NMR-Guided Directed Evolution

Ivan Korendovych (ikorendo@syr.edu)
Syracuse University https://orcid.org/0000-0001-8144-783X
Sagar Bhattacharya
Syracuse University
Eleanora Margheritis
Yokohama City University
Katsuya Takahashi
Yokohama City University
Alona Kulesha
Syracuse University
Areetha D'Souza
Syracuse University
Inhye Kim
Syracuse University
Jeremy Tame
Yokohama city university
Jennifer Yoon
Syracuse University
Alexander Volkov
VIB Centre for Structural Biology
Olga Makhlynets
https://orcid.org/0000-0002-2763-7891

Biological Sciences - Article

Keywords:

Posted Date: December 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1150236/v1

Version of Record: A version of this preprint was published at Nature on October 5th, 2022. See the published version at https://doi.org/10.1038/s41586-022-05278-9.

Abstract

Directed evolution can rapidly achieve dramatic improvements in the properties of a protein or bestow entirely new functions on it. We have discovered a strong correlation between the probability of finding a productive mutation at a particular position of a protein and a chemical shift perturbation in Nuclear Magnetic Resonance spectra upon addition of an inhibitor for the chemical reaction it promotes. In a proof-of-concept study we converted myoglobin, a non-enzymatic protein, into the most active Kemp eliminase reported to date using only three mutations. The observed levels of catalytic efficiency are on par with the levels shown by natural enzymes. This simple approach, that requires no a priori structural or bioinformatic knowledge, is widely applicable and will unleash the full potential of directed evolution.

Full Text

Directed evolution is a powerful tool for improving existing properties and imparting completely new functionalities onto proteins.[1-4] Nonetheless, even in small proteins its potential is inherently limited by the astronomical number of possible amino acid sequences. Sampling the complete sequence space of a 100-residue protein would require testing of 20^{100} combinations, which is currently beyond any existing experimental approach. Fortunately, in practice, selective modification of relatively few residues is sufficient for efficient improvement, functional enhancement and repurposing of existing proteins.[5] Moreover, computational methods have been developed to predict the location, and, in certain cases, identities of potentially productive mutations.[6-9] Importantly, all current approaches for prediction of hot spots and productive mutations rely heavily on structural information and/or bioinformatics, which is not always available for proteins of interest. Moreover, they offer limited ability to identify beneficial mutations far from the active site, even though such changes may dramatically improve the catalytic properties of an enzyme.[10] Here we show that mutagenic hot spots in enzymes can be identified using Nuclear Magnetic Resonance (NMR) spectroscopy. In a proof-of-concept study we converted myoglobin, a non-enzymatic oxygen storage protein, into a highly efficient Kemp eliminase using only three mutations. The observed levels of catalytic efficiency (k_{cat}/K_M of 2.8 x 10^6 $M^{-1}s^{-1}$ and $k_{cat}/k_{uncat} > 10^8$) are the highest reported for any designed protein and are on par with the levels shown by natural enzymes for the reactions they are evolved to catalyze. Given the simplicity of this experimental approach, which requires no *a priori* structural or bioinformatic knowledge, we expect it to be widely applicable and to unleash the full potential of directed enzyme evolution.

Recent paradigm shifting advances in understanding the fundamental principles that drive enzyme evolution point to a major role of global conformational selection for productive arrangements of functional groups to perfect transition state stabilization, as well as steric and electrostatic interactions[11-15]. Here we seek to build on this recent work to predict experimentally the locations of the productive mutations that can minimize non-essential protein dynamics to achieve high catalytic efficiency. Efficient catalysis relies on tight and specific association of the substrate with the enzyme, placing it in a unique anisotropic environment (often with a high dipole moment, considered to be important for activity[16]).

Experimentally, such an environment can be evaluated using NMR, which provides residue-level information under catalytic conditions without the need for full structural characterization. In a conformational ensemble, residues that require substantial reorganization to adopt or to increase the population of a specific rotamer to support the transition state, should experience a large change in their NMR chemical shift upon addition of the corresponding transition state analog (usually a competitive inhibitor). Thus, analysis of the chemical shift perturbation (CSP) upon inhibitor titration may help identify mutagenic hot spots in the protein structure, both near and far from the active site.

Kemp elimination (Fig. 1) is a well-established and benchmarked model reaction for testing protein design and evolution methodologies[17-25], thus we set up to explore whether an NMR-guided approach can be successfully used to evolve a Kemp eliminase. Inspired by the recent discovery of redox-mediated Kemp elimination promoted by cytochrome P450[26], we sought to use a non-enzymatic heme protein as a starting point for the evolution. For the unbiased test of the approach, we chose not to perform any computational pre-selection of possible candidates, but rather focused on the simplest proteins. Myoglobin (Mb), arguably the most well characterized heme protein, adopts catalytic functionalities upon replacement of distal histidine His64[27], which in the native protein controls oxygen binding and slows heme oxidation. Mb-H64V has been extensively studied before[28], so we experimentally tested this mutant for the ability to promote Kemp elimination. In the reduced form, Mb-H64V demonstrated catalytic efficiency of 104 $M^{-1}s^{-1}$ at pH 8.0 presenting itself as a promising candidate for NMR-guided directed evolution (Table 1). Even with paramagnetism and high helical content of the reduced protein, a nearly full backbone assignment was possible, which enabled us to perform a CSP study using 6-nitrobenzotriazole (6-NBT), an inhibitor of Kemp elimination (Fig. 1). The data show 15 hot spots, defined as regions with residues CSP Z-score of above 1, dispersed around the protein, both near to and away from the heme cofactor (Fig. 2a,d). Next, we prepared saturation mutagenesis libraries in all positions with Z>1 and their immediate neighbors (except for the proximal His93 that was not considered as it is absolutely required for the heme cofactor binding). Crude lysate screening of the saturation mutagenesis libraries showed hits in all hot spots. Purification of the identified proteins confirmed the screening results in all cases (all showing large increase in catalytic efficiencies ranging from 2.3-fold to 93-fold, with an average of 21-fold) except in one instance (Mb-H64V/Q152M), where we were unable to produce enough soluble protein for kinetic characterization. Nine of the 19 identified productive mutations were located away from the active site (Fig. 2d).

Saturation mutagenesis performed in 12 randomly selected positions with small CSP yielded no hits (Fig. 2a, blue asterisks). In a subsequent non-exhaustive gene shuffling experiment, we found that L29I, H64G and V68A can be productively combined with positive synergy (the observed rate is 3-fold higher relative to the independent contribution of all mutations), a trait quite uncommon in traditional directed evolution experiments. The resulting triple mutant Mb-L29I/H64G/V68A, named FerrElCat for FERRous Kemp ELimination CATalyst, showed a remarkable Kemp elimination activity with catalytic efficiency of 2,796,000 $M^{-1}s^{-1}$ at pH 8 (Table 1). This level of catalytic efficiency is more than an order of magnitude higher than that of the most active reported Kemp eliminase HG3.17, evolved in 17 rounds of directed

evolution[19], and is on par with the levels shown by natural enzymes for the reactions they are evolved to catalyze. Importantly, NMR-guided approach yields mutants with high k_{cat} values (1,398 s^{-1} for FerrElCat), a trait that is often hard to achieve in traditional approaches to directed evolution, where high levels of catalytic efficiency are often achieved by lowering the K_M. FerrElCat is capable of at least 10,000 turnovers before showing signs of product inhibition (Extended Data Fig. 1). The unprecedented, experimentally guided *ca.* 27,000-fold improvement in catalytic efficiency (Extended Data Fig. 2) over the starting design in directed evolution of a catalyst for an unnatural reaction was obtained with only three mutations in a non-enzymatic protein (Fig. 2c). The crystal structure of FerrElCat shows remarkable similarity to the starting point of the evolution[29] (RMSD of 0.16 Å, Fig. 2e). While we were unable to obtain a crystal structure of FerrElCat with an inhibitor, docking studies (Fig. 2e) show that directed evolution results in creation of a tight binding pocket bringing the substrate into proximity to the heme iron. Strikingly, we were unable to dock neither 5-NBI nor 6-NBT into the crystal structure of Mb-H64V, that lacks a sufficiently large binding pocket (Fig. 2d). Yet, CSP analysis clearly shows association of the inhibitor with the protein in the place of the distal histidine, highlighting the power of NMR to easily identify productive arrangements of molecules that may not be apparent in modelling based on static crystal structures.

In conclusion, we have discovered a strong correlation between the degree of NMR CSP of backbone amide resonances in ^{15}N-^{1}H HSQC spectra of enzymes by an inhibitor and the probability of finding a beneficial mutation in the vicinity of that residue. The chemical shift perturbation maps are highly sensitive to minor changes in protein sequences, and pinpoint areas likely to affect catalytic activity, even if located far from the active site. In a proof-of-concept study we converted myoglobin, a non-enzymatic oxygen storage protein, into a highly efficient Kemp eliminase using only three mutations. To our knowledge this represents the first example of an experimental approach to guide directed evolution that does not rely on an *a priori* structural or bioinformatic analyses, and only requires reliable backbone amide assignments and an appropriate inhibitor. Such NMR data can usually be easily obtained for soluble, folded proteins with fewer than about 300 residues, a criterion that is true for many enzymes selected for directed evolution. Given the simplicity of this experimental approach, we expect it to be widely applicable to other proteins and to unleash the full potential of directed evolution to rapidly create new enzymes for practically important chemical transformations. These results also highlight the power of the minimalist approach to design of protein catalysts[30] that allows for quick and inexpensive identification of starting points for subsequent directed evolution without detailed consideration of the reaction mechanism as well as extensive computation, and instead exploits the incredible plasticity of proteins to adopt new functions. Last, but not least, our results contribute to the ongoing debate about the role of dynamics in enzymatic catalysis[11-15] by prospectively validating the importance of conformational selection in protein evolution. This opens the path to new high value fundamental studies of enzymatic function and evolution.

References

1. Bornscheuer, U. T. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185-194 (2012).

2. Reetz, M. T. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew. Chem. Int. Ed.* **50**, 138-174 (2011).

3. Denard, C. A., Ren, H. & Zhao, H. Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.* **25**, 55-64 (2015).

4. Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat. Catal.* **3**, 203-213 (2020).

5. Reetz, M. T., Wilensek, S., Zha, D. & Jaeger, K.-E. Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angew. Chem. Int. Ed.* **40**, 3589-3591 (2001).

6. Wijma, H. J., Floor, R. J. & Janssen, D. B. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.* **23**, 588-594 (2013).

7. Planas-Iglesias, J. et al. Computational design of enzymes for biotechnological applications. *Biotech. Adv.* **47**, 107696 (2021).

8. Verma, R., Schwaneberg, U. & Roccatano, D. Computer-aided protein directed evolution: a review of web servers, databases and other computational tools for protein engineering. *Comput. Struct. Biotech. J.* **2**, e201209008 (2012).

9. Ebert, M. C. & Pelletier, J. N. Computational tools for enzyme improvement: why everyone can – and should – use them. *Curr. Opin. Chem. Biol.* **37**, 89-96 (2017).

10. Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Comp. Mol. Sci.* **11**, e1502 (2021).

11. Acevedo-Rocha, C. G. et al. Pervasive cooperative mutational effects on multiple catalytic enzyme traits emerge via long-range conformational dynamics. *Nat. Commun.* **12**, 1621 (2021).

12. Otten, R. et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442-1446 (2020).

13. Campbell, E. et al. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944-950 (2016).

14. Hong, N.-S. et al. The evolution of multiple active site configurations in a designed enzyme. *Nat. Commun.* **9**, 3900 (2018).

15. Broom, A. et al. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat. Commun.* **11**, 4808 (2020).

16. Warshel, A. et al. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210-3235 (2006).

17. Kemp, D. S. & Casey, M. L. Physical organic chemistry of benzisoxazoles. II. Linearity of the Bronsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **95**, 6670-6680 (1973).

18. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).

19. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418-421 (2013).

20. Risso, V. A. et al. *De novo* active sites for resurrected Precambrian enzymes. *Nat. Commun.* **8**, 16113 (2017).

21. Merski, M. & Shoichet, B. K. Engineering a model protein cavity to catalyze the Kemp elimination. *Proc. Nat. Acad. Sci. U.S.A.* **109**, 16179–16183 (2012).

22. Debler, E. W., Müller, R., Hilvert, D. & Wilson, I. A. An aspartate and a water molecule mediate efficient acid-base catalysis in a talored antibody pocket. *Proc. Nat. Acad. Sci. U. S. A.* **106**, 18539-18544 (2009).

23. Vaissier, V., Sharma, S. C., Schaettle, K., Zhang, T. & Head-Gordon, T. Computational optimization of electric fields for improving catalysis of a designed Kemp eliminase. *ACS Cat.* **8**, 219-227 (2018).

24. Lamba, V. et al. Kemp eliminase activity of Ketosteroid Isomerase. *Biochemistry* **56**, 582-591 (2017).

25. Risso, V. A. et al. Enhancing a *de novo* enzyme activity by computationally-focused ultra-low-throughput screening. *Chem. Sci.* **11**, 6134-6148 (2020).

26. Li, A. et al. A redox-mediated Kemp eliminase. *Nat. Commun.* **8**, 14876 (2017).

27. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54**, 1744-1748 (2015).

28. Yi, J., Heinecke, J., Tan, H., Ford, P. C. & Richter-Addo, G. B. The distal pocket histidine residue in horse heart myoglobin directs the *O*-binding mode of nitrite to the heme iron. *J. Am. Chem. Soc.* **131**, 18119-18128 (2009).

29. Wang, B. et al., Nitrosyl myoglobins and their nitrite precursors: crystal structural and quantum mechanics and molecular mechanics theoretical investigations of preferred Fe–NO ligand orientations in myoglobin distal pockets. *Biochemistry* **57**, 4788-4802 (2018).

30. Marshall, L. R., Zozulia, O., Lengyel-Zhand, Z. & Korendovych, I. V. Minimalist *de novo* design of protein catalysts. *ACS Cat.* **9**, 9265-9275 (2019).

Table 1

Table 1 | Kinetic parameters for Kemp elimination promoted by selected myoglobin mutants at pH 8.0 in the reduced form, unless noted.

Proteins	k_{cat}, s^{-1}	K_M, mM	k_{cat}/K_M, $M^{-1}s^{-1}$
Mb-H64V [a]	-	-	7 ± 1 [b]
Mb-H64V	-	-	104 ± 3 [b]
Mb-H64V/F43L	35.62 ± 3.93	3.70 ± 0.48	9,626 ± 1,641
Mb-H64V/V68A	-	-	2,631 ± 61 [b]
Mb-H64V/L29I	-	-	528 ± 7 [b]
Mb-H64G	10.0 ± 1.42	1.39 ± 0.29	7,194 ± 1,799
Mb-H64G/V68A	-	-	442,560 ± 1,883 [b]
FerrElCat (Mb-L29I/H64G/V68A)	1,398 ± 132	0.5 ± 0.1	2,796,000 ± 264,000

[a] Oxidized form. [b]Individual k_{cat} and K_M values could not be determined due to substrate solubility.

Methods

Chemicals and reagents

Reagents and buffers were purchased from Biobasic, Inc. and Santa Cruz Biotechnology, Inc. Buffers were made using MilliQ water (Millipore Elix 3 instrument). DNA oligonucleotides were purchased from Integrated DNA technologies (IDT). All enzymes for cloning and mutagenesis were obtained from Thermo Fischer Scientific. *E. coli* BL21 (DE3), BL21 (DE3) pLysS, and NEB5a cells were purchased from Promega and New England Biolabs (NEB). 5-nitrobenzisoxazole (5-NBI) was prepared according to the literature procedure[31] while 6-nitrobenzotriazole (6-NBT) was purchased from AK scientific. pET28a(+) vector was obtained from Novagen. (L)-Ascorbic acid, superoxide dismutase and catalase (from bovine liver) were obtained from Sigma Aldrich. Labeled ^{15}N-ammonium chloride and $^{13}C_6$-D-glucose were purchased from Cambridge Isotope Laboratories.

Protein expression and purification

His64 in Myoglobin gene was mutated to Valine (H64V) using splicing by overlap extension (SOE) PCR. At the first PCR, the combination of mutagenic primers targeting the desired site and primer pair overlapping the 5' and 3' termini of the gene generated two mutant fragments (primer sequences for two pairs of primers are shown in Supplementary Table 1: NcoI_XhoI_F + H64V_R and NcoI_XhoI_R + H64V_F) which served as templates for the second PCR. Product of the second PCR was cloned into pET28a vector using *Nco*I and *Xho*I restriction sites. Mutations were introduced as needed using SOE protocol instead of site-directed mutagenesis. Myoglobin proteins contain many histidine residues (11 His residues for H64V mutant) and bind to the Ni-NTA column without the requirement of any additional His-tag. Plasmids encoding the appropriate genes in pET28a vector were transformed into *E. coli* BL21 (DE3) and plated on LB agar with 50 µg/mL kanamycin (Kan). In all subsequent experiments this concentration of antibiotic was used. Single colonies were inoculated into LB media containing Kan and grown at 37 °C for 5-6 h. Starter culture (10 mL) was inoculated into LB (1 L) with Kan and allowed to grow at 37 °C until OD_{600} reached 0.6-0.8. Next, δ-aminolevulinic acid (Tokyo Chemical Industry, 0.3 mM) was added and the culture was induced by the addition of 0.25 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and grown at 25 °C for 20 h. Cells were harvested by centrifugation at 4 °C, 4,000 g, flash frozen in liquid nitrogen and stored at -80 °C. For purification of protein, cells were resuspended in buffer A (25 mM TRIS, pH 8.0), lysed by sonication, and centrifuged to separate the soluble fraction. The lysate was loaded onto a Ni-NTA column (Clontech) pre-equilibrated with buffer A, washed and eluted using a gradient of 20-250 mM imidazole in buffer A. Protein fractions were exchanged into buffer B (20 mM HEPES, pH 7.0) using desalting column. Further purification was performed on cation exchange column (HiTrap SP HP, GE Healthcare) and FPLC in buffer B using NaCl gradient of 0-600 mM. Eluted protein fractions were exchanged into buffer B. Protein concentrations were determined from Soret band maxima using extinction coefficients experimentally determined pyridine hemochromagen assay[32]. SDS-PAGE gels for the proteins are shown in Extended Data Figure 3. For the expression of isotopically labelled proteins. plasmid pET28a(+) encoding gene of myoglobin was transformed into *E. coli* BL21 (DE3) cells and plated on LB agar containing Kan. Individual colonies were inoculated into LB (20 mL) with Kan and then the culture was incubated with shaking (200 rpm) at 37 °C for 5-6 h. This culture was diluted with Terrific broth (TB, 2 L) and Kan and grown at 37 °C until OD_{600} reached 0.6-0.8. Cells were collected by centrifugation, resuspended in unlabeled M9 minimal media (18 mL) and transferred to labelled M9 minimal media (1 L) prepared with $^{15}NH_4Cl$ and dextrose (or $^{13}C_6$-glucose) containing Kan and then the culture was allowed to grow at 37 °C. After 3-4 h of incubation, δ-aminolevulinic acid (0.3 mM) was added as heme precursor and the culture was induced by adding IPTG (0.25 mM). The culture was further grown at 25 °C for 20 h. Cells were collected by centrifugation and preserved at -80 °C. Purification was performed following the same protocol as for unlabeled protein.

Reduction and concentration determination of myoglobin variants

For standardization of dithionite, 20-30 mg of solid dithionite (Riedel-de Haen, Germany) as well as potassium ferricyanide (Sigma) were brought into the glovebox. Both the solid reagents were dissolved in 1 mL of degassed MilliQ water to prepare stock solutions. Dithionite stock was further diluted by 20-fold.

Next, two 1 mL solutions were prepared where in the first one, potassium ferricyanide stock was diluted by 100-fold while in the second one, a 1:1 mixture of ferricyanide stock and 20-fold diluted dithionite solution was prepared with subsequent dilution of each of them by 100-fold. Absorbances of both the solutions were measured at λ_{max} = 420 nm using UV-Vis diode array spectrophotometer (Agilent 8453). The reducing equivalence of 20-fold diluted dithionite solution was calculated from the difference in absorbances of the solutions using extinction coefficient of 1,020 $M^{-1}cm^{-1}$ at 420 nm. Protein concentrations were determined using the extinction coefficient of 157,000 $M^{-1}cm^{-1}$ at absorption maxima at 434 nm for reduced species. Absorbance spectra of Mb-H64V and Mb-L29I/H64G/V68A (FerrElCat) in the oxidized and reduced forms are shown in Extended Data Fig. 4.

Kinetic characterization

Extinction coefficients of 15,800 $M^{-1}cm^{-1}$ at 380 nm was used for the product of Kemp elimination, 2-hydroxybenzonitrile. Catalytic efficiency (k_{cat}/K_M) as well as individual kinetic parameters (k_{cat} and K_M, where possible), were determined by fitting the dependence of initial rates on substrate concentration (final concentration of 100-1000 μM) to the Michaelis-Menten equation $v_0 = k_{cat}[E][S]/(K_M+[S])$. For proteins with high K_M values, where substrate saturation could not be achieved due to solubility limits, k_{cat}/K_M was determined by fitting data to $v_0 = (k_{cat}/K_M)[E][S]$. All measurements were done at least in triplicate with different batches of proteins. In all cases the rates are corrected for the background rates in the appropriate buffers without the enzymes. Degassed protein samples were reduced by adding *ca.* 10 equivalents of sodium dithionite (Riedel-de Haen, Germany) inside a glovebox (MBRAUN). Concentrations of reduced proteins were determined using a Soret band (typically around 434 nm) and the extinction coefficient of 157,000 $M^{-1}cm^{-1}$. Reactions were initiated by mixing reduced enzyme in buffer and solutions of 5-nitrobenzisoxazole in water in 1:1 ratio using stopped-flow spectrometer (Applied Photophysics SX20). The final reaction mixtures contained reduced protein (5 nM-1 μM) in 20 mM TRIS, pH 8.0 with ascorbate (1 mM), SOD (0.1 μM), catalase (20 nM), 1.5 % acetonitrile and variable concentrations of the substrate. Product formation was monitored at 25 °C (temperature was controlled by Thermo Scientific water bath) for the time interval of 0.1-10 s. The resulting Michaelis-Menten plots of Kemp elimination catalyzed by reduced myoglobin mutant proteins using stopped-flow spectrophotometry are in Extended Data Fig. 5 and the corresponding kinetic parameters are in Extended Data Table 2.

Library design

The gene encoding sperm whale myoglobin was cloned into pET-28a(+) (Novagen) with simultaneous introduction of the H64V mutation using standard protocols. Site-specific saturation mutagenesis targeting defined site was achieved using megaprimer PCR protocol[33] with primer sets (Integrated DNA Technologies) which overlapped on the 5’ terminus of the randomized position, together with flanking primer (T7 forward or T7 reverse), as appropriate (NNK primer sequences are shown in Supplementary Table 1). The size of PCR product was verified using agarose gel electrophoresis. DNA sample was

digested with *DpnI* (New England Biolabs) at 37 °C for 10-12 h to eliminate parental clone. Digested sample was transformed into *E. coli* NEB5α cells (New England Biolabs) and subsequently plated on Luria Bertani (LB) agar plate containing kanamycin (50 μg/mL). After incubation at 37 °C for 10-12 h, colonies obtained from the plate were allowed to grow in LB with Kan at 37 °C for 5-6 h. Cells were harvested, and plasmids were extracted using DNA extraction kit (Monarch, New England Biolabs). Library quality was confirmed by Sanger sequencing analysis (Genewiz, Inc.) (Extended Data Fig. 6).

Library screening

Myoglobin NNK libraries were transformed into *E. coli* BL21 (DE3) pLysS cells and plated on LB agar with Kan and chloramphenicol (CHL, 34 μg/mL). Individual colonies were inoculated into LB (200 μL) containing Kan and CHL in 96-well plate. Cultures were incubated at 37 °C until OD_{600} 0.6-0.8 and replica plate was generated where cultures were inoculated into LB with Kan and CHL. δ-aminolevulinic acid (0.3 mM) as heme precursor and 0.25 mM IPTG for induction were added to the cultures and grown at 25 °C for 20 h. Cells were harvested by centrifugation. Pellets were resuspended with buffer (25 mM TRIS, pH 8.0), centrifuged again, and the supernatant was discarded. A buffer containing 25 mM TRIS, pH 8.0, 0.5% triton X was used to lyse the cells and supernatant was separated by centrifugation. Activity of the clones was tested using 96-well plates in the buffer (20 mM TRIS, pH 8.0, 100 mM NaCl, 1 mM ascorbate, 0.1 μM superoxide dismutase and 20 nM catalase) by measuring absorbance at 380 nm at 22 °C on a plate reader (BioTek Eon3). The activities of clones showing large increase over the starting templates were confirmed by rescreening them in triplicate. Plasmids extracted from the colonies demonstrating improved activity were sequenced (Genewiz, Inc.) to determine the identities of beneficial mutations. Sequences of Mb-H64V and FerrElCat are given in Extended Data Table 1.

NMR spectroscopy

All NMR spectra were acquired at 298 K on a Bruker Avance III HD 800 MHz spectrometer equipped with a TCI cryoprobe. The protein samples were prepared in 20 mM HEPES pH 7.0 and contained 5-10 % D_2O for the lock. The assignments of backbone amide resonances were obtained from 0.7-1.0 mM U-[^{13}C,^{15}N] protein samples using a standard set of 3D BEST HNCACB, HN(CO)CACB, HNCO and HN(CA)CO experiments. All myoglobin samples were reduced with sodium dithionite under a nitrogen atmosphere and the NMR tubes were flame sealed. The NMR data were processed in NMRPipe[34] and analyzed in CCPNMR[35]. The CSP experiments were performed by stepwise addition of a concentrated stock solution of 6-NBT in CH_3CN to U-[^{15}N] or U-[^{13}C,^{15}N] protein samples at the initial concentration of 0.2 mM. At each increment, changes in chemical shifts of the protein resonances were monitored in 2D [^{1}H,^{15}N] HSQC spectra. The average amide CSPs ($\Delta\delta_{avg}$) were obtained at two-fold molar excess of 6-NBT as $\Delta\delta_{avg} = (\Delta\delta_N^2/50+\Delta\delta_H^2/2)^{0.5}$, where $\Delta\delta_N$ and $\Delta\delta_H$ are the chemical shift perturbations of the amide nitrogen and proton, respectively (Supplementary Table S4). For each observed resonance, the Z score was calculated as $Z = (\Delta\delta_{avg} - \mu)/\sigma$, where μ and σ are, respectively, the average and the standard deviation of $\Delta\delta_{avg}$ values for a given CSP experiment.

Crystallographic methods

Crystals of FerrElCat were grown by hanging drop vapor diffusion method at 20 °C upon mixing protein solution and reservoir solution (100 mM TRIS, pH 8.0, 2.4 M ammonium sulfate). X-ray diffraction data were collected using a Pilatus-200 K detector on a Rigaku Micromax-007 rotating anode X-ray generator. The protein was crystallized in hexagonal space group (P6). Diffraction data were processed with the CrysAlisPro software suite (Rigaku). All structures were determined by molecular replacement using PHASER[36] starting with the deposited model of myoglobin (PDB code 1mbn). Refinement was performed by COOT[37] and PHENIX[38], and the final structures were validated with MolProbity[39]. The crystallographic data and refinement statistics were deposited with the entry code 7vuc. Crystallographic data and refinement statistics for Mb-L29I/H64G/V68A (FerrElCat) are in Extended Data Table S3.

Kemp substrate docking

Kemp substrate was docked into FerrElCat using AutoDock Vina[8] and described protocol[9]. The imidazole molecule associated with the heme in the crystal structure was removed prior to docking.

Circular dichroism (CD) spectroscopy

All CD spectra of myoglobin variants were recorded using Jasco J-715 CD spectrometer in continuous mode with 1 nm bandwidth, 2 nm data pitch, scan rate of 50 nm/min with 8 s averaging time. The final spectra represent a buffer-subtracted average of three runs. The CD spectra of non-reduced proteins in the far-UV region (200-260 nm) were collected using quartz cuvette with 1 mm pathlength while for the Soret band region (390-470 nm) quartz cuvette of 1 cm pathlength was used. The spectra at Soret band region (390-470 nm) were obtained to determine mean residue ellipticity values (MRE) assuming protein binds heme in a 1:1 ratio. Protein stocks were diluted in 2 mM TRIS (pH 8.0) to 5 μM and the spectra were recorded for oxidized protein. For the analysis of the reduced protein, the stock was diluted in the same way as for the oxidized sample and ten equivalents of sodium dithionite were added to the protein inside the glovebox. The concentration of the reduced protein was calculated based on the Soret band maxima using the corresponding extinction coefficients. Sample absorbance never exceeded 2 at all wavelengths. The mean residue ellipticity (MRE, deg*cm^{2}*dmol^{-1}) values were calculated using the following equation (MRE = θ/(10*c*l*N)), where θ (mdeg) is ellipticity, l (cm) is the pathlength of the cuvette, c (M) is the protein concentration and N is the number of residues (Extended Data Fig. 7).

Data availability

The crystal structure of FerrElCat (Mb-L29I/H64G/V68A) was deposited in the Worldwide Protein Data Bank (wwPDB) under the accession number 7VUC.

Methods References

31. Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.***38**, 2294-2301 (1973).

32. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.***161**, 1-15 (1987).

33. Barik, S. in *PCR Cloning Protocols* (eds Bing-Yuan Chen & Harry W. Janes), 189-196 (Humana Press, 2002).

34. Delaglio, F. et al. NMRpipe: a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR***6**, 277-293, (1995).

35. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins***59**, 687-696, (2005).

36. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D Biol. Crystallogr.***60**, 432-438, (2004).

37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.***66**, 486-501, (2010).

38. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.***66**, 213-221, (2010).

39. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.***66**, 12-21, (2010).

40. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput. Chem.***31**, 455-461, (2010).

41. Makhlynets, O. V. & Korendovych, I. V. Minimalist design of allosterically regulated protein catalysts. *Methods Enzymol.***580**, 191-202, (2016).

Declarations

Acknowledgments This work was supported by the National Institutes of Health grant (grant no. GM119634) and the Alexander von Humboldt Foundation. The authors thank Prof. Rudi Fasan for the gift of the plasmid containing the Mb gene. J.R.H.T. thanks OpenEye Scientific Software.

Author contributions I.V.K., O.V.M. and A.N.V. designed the experiments. S.B., A.K., J.H.Y. and I.K. performed directed evolution, protein expression and characterization as well as kinetic studies. A.D., A.N.V., O.V.M. and I.V.K. performed NMR titrations, backbone and side chain assignment and NMR data analysis. E.M., K.T. and J.R.H.T. expressed and crystallized the proteins and solved their structures. I.V.K.

and O.V.M. wrote the manuscript with input of all the authors. [5]These authors contributed equally to this work: Eleanor Margheritis, Katsuya Takahashi, Alona Kulesha

Competing interests The authors declare that they have no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/

Correspondence and requests for materials should be addressed to Ivan V. Korendovych

Figures

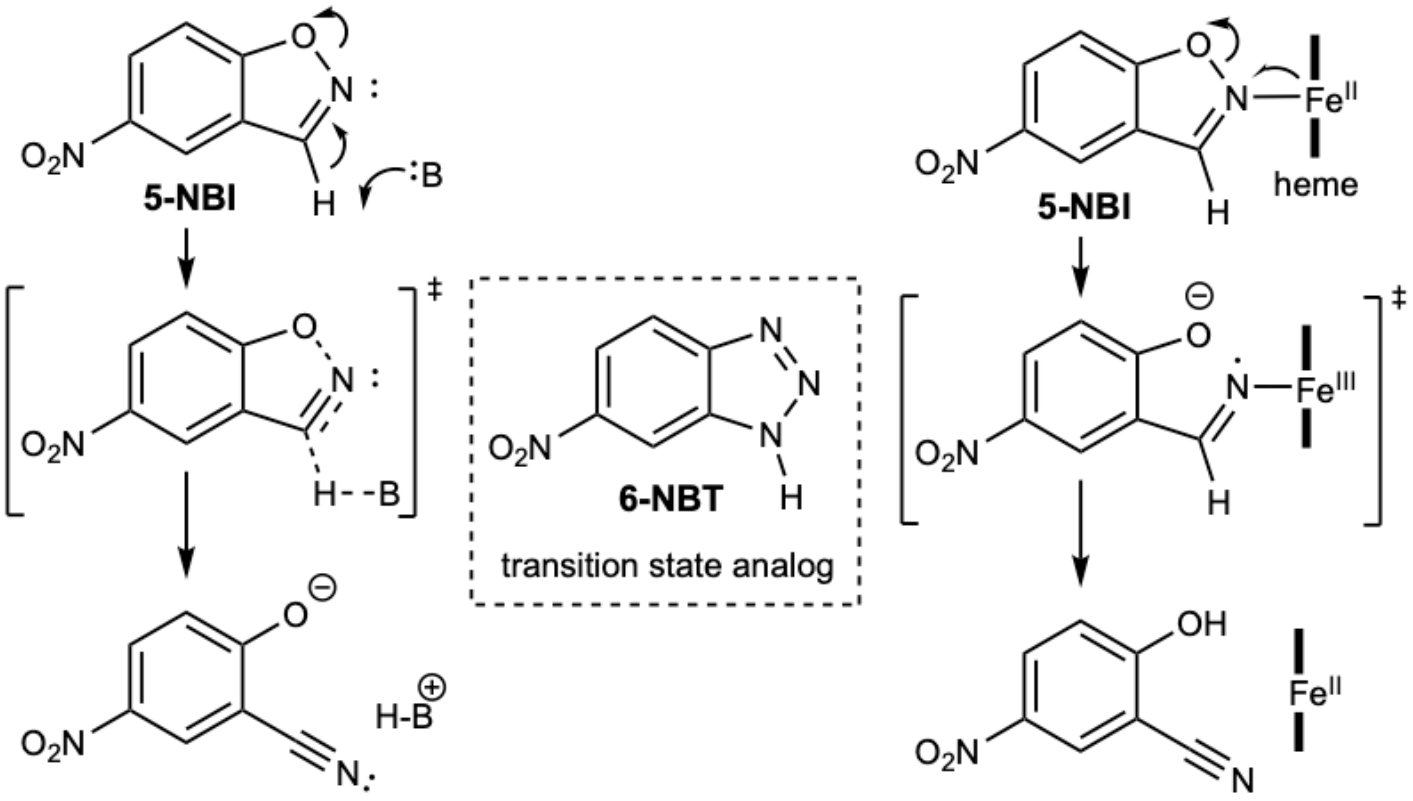

Figure 1

Kemp elimination promoted by acid-base or redox mechanisms.

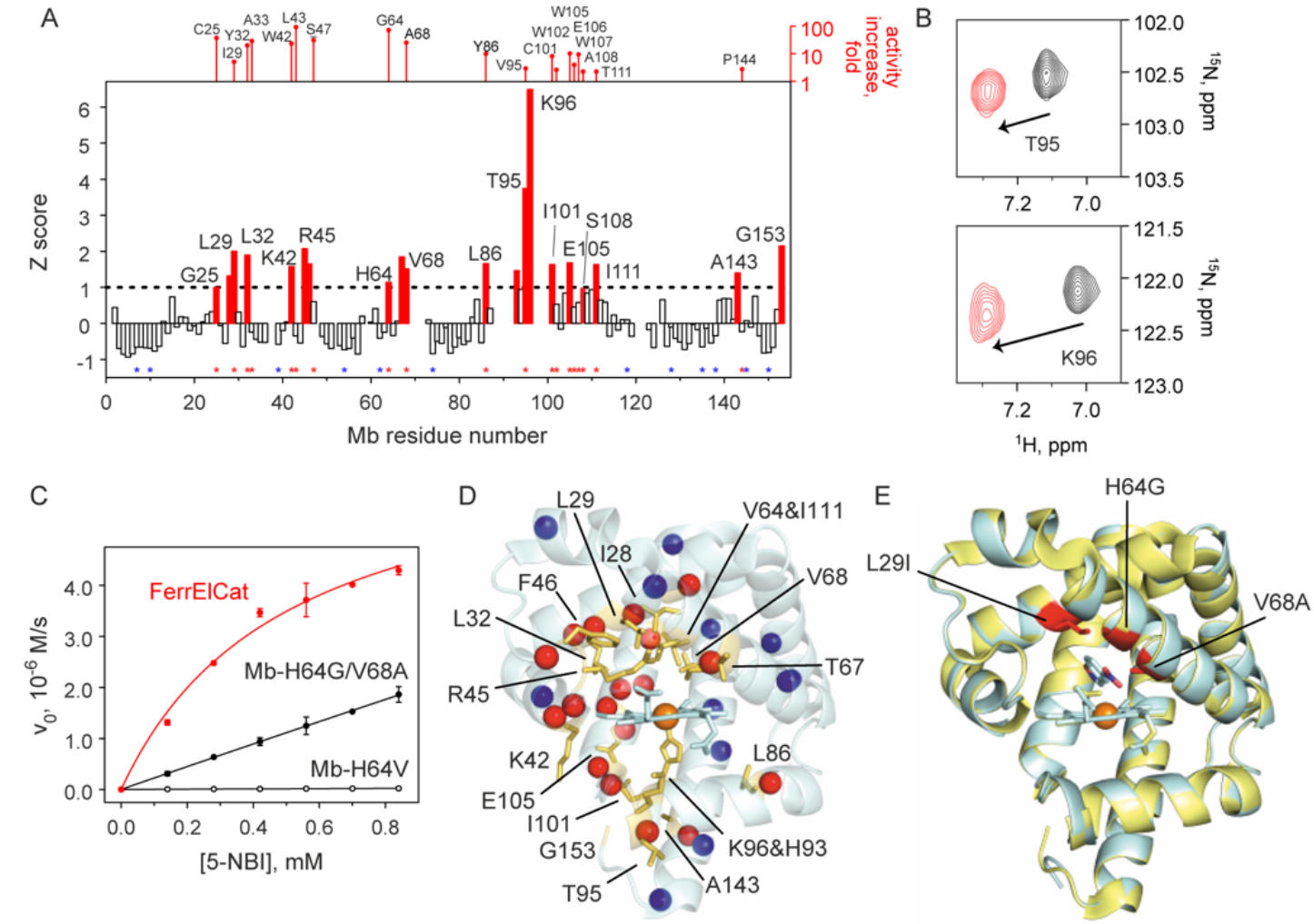

Figure 2

NMR-guided evolution of myoglobin. a, Backbone amide CSP of Mb-H64V upon addition of 2 molar equivalents of 6-NBT. The red bars indicate the protein regions experiencing large chemical shift perturbation (Z > 1). No bars are provided where no backbone resonance assignment could be made. The positions where productive mutations were found are marked by red asterisks (along with the corresponding increase in kcat/KM relative to Mb-H64V, top). Positions where screening identified no productive mutations are marked by blue asterisks. The corresponding representative 1H-15N HSQC spectral regions are shown in panel b. c, Michaelis-Menten plots for representative proteins. d, NMR CSP data mapped on the X-ray crystal structure of Mb-H64V (PDB 6cf0) showing the residues with Z > 1 as yellow sticks. The spheres show backbone nitrogen atoms of the residues with identified productive mutations (red) or those for which no productive mutations could be found (blue). e, Overlay of the crystal structures of Mb-H64V (yellow) and FerrElCat with the docked inhibitor (cyan). The newly introduced mutations are shown in red.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.docx
- ExtendedData.docx