fold slower than wild-type tP4H, the much larger improvement over three rounds of directed evolution produced an R2_11 triple mutant with a 4.5-fold higher TTN than the triple mutant obtained from wild-type tP4H (Figure 3).

In addition to a more efficient directed evolution trajectory, the R2_11 triple mutant maintains high stability relative to the wild-type tP4H and associated variants from directed evolution. (Figure 5A, Figure S10). Higher stability allows reactions to be run more efficiently, both at higher temperatures and for less time. For example, after 6 hours at 35 °C, the R2_11 triple mutant reaches a mean TTN of 142 for the non-native reaction with carboxylate **1** to form product **4** with 4:1 selectivity for the *trans* reaction product. The TTN after 6 hours at 35 °C is comparable to the TTN after 24 hours at 25 °C. In contrast, the tP4H triple mutant shows a slight decrease in TTN at 35 °C, likely due to enzyme instability at higher temperatures (Figure 5B). The stability profile of the R2_11 triple mutant suggests that this enzyme will be more robust towards further

engineering compared to the tP4H triple mutant. Future engineering efforts with the R2_11 mutant could include improvements to key reaction metrics like turnover, selectivity, and increased substrate scope.

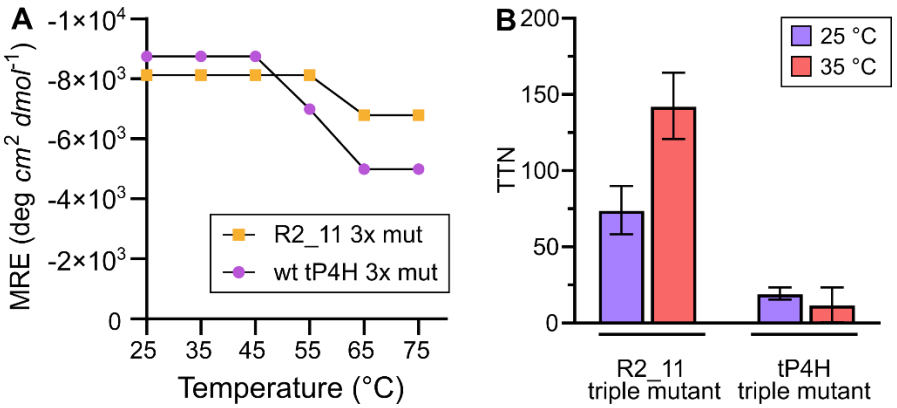

Figure 5 A) Temperature dependent CD of the R2_11 and tP4H triple mutants. B) TTN for formation of **4** (*d.r.* 4:1) with the R2_11 and tP4H triple mutants at two different temperatures. Reactions were carried out for 6 hrs. at 25 °C and 35 °C using purified enzyme (15 μM) in MES buffer (50 mM, pH 6.8), with 20 mM cyclohexane carboxylic acid **1**, 40 mM αKG, 1 mM ferrous ammonium sulfate, and 1 mM ascorbic acid.

Conclusions

Directed evolution is a powerful tool to engineer enzymes for new-to-nature reactions. However, many enzyme starting points for evolution may lack the stability required to reach user-defined optimum fitness after multiple rounds of mutagenesis. Here we show that the deep learning tool ProteinMPNN can be used to stabilize the Fe(II)/αKG enzyme superfamily member tP4H, which exhibits novel reactivity towards remote C-H hydroxylation of free carboxylic acids. We directly compared evolutionary trajectories of the wild-type enzyme with a stabilized variant and demonstrated superior performance of the stabilized redesign variant. Consistent with previous results using ProteinMPNN, our top design was identified by using the most conservative of our chosen methods for fixing residues during sequence redesign. Future work will determine if this design method is generalizable to other enzymes and enzyme families for optimized directed evolution. User-friendly deep learning tools like ProteinMPNN and related tools like MutCompute

are rapidly emerging,[22,29,30] and our work suggests that these tools should be routinely incorporated into enzyme engineering workflows to efficiently optimize catalytic fitness for new biocatalysts.[31]

Experimental Procedures

See Supplementary Materials. The Supplementary Information contains all experimental procedures and materials used in this work. An additional Supplementary Spreadsheet supplies the protein and DNA sequences used in this work, as well as ProteinMPNN sequences and metrics.

Acknowledgements

We would like to thank Dr. Wolfgang Hüttel at the University of Freiburg for the donation of the wild-type tP4H expression vector and for his advice on tP4H reactions in various formats. We also thank Jonathan Zhang and Susanna Vazquez Torres for their early contributions to Fe(II)/αKG reaction screening, and Dr. Martin Sadilek in University of Washington Mass Spectrometry Facility for his continued support and helpful advice in analytical method development. This work was supported by U.S. National Institutes of Health grants T32 GM008268 (B.R.K., J.L.C.) and R35 GM124773 (J.G.Z.), and by the Open Philanthropy Project Improving Protein Design Fund (K.H.S., D.B.).

Competing Interests

The authors declare that they have no competing interests.

References

(1) Renata, H.; Wang, Z. J.; Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chem. Int. Ed.* **2015**, *54* (11), 3351–3367. https://doi.org/10.1002/anie.201409470.

(2) Zeymer, C.; Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018**, *87* (1), 1–27. https://doi.org/10.1146/annurev-biochem-062917-012034.

(3) Meiering, E. M.; Serrano, L.; Fersht, A. R. Effect of Active Site Residues in Barnase on Activity and Stability. *J. Mol. Biol.* **1992**, *225* (3), 585–589. https://doi.org/10.1016/0022-2836(92)90387-y.

(4) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. A Relationship between Protein Stability and Protein Function. *Proc. Natl. Acad. Sci.* **1995**, *92* (2), 452–456. https://doi.org/10.1073/pnas.92.2.452.

(5) Giver, L.; Gershenson, A.; Freskgard, P.-O.; Arnold, F. H. Directed Evolution of a Thermostable Esterase. *Proc. Natl. Acad. Sci.* **1998**, *95* (22), 12809–12813. https://doi.org/10.1073/pnas.95.22.12809.

(6) Beadle, B. M.; Shoichet, B. K. Structural Bases of Stability–Function Tradeoffs in Enzymes. *J. Mol. Biol.* **2002**, *321* (2), 285–296. https://doi.org/10.1016/s0022-2836(02)00599-5.

(7) Nagatani, R. A.; Gonzalez, A.; Shoichet, B. K.; Brinen, L. S.; Babbitt, P. C. Stability for Function Trade-Offs in the Enolase Superfamily "Catalytic Module". *Biochemistry* **2007**, *46* (23), 6688–6695. https://doi.org/10.1021/bi700507d.

(8) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLoS Comput. Biol.* **2008**, *4* (2), e1000002. https://doi.org/10.1371/journal.pcbi.1000002.

(9) Tokuriki, N.; Jackson, C. J.; Afriat-Jurnou, L.; Wyganowski, K. T.; Tang, R.; Tawfik, D. S. Diminishing Returns and Tradeoffs Constrain the Laboratory Optimization of an Enzyme. *Nat. Commun.* **2012**, *3* (1), 1257. https://doi.org/10.1038/ncomms2246.

(10) Goldsmith, M.; Tawfik, D. S. Enzyme Engineering: Reaching the Maximal Catalytic Efficiency Peak. *Curr. Opin. Struct. Biol.* **2017**, *47*, 140–150. https://doi.org/10.1016/j.sbi.2017.09.002.

(11) Tsuboyama, K.; Dauparas, J.; Chen, J.; Laine, E.; Behbahani, Y. M.; Weinstein, J. J.; Mangan, N. M.; Ovchinnikov, S.; Rocklin, G. J. Mega-Scale Experimental Analysis of Protein Folding Stability in Biology and Design. *Nature* **2023**, *620* (7973), 434–444. https://doi.org/10.1038/s41586-023-06328-6.

(12) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci.* **2006**, *103* (15), 5869–5874. https://doi.org/10.1073/pnas.0510098103.

(13) Gumulya, Y.; Baek, J.-M.; Wun, S.-J.; Thomson, R. E. S.; Harris, K. L.; Hunter, D. J. B.; Behrendorff, J. B. Y. H.; Kulig, J.; Zheng, S.; Wu, X.; Wu, B.; Stok, J. E.; Voss, J. J. D.; Schenk, G.; Jurva, U.; Andersson, S.; Isin, E. M.; Bodén, M.; Guddat, L.; Gillam, E. M. J. Engineering Highly Functional Thermostable Proteins Using Ancestral Sequence Reconstruction. *Nat. Catal.* **2018**, *1* (11), 878–888. https://doi.org/10.1038/s41929-018-0159-5.

(14) Trudeau, D. L.; Tawfik, D. S. Protein Engineers Turned Evolutionists—the Quest for the Optimal Starting Point. *Curr. Opin. Biotechnol.* **2019**, *60*, 46–52. https://doi.org/10.1016/j.copbio.2018.12.002.

(15) Besenmatter, W.; Kast, P.; Hilvert, D. Relative Tolerance of Mesostable and Thermostable Protein Homologs to Extensive Mutation. *Proteins: Struct., Funct., Bioinform.* **2007**, *66* (2), 500–506. https://doi.org/10.1002/prot.21227.

(16) Socha, R. D.; Tokuriki, N. Modulating Protein Stability – Directed Evolution Strategies for Improved Protein Function. *FEBS J.* **2013**, *280* (22), 5582–5595. https://doi.org/10.1111/febs.12354.

(17) Stimple, S. D.; Smith, M. D.; Tessier, P. M. Directed Evolution Methods for Overcoming Trade-offs between Protein Activity and Stability. *AIChE J.* **2020**, *66* (3). https://doi.org/10.1002/aic.16814.

(18) Li, Y.; Drummond, D. A.; Sawayama, A. M.; Snow, C. D.; Bloom, J. D.; Arnold, F. H. A Diverse Family of Thermostable Cytochrome P450s Created by Recombination of Stabilizing Fragments. *Nat. Biotechnol.* **2007**, *25* (9), 1051–1056. https://doi.org/10.1038/nbt1333.

(19) Heinzelman, P.; Snow, C. D.; Wu, I.; Nguyen, C.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. A Family of Thermostable Fungal Cellulases Created by Structure-Guided Recombination. *Proc. Natl. Acad. Sci.* **2009**, *106* (14), 5610–5615. https://doi.org/10.1073/pnas.0901417106.

(20) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **2016**, *63* (2), 337–346. https://doi.org/10.1016/j.molcel.2016.06.012.

(21) Musil, M.; Konegger, H.; Hon, J.; Bednar, D.; Damborsky, J. Computational Design of Stable and Soluble Biocatalysts. *ACS Catal.* **2019**, *9* (2), 1033–1054. https://doi.org/10.1021/acscatal.8b03613.

(22) Lu, H.; Diaz, D. J.; Czarnecki, N. J.; Zhu, C.; Kim, W.; Shroff, R.; Acosta, D. J.; Alexander, B. R.; Cole, H. O.; Zhang, Y.; Lynd, N. A.; Ellington, A. D.; Alper, H. S. Machine Learning-Aided Engineering of Hydrolases for PET Depolymerization. *Nature* **2021**, *604* (7907), 662–667. https://doi.org/10.1038/s41586-022-04599-z.

(23) Islam, M. S.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* **2018**, *87*, 585–620. https://doi.org/10.1146/annurev-biochem-061516-044724.

(24) Herr, C. Q.; Hausinger, R. P. Amazing Diversity in Biochemical Roles of Fe(II)/2-Oxoglutarate Oxygenases. *Trends Biochem Sci* **2018**, *43* (7), 517–532. https://doi.org/10.1016/j.tibs.2018.04.002.

(25) Papadopoulou, A.; Meyer, F.; Buller, R. M. Engineering Fe(II)/α-Ketoglutarate-Dependent Halogenases and Desaturases. *Biochemistry* **2022**. https://doi.org/10.1021/acs.biochem.2c00115.

(26) Zwick, C. R.; Renata, H. Harnessing the Biocatalytic Potential of Iron- and α-Ketoglutarate-Dependent Dioxygenases in Natural Product Total Synthesis. *Nat. Prod. Rep.* **2020**, *37* (8), 1065–1079. https://doi.org/10.1039/c9np00075e.

(27) Liu, C.; Zhao, J.; Liu, J.; Guo, X.; Rao, D.; Liu, H.; Zheng, P.; Sun, J.; Ma, Y. Simultaneously Improving the Activity and Thermostability of a New Proline 4-Hydroxylase by Loop Grafting and Site-Directed Mutagenesis. *Appl. Microbiol. Biotechnol.* **2019**, *103* (1), 265–277. https://doi.org/10.1007/s00253-018-9410-x.

(28) Cheung-Lee, W. L.; Kolev, J. N.; McIntosh, J. A.; Gil, A. A.; Pan, W.; Xiao, L.; Velásquez, J. E.; Gangam, R.; Winston, M. S.; Li, S.; Abe, K.; Alwedi, E.; Dance, Z. E. X.; Fan, H.; Hiraga, K.; Kim, J.; Kosjek, B.; Le, D. N.; Marzijarani, N. S.; Mattern, K.; McMullen, J. P.; Narsimhan, K.; Vikram, A.; Wang, W.; Yan, J.; Yang, R.; Zhang, V.; Zhong, W.; DiRocco, D. A.; Morris, W. J.; Murphy, G. S.; Maloney, K. M. Engineering Hydroxylase Activity, Selectivity, and Stability for a Scalable Concise Synthesis of a Key Intermediate to Belzutifan. *Angew. Chem. Int. Ed.* **2024**, *63* (13), e202316133. https://doi.org/10.1002/anie.202316133.

(29) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; Haas, R. J. de; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378* (6615), 49–56. https://doi.org/10.1126/science.add2187.

(30) Sumida, K. H.; Núñez-Franco, R.; Kalvet, I.; Pellock, S. J.; Wicky, B. I. M.; Milles, L. F.; Dauparas, J.; Wang, J.; Kipnis, Y.; Jameson, N.; Kang, A.; Cruz, J. D. L.; Sankaran, B.; Bera, A. K.; Jiménez-Osés, G.; Baker, D. Improving Protein Expression, Stability, and Function with ProteinMPNN. *J. Am. Chem. Soc.* **2024**. https://doi.org/10.1021/jacs.3c10941.

(31) Yang, J.; Li, F.-Z.; Arnold, F. H. Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering. *ACS Cent. Sci.* **2024**, *10* (2), 226–241. https://doi.org/10.1021/acscentsci.3c01275.

(32) Roduner, E.; Kaim, W.; Sarkar, B.; Urlacher, V. B.; Pleiss, J.; Gläser, R.; Einicke, W.; Sprenger, G. A.; Beifuß, U.; Klemm, E.; Liebner, C.; Hieronymus, H.; Hsu, S.; Plietker, B.; Laschat, S. Selective Catalytic Oxidation of C□H Bonds with Molecular Oxygen. *ChemCatChem* **2013**, *5* (1), 82–112. https://doi.org/10.1002/cctc.201200266.

(33) He, C.; Whitehurst, W. G.; Gaunt, M. J. Palladium-Catalyzed C(Sp3)–H Bond Functionalization of Aliphatic Amines. *Chem* **2019**, *5* (5), 1031–1058. https://doi.org/10.1016/j.chempr.2018.12.017.

(34) Klein, C.; Hüttel, W. A Simple Procedure for Selective Hydroxylation of L-Proline and L-Pipecolic Acid with Recombinantly Expressed Proline Hydroxylases. *Adv. Synth. Catal.* **2011**, *353* (8), 1375–1383. https://doi.org/10.1002/adsc.201000863.

(35) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19* (6), 679–682. https://doi.org/10.1038/s41592-022-01488-1.

(36) Hu, X.; Huang, X.; Liu, J.; Zheng, P.; Gong, W.; Yang, L. Structures of L-Proline *Trans*-Hydroxylase Reveal the Catalytic Specificity and Provide Deeper Insight into AKG-Dependent Hydroxylation. *Acta Crystallogr. Sect. D* **2023**, *79* (Pt 4), 318–325. https://doi.org/10.1107/s2059798323001936.

(37) Lukat, P.; Katsuyama, Y.; Wenzel, S.; Binz, T.; König, C.; Blankenfeldt, W.; Brönstrup, M.; Müller, R. Biosynthesis of Methyl-Proline Containing Griselimycins, Natural Products with Anti-Tuberculosis Activity. *Chem Sci* **2017**, *8* (11), 7521–7527. https://doi.org/10.1039/c7sc02622f.

(38) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(39) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2* (2), 83–92. https://doi.org/10.1021/sb300037w.

(40) Goldberg, N. W.; Knight, A. M.; Zhang, R. K.; Arnold, F. H. Nitrene *Trans*fer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Ligands. *J. Am. Chem. Soc.* **2019**, *141* (50), 19585–19588. https://doi.org/10.1021/jacs.9b11608.