Article

NMR-guided directed evolution

https://doi.org/10.1038/s41586-022-05278-9

Received: 7 December 2021

Accepted: 25 August 2022

Published online: 5 October 2022

Check for updates

Sagar Bhattacharya[1], Eleonora G. Margheritis[2,5], Katsuya Takahashi[2,5], Alona Kulesha[1,5], Areetha D'Souza[1], Inhye Kim[1], Jennifer H. Yoon[1], Jeremy R. H. Tame[2], Alexander N. Volkov[3,4]✉, Olga V. Makhlynets[1]✉ & Ivan V. Korendovych[1]✉

Directed evolution is a powerful tool for improving existing properties and imparting completely new functionalities to proteins[1–4]. Nonetheless, its potential in even small proteins 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 beyond any existing experimental approach. 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 locations 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 a limited ability to identify beneficial mutations far from the active site, even though such changes may markedly improve the catalytic properties of an enzyme[10]. Machine learning methods have recently showed promise in predicting productive mutations[11], but they frequently require large, high-quality training datasets, which are difficult to obtain in directed evolution experiments. Here we show that mutagenic hot spots in enzymes can be identified using 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 exceed those of proteins designed using current approaches and are similar with those of natural enzymes for the reactions that they are evolved to catalyse. 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 enable the full potential of directed enzyme evolution.

Recent 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 optimize transition-state stabilization, as well as steric and electrostatic interactions[12–16]. 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, which is considered to be important for activity[17]). Experimentally, such an environment can be evaluated using NMR, which provides residue-level information under catalytic conditions without the need for a 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 analogue (usually a competitive inhibitor). Thus, analysis of the chemical shift perturbation (CSP) upon addition of an inhibitor may help to 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[18–26]. Inspired by the recent discovery of redox-mediated Kemp elimination promoted by cytochrome P450 (ref. [27]) and aldoxime dehydratases[28], we set out to explore whether an NMR-guided approach can be successfully used to evolve a novel Kemp eliminase from a non-enzymatic haem protein. For an unbiased test of the approach, we chose to not perform any computational pre-selection of possible candidates, but rather focused on the simplest proteins. Myoglobin (Mb), arguably the most well-characterized haem protein, adopts catalytic functions upon replacement of distal histidine 64 ref. [29]), which controls oxygen binding and slows haem oxidation in the native protein. Mb(H64V) has been extensively studied[30], so we experimentally tested

[1]Department of Chemistry, Syracuse University, Syracuse, NY, USA. [2]Graduate School of Medical Life Science, Yokohama City University, Yokohama, Kanagawa, Japan. [3]VIB Centre for Structural Biology, Vlaams Instituut voor Biotechnologie (VIB), Brussels, Belgium. [4]Jean Jeener NMR Centre, Vrije Universiteit Brussel (VUB), Brussels, Belgium. [5]These authors contributed equally: Eleonora G. Margheritis, Katsuya Takahashi, Alona Kulesha. ✉e-mail: ovolkov@vub.be; ovmakhly@syr.edu; ikorendo@syr.edu

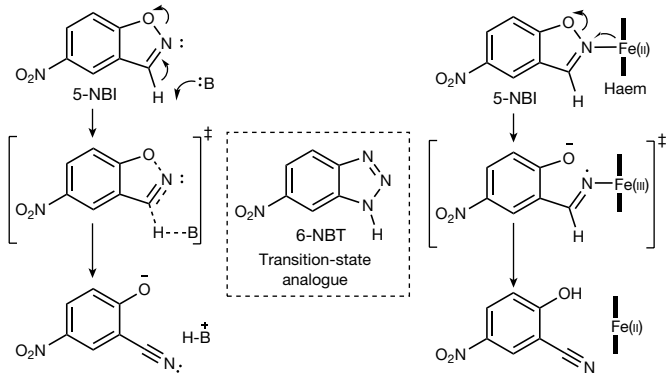


Fig. 1 | Kemp elimination. 5-Nitrobenzisoxazole (5-NBI) ring opening promoted by acid–base (left) and redox (right) mechanisms. 6-Nitrobenzotriazole (6-NBT) is a transition-state analogue for this reaction.

this mutant for the ability to promote Kemp elimination. In the reduced form, Mb(H64V) had a catalytic efficiency of 255 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 the high helical content of the reduced protein, a nearly full backbone assignment was possible, which enabled us to perform a CSP study using 6-NBT, an inhibitor of Kemp elimination (Fig. 1). The data show 15 hot spots, defined as regions with residue CSP *Z*-scores of above approximately 1, dispersed around the protein, both near to and away from the haem cofactor (Fig. 2a,d). Next, we prepared saturation mutagenesis libraries in all positions with $Z \geq 1$ and their immediate neighbours (except for the proximal His93, which was not considered as it is required for the binding of the haem cofactor). 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 (with increases in catalytic efficiencies ranging from 2-fold to 71-fold, with an average of 20-fold), except in one instance (Mb(H64V/Q152M)) where we were unable to produce enough soluble protein for kinetic characterization. Nine out of the 19 identified productive mutations were located away from the active site (Fig. 2d).

Saturation mutagenesis performed at 18 randomly selected positions with small CSP yielded no hits (Fig. 2a, blue asterisks). Since the probability of finding productive mutations is highest close to the active site, we sampled all cold spots in the immediate vicinity of the active site and tried to provide a representative sampling of the positions located further away (Extended Data Fig. 1). In a subsequent non-exhaustive gene-shuffling experiment, we found that L29I, H64G and V68A can be productively combined with positive synergy (the triple mutant is threefold more active than predicted from the three individual mutations), an uncommon trait in traditional directed evolution experiments. The resulting enzyme Mb(L29I/H64G/V68A), which we named ferrous Kemp elimination catalyst (FerrElCat), showed a remarkable Kemp elimination activity with a catalytic efficiency of 15,721,000 M^{-1} s^{-1} at pH 8.0 (Table 1). This level of catalytic efficiency is almost two orders of magnitude higher than that of the most active reported Kemp eliminase, HG3.17, which was evolved over 17 rounds of directed evolution[20], and is similar to levels exhibited by the most active natural enzymes for the reactions they have evolved to catalyse. It is also only one or two orders of magnitude from the diffusion limit. Notably, this NMR-guided approach yields mutants with high values of the catalytic rate constant (k_{cat}) (3,656 s^{-1} for FerrElCat), a trait that is often difficult to achieve using traditional approaches to directed evolution, where high levels of catalytic efficiency are often obtained by lowering the Michaelis constant (K_M). FerrElCat is capable of at least 10,000 turnovers before showing signs of product inhibition (Extended Data Fig. 2). This experimentally guided improvement of approximately 62,000-fold in catalytic efficiency (Extended Data Fig. 3) over the starting design was obtained with only three mutations of a non-enzymatic protein (Fig. 2c). The crystal structure of FerrElCat exhibits remarkable similarity to the starting point of the evolution[31] (backbone root mean squared deviation (r.m.s.d.) of 0.16 Å; Fig. 2e) and the newly introduced mutations had only a minor effect on the cofactor redox potential (Extended Data Fig. 4). Although we were unable to obtain a crystal structure of FerrElCat with an inhibitor, docking studies (Fig. 2e) show that directed evolution results in the creation of a tight binding pocket, bringing the substrate into proximity with the haem iron. Of note, we were unable to dock either 5-NBI or 6-NBT into the crystal structure of Mb(H64V) because the computationally predicted binding pocket is too small (Fig. 2d). Yet CSP analysis clearly shows association of the inhibitor with the protein, highlighting the power of NMR to readily identify productive arrangements of molecules that may not be apparent in modelling based on static crystal structures.

To test the general applicability of the NMR-guided directed evolution we applied it to the Kemp eliminases of the AlleyCat family that promote benzisoxazole ring opening using base-facilitated catalysis[18–24].

Table 1 | Kinetic parameters for Kemp elimination promoted by selected Kemp eliminases at pH 8.0

Protein	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
Mb(H64V)[a]	NA	NA	7±1
Mb(H64V)	NA	NA	255±8
Mb(H64V/F43L)	26.10±3.85	1.94±0.38	13,458±670
Mb(H64V/V68A)	NA	NA	12,939±622
Mb(H64V/L29I)	NA	NA	1,550±55
Mb(H64G)	NA	NA	18,152±519
Mb(H64G/V68A)	2,557±372	1.28±0.28	1,992,300±143,420
FerrElCat (Mb(L29I/H64G/V68A))	3,656±667	0.23 ± 0.13	15,721,000±6,035,800
AlleyCat	NA	NA	5.8±0.3[b]
AlleyCat7	3.2±0.2[b]	2.4±0.2[b]	1,283±13[b]
AlleyCat8	10.1±1.5	4.1±0.7	2,451±15, 2,369±99[c]
AlleyCat8(T146R)	5.8±0.7	1.6±0.3	3,563±39
AlleyCat9	18.9±3.9	4.9±1.2	3,857±27, 3,894±61[c]
AlleyCat10	21.2±2.8	4.8±0.7	4,378±20, 4,392±83[c]

[a]Oxidized form. [b]from ref. [31]. [c]$(k_{cat}/K_M)_{max}$ obtained from the pH activity profile. NA, not applicable (individual k_{cat} and K_M values could not be determined owing to low substrate solubility). The myoglobin mutants were reduced unless stated otherwise.

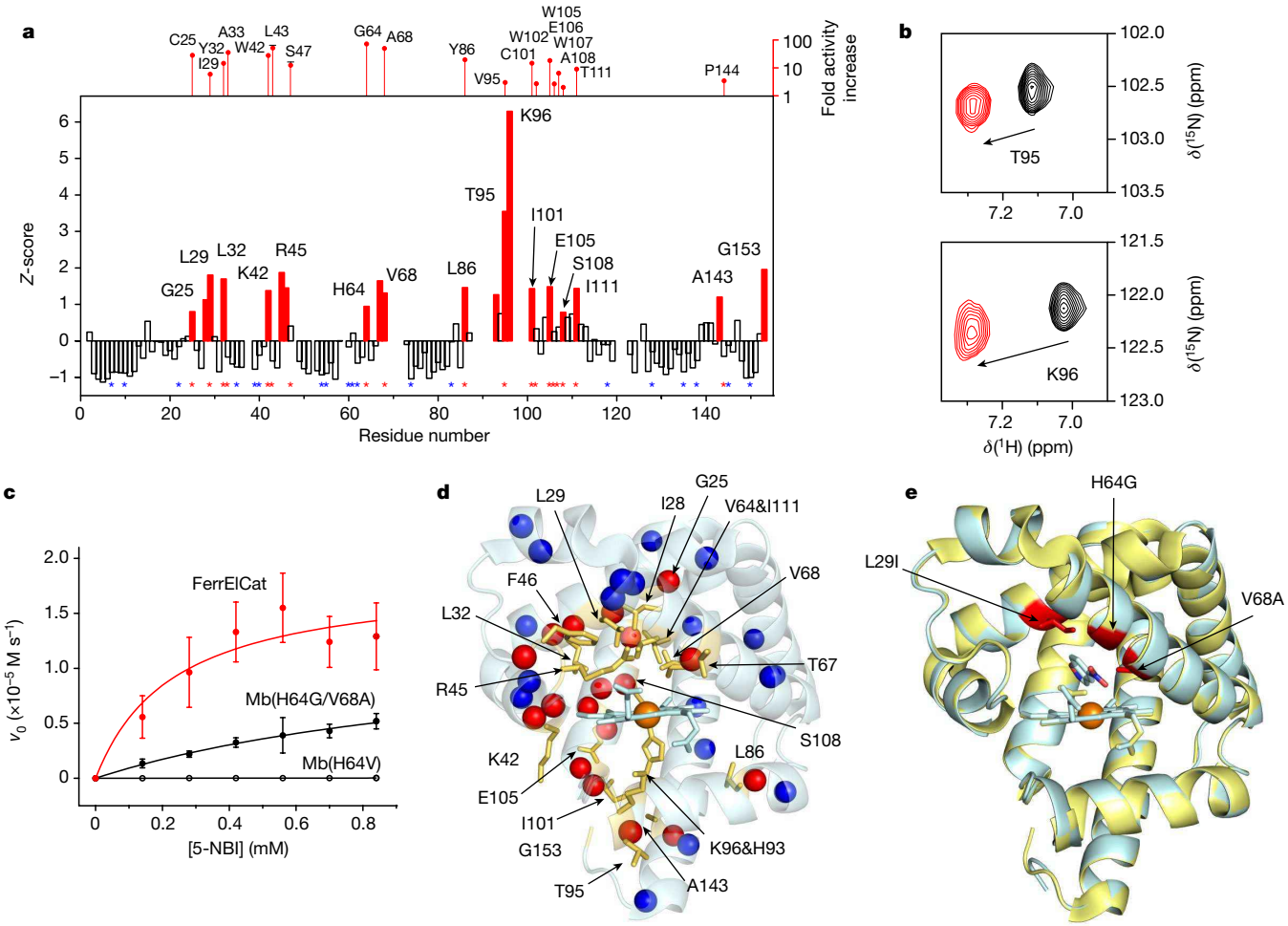


Fig. 2 | NMR-guided evolution of myoglobin. a, Backbone amide CSP of Mb(H64V) upon addition of 2 molar equivalents of 6-NBT. The red bars indicate the protein regions experiencing large CSP ($Z \gtrsim 1$). No bars are shown where no backbone resonance could be assigned. The positions where productive mutations were found are marked with red asterisks, along with the corresponding increase in k_{cat}/K_M relative to Mb(H64V) (top). Positions where screening did not identify any productive mutations are marked by blue asterisks. The corresponding representative ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectral regions are shown in **b**. **c**, Michaelis–Menten plots for representative proteins. Data are mean ± s.d. from six (nine for FerrElCat) independent measurements. **d**, NMR CSP data mapped on the X-ray crystal structure of Mb(H64V) (Protein Data Bank (PDB) ID: 6CF0) showing the residues with prominent changes ($Z \gtrsim 1$) 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.

AlleyCat was designed using a minimalist approach by introducing a single glutamate residue into the 74-residue C-terminal domain of calmodulin (cCaM), a non-enzymatic protein[32]. Subsequently, in seven rounds of directed evolution using saturation mutagenesis, error-prone PCR and gene shuffling, we evolved AlleyCat into AlleyCat7, which showed turnover numbers similar to some of the best examples of Kemp eliminases[33]. Owing to its small size, diamagnetism, extensive previous characterization and a wealth of functional data obtained through traditional approaches to directed evolution, the AlleyCat proteins provide an excellent and unbiased test bed for the NMR-guided directed evolution both retrospectively, to evaluate the performance of CSP-based approaches, and prospectively, to test the limits of the method. CSP maxima observed upon titrating 6-NBT into the C-terminal domain of calmodulin that was used as a starting point for the design (Fig. 3a) are in excellent agreement with the first three mutations introduced into the protein during the design and subsequent directed evolution[33]: F92E, M144R and H107I. Residue 107 is notably not part of the substrate-binding pocket. Upon introduction of the F92E mutation, a new hot spot consistent with the previously found productive A88Q mutation in AlleyCat appears (Fig. 3a). Notably, we observed a drop in CSP Z-values in the C-terminal region of the protein, where beneficial mutations in positions 144 and 145 were found in AlleyCat, potentially related to a more than threefold decrease in affinity for the inhibitor (dissociation constant (K_d) of 3.3 mM for AlleyCat versus 1.0 mM for cCaM). Encouraged by the similarity of the trajectories between the traditional and NMR-guided evolution, we undertook a prospective study to determine whether CSP analysis could be used to improve the catalytic efficiency of AlleyCat7. The CSP data for AlleyCat7 (Fig. 3b) are quite different from those of cCaM and AlleyCat both in terms of positions of the major peaks as well as their relative magnitude. We chose to not pursue residues in the calcium-binding EF hand domains that are essential for both the fold and allosteric regulation. Since we have already introduced mutations at positions 124, 128 and 144, we performed saturation mutagenesis at position 125. AlleyCat7(I125H) (named AlleyCat8), which we identified in the screening, showed a threefold increase in k_{cat} (Table 1). No beneficial mutations were found by saturation mutagenesis of any positions that did not show significant CSP (Fig. 3b, blue asterisks). The CSP graph for AlleyCat8 again shows significant changes (Fig. 3c). The most prominent shifts for AlleyCat8 are observed for residues 114–116, which were little affected in previous

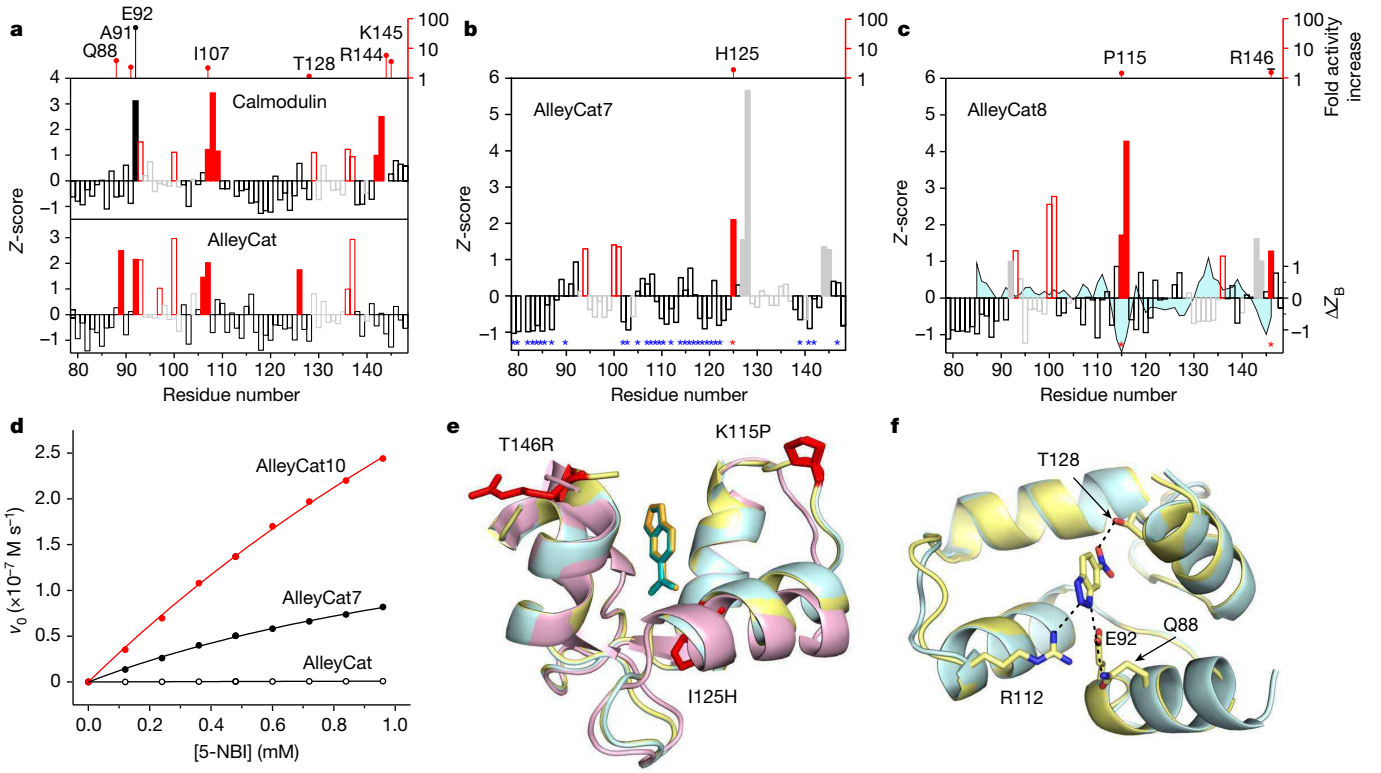


Fig. 3 | NMR-guided evolution of calmodulin. a–**c**, Backbone amide CSP of cCaM and AlleyCat (**a**), AlleyCat7 (**b**) and AlleyCat8 (**c**) upon addition of 2 molar equivalents of 6-NBT. The red bars indicate the protein regions experiencing large CSP ($Z \geq 1$). The open red and grey bars identify EF hand residues. The catalytic F92E mutation is shown with a solid black bar. The positions where productive mutations were found are marked by red asterisks in **b**,**c** along with the corresponding increase in k_{cat}/K_M relative to the previous round of design (top). Positions where screening identified no productive mutations are marked by blue asterisks in **b**. The solid grey bars in **b**,**c** refer to residues already mutated in previous rounds. The difference in *Z*-score of crystallographic B-factors (Cα) for the inhibitor-bound and free AlleyCat9 is mapped onto the CSP data on AlleyCat8 (ΔZ_B trace in **c**). **d**, Michaelis–Menten plots for representative proteins. Data are mean ± s.d. from three independent measurements. **e**, Overlay of the crystal structures of cCaM (magenta), AlleyCat9 with the inhibitor (cyan) and AlleyCat10 with the inhibitor (yellow). The residues identified in CSP analysis are shown in red. **f**, Overlay of the crystal structures of apo (cyan) and inhibitor-bound (yellow) AlleyCat10.

generations of the protein, as well as residues 143 and 146. Saturation mutagenesis in positions 114, 115, 116 and 146 (position 143 is next to the previously mutated Met144) yielded the productive mutations K115P (a variant subsequently called AlleyCat9) and T146R, which show significant improvements in k_{cat}/K_M driven by the increase in k_{cat} for K115P and the decrease in K_M for T146R. The effect of these two mutations is additive, so that the resulting protein AlleyCat10 has a k_{cat}/K_M of 4,378 M^{-1} s^{-1} and a k_{cat} of 21.0 s^{-1}. This represents a 750-fold improvement in catalytic efficiency over the starting design (a value second only to that of FerrElCat; Extended Data Fig. 3) that was achieved almost exclusively by improving k_{cat}.

The new members of the AlleyCat family fully preserve allosteric regulation by calcium (Extended Data Fig. 5). Crystallographic characterization of AlleyCat9 and AlleyCat10 both in the absence and in the presence of the inhibitor shows that these highly evolved variants, with more than 10% of their sequence mutated, remain structurally very similar (backbone r.m.s.d. 0.6–0.9 Å) to the C-terminal domain of calmodulin that served as a basis for the design[34] (Fig. 3e,f). Two of the three newly introduced productive mutant residues—Pro115 and Arg146—are located away from the active site (more than 10 Å from the inhibitor), showing the practical utility of CSP patterns for identifying effectively mutagenic hot spots across the whole protein. Analysis of the difference in crystallographic B-factors in the crystal structures of the apo and inhibitor-bound AlleyCat9 (Fig. 3c) shows significant rigidification of the protein structure upon binding of the transition-state analogue, consistent with the contribution of protein dynamics to the observed CSP patterns.

In conclusion, we have found a strong correlation between the degree of NMR CSP of backbone amide resonances in ^{1}H-^{15}N HSQC spectra of enzymes by an inhibitor and the probability of finding a beneficial mutation in the vicinity of that residue. The CSP maps are highly sensitive to minor changes in protein sequence and pinpoint areas that are likely to affect catalytic activity, even if they are 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. This approach only requires reliable backbone amide assignments and an appropriate inhibitor. To our knowledge, this represents the first example of an experimental approach to guide directed evolution that does not rely on a priori structural or bioinformatic analyses. 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, which we applied successfully to two unrelated proteins that utilize different mechanisms, we expect the CSP-guided methodology to be widely applicable to other proteins and to realize 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[35], which 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. Finally, our results contribute to the ongoing debate about the role of dynamics in enzymatic catalysis[12–16] by

prospectively validating the importance of conformational flexibility in protein evolution. This opens a path to new fundamental studies of enzymatic function and evolution.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05278-9.

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. Engl.* **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. Engl.* **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. *Biotechnol. 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. Biotechnol. 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 Comput. Mol. Sci.* **11**, e1502 (2021).
11. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl Acad. Sci. USA* **116**, 8852–8858 (2019).
12. 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).
13. Otten, R. et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446 (2020).
14. Campbell, E. et al. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944–950 (2016).
15. Hong, N. S. et al. The evolution of multiple active site configurations in a designed enzyme. *Nat. Commun.* **9**, 3900 (2018).
16. Broom, A. et al. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nat. Commun.* **11**, 4808 (2020).
17. Warshel, A. et al. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
18. 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).
19. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
20. Blomberg, R. et al. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **503**, 418–421 (2013).
21. Risso, V. A. et al. De novo active sites for resurrected Precambrian enzymes. *Nat. Commun.* **8**, 16113 (2017).
22. Merski, M. & Shoichet, B. K. Engineering a model protein cavity to catalyze the Kemp elimination. *Proc. Natl Acad. Sci. USA* **109**, 16179–16183 (2012).
23. Debler, E. W., Muller, R., Hilvert, D. & Wilson, I. A. An aspartate and a water molecule mediate efficient acid–base catalysis in a tailored antibody pocket. *Proc. Natl Acad. Sci. USA* **106**, 18539–18544 (2009).
24. 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 Catal.* **8**, 219–227 (2018).
25. Lamba, V. et al. Kemp eliminase activity of ketosteroid isomerase. *Biochemistry* **56**, 582–591 (2017).
26. Risso, V. A. et al. Enhancing a de novo enzyme activity by computationally-focused ultra-low-throughput screening. *Chem. Sci.* **11**, 6134–6148 (2020).
27. Li, A. et al. A redox-mediated Kemp eliminase. *Nat. Commun.* **8**, 14876 (2017).
28. Miao, Y., Metzner, R. & Asano, Y. Kemp elimination catalyzed by naturally occurring aldoxime dehydratases. *ChemBioChem* **18**, 451–454 (2017).
29. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed. Engl.* **54**, 1744–1748 (2015).
30. 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).
31. 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).
32. Korendovych, I. V. et al. Design of a switchable eliminase. *Proc. Natl Acad. Sci. USA* **108**, 6823–6827 (2011).
33. Moroz, O. V. et al. A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew. Chem. Int. Ed. Engl.* **52**, 6246–6249 (2013).
34. Chattopadhyaya, R., Meador, W. E., Means, A. R. & Quiocho, F. A. Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.* **228**, 1177–1192 (1992).
35. Marshall, L. R., Zozulia, O., Lengyel-Zhand, Z. & Korendovych, I. V. Minimalist de novo design of protein catalysts. *ACS Catal.* **9**, 9265–9275 (2019).

Methods

Chemicals and reagents

Reagents and buffers were purchased from Biobasic and Santa Cruz Biotechnology. Buffers were made using MilliQ water (Millipore Elix 3). DNA oligonucleotides were purchased from Integrated DNA technologies (IDT). All enzymes for cloning and mutagenesis were obtained from Thermo Fischer Scientific. *Escherichia coli* BL21 (DE3), BL21 (DE3) pLysS and NEB5α cells were purchased from Promega and New England Biolabs (NEB). 5-NBI was prepared as described[36], 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 MilliporeSigma. $^{15}NH_4Cl$ and $^{13}C_6$-D-glucose were purchased from Cambridge Isotope Laboratories.

Protein expression and purification

Myoglobin variants. The gene encoding Mb(H64V) was produced by 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 NcoI and XhoI restriction sites. Mutations were introduced as needed using the SOE protocol instead of site-directed mutagenesis. Myoglobin proteins contain many histidine residues (11 in Mb(H64V)) and bind to the Ni-NTA column without 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^{-1} kanamycin (Kan). This concentration of antibiotic was used in all experiments. Single colonies were inoculated into LB medium containing kanamycin and grown at 37 °C for 5–6 h. Starter culture (10 ml) was inoculated into LB (1 l) with kanamycin and allowed to grow at 37 °C until A_{600} reached 0.6–0.8. Next, 0.3 mM δ-aminolevulinic acid (Tokyo Chemical Industry) 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 collected by centrifugation at 4 °C, 4,000*g*, flash frozen in liquid nitrogen and stored at − 80 °C. For protein purification, cells were resuspended in buffer A (25 mM Tris, pH 8.0), lysed by sonication, and centrifuged to isolate 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—this step is critical. Eluted protein fractions were analysed by UV-vis spectroscopy and only the fractions with the appropriate Soret band maxima (Supplementary Table 2) were exchanged into buffer B. Proteinconcentrations were determined from Soret band maxima using extinction coefficient of 157,000 M^{-1} cm^{-1} or coefficients experimentally determined using pyridine haemochromagen assay[37]. For the expression of the isotopically labelled proteins the plasmid was transformed into *E. coli* BL21 (DE3) cells and plated on LB agar containing kanamycin. Individual colonies were inoculated into LB (20 ml) with kanamycin and then the culture was incubated at 37 °C for 5–6 h with shaking at 200 rpm. This culture was diluted with terrific broth (TB) (2 litres) and kanamycin and grown at 37 °C until A_{600} reached 0.6–0.8. Cells were collected by centrifugation, resuspended in unlabelled M9 minimal medium (18 ml) and transferred to labelled M9 minimal medium (1 litre) prepared with $^{15}NH_4Cl$ and dextrose (or [$^{13}C_6$]glucose, as appropriate) containing kanamycin, and then the culture was grown at 37 °C for 3–4 h. Next, 0.3 mM δ-aminolevulinic acid was added and the culture was induced by adding 0.25 mM IPTG and 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 unlabelled protein.

AlleyCat variants. Plasmids encoding SUMO-AlleyCat variants were transformed into *E. coli* BL21 (DE3) cells and plated on LB agar plates containing kanamycin. An individual colony was inoculated in LB containing kanamycin and grown at 37 °C for 5–6 h. Ten millilitres of starter culture was diluted with 1 l LB medium supplemented with kanamycin and grown at 37 °C until A_{600} reached ~0.6–0.8. The culture was induced by 0.5 mM IPTG and grown at 18 °C for 20 h. Cells were collected by centrifugation and resuspended in resuspension buffer (25 mM Tris, pH 8.0, 20 mM imidazole, 10 mM $CaCl_2$, 300 mM NaCl) with 0.5 mM phenylmethylsulfonyl fluoride. Cells were lysed by sonication (Microson) and the soluble fraction was loaded on a Ni-NTA Sepharose column (HisTrap, GE Healthcare). The proteins were eluted with buffer containing 25 mM Tris, pH 8.0, 20 mM imidazole, 10 mM $CaCl_2$, 300 mM NaCl and 20–500 mM imidazole applied in a gradient fashion on a FPLC system (NGC, BioRad). After buffer exchange on a desalting column (BioRad, 10 DG) into a cleavage buffer (50 mM Tris, pH 8.0, 75 mM NaCl), SUMO protease (at protein-to-protease absorbance (A_{280}) ratio of 200:1) along with EDTA (Invitrogen, 0.5 mM) and DTT (Sigma, 1 mM) was added to cleave the SUMO fusion tag. After incubation at 30 °C for 3–4 h, the protein was exchanged into storage buffer (20 mM HEPES, pH 7.0, 10 mM $CaCl_2$, 100 mM NaCl). Protein was further purified by anion exchange chromatography (Q HP FF, GE Healthcare) in storage buffer with elution gradient of 100–800 mM NaCl. Protein fractions were exchanged into storage buffer, concentrated using 5K MWCO spin concentrator (Corning). Protein concentrations were determined by measuring the absorbance at 280 nm using a calculated extinction coefficient of 2,980 M^{-1} cm^{-1}. For the expression of isotopically labelled AlleyCat variants, the plasmid was transformed into *E. coli* BL21 (DE3) cells and plated on LB agar plate containing kanamycin. Single colony was inoculated with 2 ml of LB medium containing kanamycin and grown at 37 °C for 5–6 h, then 18 ml of unlabelled M9 minimal medium with kanamycin was added and the culture was grown at 37 °C for an additional 5–6 h. The resulting 20 ml starter culture was diluted with 1 l of M9 medium made with $^{15}NH_4Cl$ as ^{15}N and [$^{13}C_6$]glucose, supplemented with kanamycin, and grown at 37 °C until A_{600} reached 0.6–0.8. The culture was then induced by 0.5 mM IPTG and grown at 18 °C for 20 h. The cells were then collected by centrifugation and the isotopically labelled protein was purified as discussed above. SDS–PAGE gels for all proteins used in this study are shown in Supplementary Fig. 1.

Reduction and concentration determination of myoglobin variants

For standardization of dithionite, 20–30 mg of solid dithionite (Riedel-de Haen) as well as potassium ferricyanide (MilliporeSigma) were brought into the MBRAUN glovebox with a nitrogen atmosphere (keeping oxygen under 2 ppm at all times). 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 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. 6. It is very important to ensure that the reduced protein has a sharp peak at ~434 nm without any pronounced shoulders (both in the reduced and oxidized forms) that could be indicative of multiple protein species present.

Kinetic characterization

In all cases the rates are corrected for the background rates in the appropriate buffers without the enzymes. Extinction coefficients of 15,800 M^{-1} cm^{-1} at 380 nm were 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 140–840 μM) to the Michaelis–Menten equation $v_0 = k_{cat}[E][S]/(K_M + [S])$, where v_0 is the initial velocity, and [E] and [S] are enzyme and substrate concentrations, respectively. For proteins with high K_M values, where substrate saturation could not be achieved due to substrate solubility limits, k_{cat}/K_M was determined by fitting data to $v_0 = (k_{cat}/K_M)[E][S]$. Myoglobin mutants: the substrate, 5-NBI, was prepared as 100 mM stock in acetonitrile inside the glovebox. Degassed protein samples were reduced by adding ~10 equivalents of sodium dithionite inside a glovebox. Concentrations of reduced proteins were determined using the extinction coefficient of 157,000 M^{-1} cm^{-1} for the Soret band. The quality of the reduced protein was assessed by examining the position and shape of Soret and Q-bands. Protein solutions containing reduced myoglobin mutants (10–200 nM) in 40 mM Tris, pH 8.0 with ascorbate (2 mM), superoxide dismutase (SOD) (0.2 μM) and catalase (40 nM) introduced to suppress side reactions stemming from any dioxygen present during the measurements, were prepared inside the glovebox in glass vials. To achieve best reproducibility, it is critical to prepare the working solution containing ascorbate, SOD, catalase and the reduced protein prior to each kinetic measurement. Ascorbate, SOD and catalase were added as separate drops to the walls of the vials, the protein was added to the buffer inside the vials, then the reagents were mixed by swirling and immediately transferred to gas-tight syringes for kinetic measurements. All measurements were done at least in triplicate, and multiple independently prepared protein batches were characterized for key mutants (FerrElCat, Mb(H64V), Mb(H64G) and Mb(H64G/V68A)). The substrate was prepared in the glovebox as 2× solution in water containing 3% acetonitrile (created using appropriate volumes of 100 mM stock solution of 5-NBI in acetonitrile, water and acetonitrile to maintain the constant final concentration of co-solvent) and transferred in gas-tight syringes. Reactions were initiated by mixing reduced enzyme in buffer and solutions of 5-NBI in water in 1:1 ratio on the Applied Photophysics SX20 stopped-flow spectrometer; the lines prior to the introduction of the reactants were equilibrated using deoxygenated water for the substrate channel and deoxygenated buffer in the protein channel. The final reaction mixtures contained reduced protein (5–100 nM) 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 for the time interval of 0.1–10 s using manufacturer provided software (Pro-Data, version 2.5.1852.0). The corresponding kinetic parameters are given in Extended Data Table 1. The pH dependence of FerrElCat enzymatic activity was assessed using the protocol described above, except different buffers were used (all at 20 mM; MES at pH 6.5, HEPES at pH 7.0 and 7.5, Tris at pH 8.0 and pH 8.5, Supplementary Fig. 3). AlleyCat proteins: the product formation was monitored on a BioTek Eon3 plate reader (manufacturer's data acquisition and analysis software version 2.05.5) in 20 mM HEPES, pH 7.0 buffer containing 10 mM $CaCl_2$, 100 mM NaCl and acetonitrile (1.5% constant final concentration) at 22 °C in a 96-well plate (Greiner Cellstar). The final enzyme concentration was 100 nM. The Michaelis–Menten plots for all proteins are given in Extended Data Fig. 7. The maximum k_{cat}/K_M as well as effective pK_a of the active site residue in the pH studies of AlleyCat proteins were obtained by using the equation $k_{cat}/K_M = (k_{cat}/K_M)_0 + (k_{cat}/K_M)_{max} \times (10^{-pKa}/(10^{-pKa} + 10^{-pH}))$ (Supplementary Fig. 4 and Supplementary Table 3).

NNK library design

The plasmids encoding the genes of the proteins of the AlleyCat family were constructed as reported[33]. The gene encoding sperm whale myoglobin was cloned into pET28a(+) (Novagen) with simultaneous introduction of the H64V mutation using standard protocols. Site-specific saturation mutagenesis targeting defined site was achieved using megaprimer PCR protocol[38] 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. Saturation mutagenesis was performed using NNK codons (where N can be any base and K represents a mixture of G and T nucleotides) that cover the 20 genetically encoded amino acids (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. The digested sample was transformed into *E. coli* NEB5α cells (New England Biolabs) and subsequently plated on LB agar plate containing kanamycin (50 μg ml^{-1}). After incubation at 37 °C for 10–12 h, colonies obtained from the plate were allowed to grow in LB with kanamycin at 37 °C for 5–6 h. Cells were collected, and plasmids were extracted using DNA extraction kit (Monarch, New England Biolabs). Library quality was confirmed by Sanger sequencing analysis (Genewiz) (Supplementary Figs. 5 and 6).

Library screening

At least 160 independent colonies were screened for each library. Myoglobin NNK libraries were transformed into *E. coli* BL21 (DE3) pLysS cells and plated on LB agar with kanamycin and chloramphenicol (34 μg ml^{-1}). Individual colonies were inoculated into LB (200 μL) containing kanamycin and chloramphenicol in 96-well plate. Cultures were incubated at 37 °C until A_{600} 0.6–0.8 and replica plate was generated where cultures were inoculated into LB with kanamycin and chloramphenicol. δ-Aminolevulinic acid (0.3 mM) as haem precursor and 0.25 mM IPTG for induction were added to the cultures and grown at 25 °C for 20 h. Cells were collected by centrifugation. Pellets were resuspended in 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, 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) to determine the identities of beneficial mutations.

Calmodulin gene libraries were transformed into *E. coli* BL21 (DE3) pLysS cells and plated on LB agar plates containing 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol (ampicillin and chloramphenicol at these concentrations were used in experiments with calmodulin libraries). Single colonies were inoculated into 200 μl of LB containing ampicillin and chloramphenicol in 96-well plate. After incubation for 5–6 h at 37 °C, replica plates were generated. Cultures were used to inoculate 400 μl of ZYM-5052 as autoinduction medium supplemented with ampicillin and chloramphenicol and allowed to grow at 37 °C for 12–16 h. The cells were collected by centrifugation and the supernatant was discarded. The cells were resuspended with 25 mM Tris, pH 8.0, 20 mM imidazole, 10 mM $CaCl_2$, 300 mM NaCl. After clearing the lysates, pellets were lysed with a buffer containing 20 mM Tris, pH 8.0, 10 mM $CaCl_2$, 100 mM NaCl, 0.2% Triton X and centrifuged to separate the lysate. Kemp elimination activity was monitored at 380 nm in the buffer containing 20 mM Tris, pH 8.0, 10 mM $CaCl_2$ and 100 mM NaCl in 96-well plates on a BioTek Eon3 plate reader at 22 °C for 10 min. 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) to determine the identities of beneficial mutations. The C-terminal domains of the improved variants were cloned

into pET-SUMO champion vector (Invitrogen) using standard protocols for detailed protein characterization. Sequences of all newly evolved proteins are given in Extended Data Table 2.

Gene shuffling

To quickly identify the most active variant we took the most active mutant identified in the first round of screening (H64G) and tested its performance together with the other mutations found in the first round. The same procedure was repeated on the most active double mutant (Mb(H64G/V68A)) to yield FerrElCat.

NMR spectroscopy

All NMR spectra were acquired at 298 K on a Bruker Avance III HD 800 MHz spectrometer equipped with a TCI cryoprobe using Bruker TopSpin Software (version 3.6). The Mb(H64V) samples were prepared in 20 mM HEPES pH 7.0, the AlleyCat samples were prepared in 20 mM HEPES, pH 6.9, 10 mM $CaCl_2$, 100 mM NaCl with 2.5 mM NaN_3. All samples 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 (version 5.97)[39] and analysed in CCPNMR (version 2.4.2)[40]. 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 CSPs of the amide nitrogen and proton, respectively (Supplementary Data). 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.

The NMR titrations to determine K_d values for AlleyCat and cCaM proteins were performed by incremental addition of a freshly prepared 5 mM 6-NBT solution in CH_3CN to a 0.2 mM [^{13}C,^{15}N] protein sample. At each increment, changes in chemical shifts of the protein resonances were monitored in ^{1}H-^{15}N HSQC spectra. The binding curves were analysed with a two-parameter nonlinear least-squares fit using a one-site binding model corrected for the dilution effect[41]: $\Delta\delta_{binding} = 0.5\Delta\delta_0(A - (A^2 - 4R)^{0.5})$, where $A = 1 + R + K_d(([NBT]_0 + R[protein]_0)/([NBT]_0[protein]_0))$; $\Delta\delta_{binding}$ is the CSP at a given [6-NBT]/[protein] ratio; $\Delta\delta_0$ is the CSP at 100% 6-NBT bound; *R* is the [6-NBT]/[protein] ratio at a given titration point; $[protein]_0$ and $[NBT]_0$ are the initial concentrations of the protein sample and the 6-NBT titrant stock solution, respectively; and K_d is the equilibrium dissociation constant (Supplementary Fig. 7).

Crystallographic methods

Crystals of FerrElCat were grown by the hanging drop vapour diffusion method at 20 °C upon mixing protein solution and reservoir solution (100 mM Tris, pH 8.0, 2.4 M ammonium sulfate). Crystal screening for AlleyCat9 as well as AlleyCat10 (~15 mg ml^{-1} in storage buffer) were performed in 96-well plate (Violamo) using sitting drop vapour diffusion method. Crystals were further grown at 293 K using hanging drop vapour diffusion method by 1:1 mixing of protein sample and reservoir buffer containing (4*S*)-2-methyl-2,4-pentanediol (MPD) (47%) and TBU (2%) for AlleyCat9 while 50% MPD was used for AlleyCat10. To obtain co-crystals with inhibitor, protein samples were mixed with 6-NBT (5–10 mM) and incubated on ice for 30 min. Crystals were then grown using the same conditions used for apo-protein crystallization. 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 (version 1.171.39.46e, Rigaku). All structures were determined by molecular replacement using PHASER[42] (version 2.8.2) starting with the deposited model of myoglobin (PDB ID: 1MBN) or AlleyCat (PDB ID: 2KZ2). Refinement was performed by COOT[43] (version 0.8.9.1) and PHENIX[44] (version 1.17.1-3660), and the final structures were validated with MolProbity[45]. Crystallographic data and refinement statistics are given in Extended Data Table 3.

Computational docking

5-NBI was docked into FerrElCat with AutoDock Vina[46] (version 1.1.2) with previously described protocol[47]. The imidazole molecule coordinated to the iron in the crystal structure was removed prior to docking.

Circular dichroism spectroscopy

All circular dichroism (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^{-1} 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 haem 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) (in deg cm^2 dmol^{-1}) values were calculated using the equation $MRE = \theta/(10 \times c \times l \times N)$, 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 (Supplementary Fig. 8). Chemical denaturation studies on AlleyCat proteins were performed by monitoring sample absorbance at 222 nm in presence of varying concentration of guanidine hydrochloride (0–6 M) as the denaturant (Supplementary Fig. 9) and thermodynamic parameters for protein unfolding were determined (Supplementary Table 4).

Spectroelectrochemical determination of redox potential of myoglobin mutants

The redox potentials of myoglobin variants were measured using a platinum honeycomb spectrochemical electrode (0.17 cm) with a Ag/AgCl reference electrode (reported as +199 mV vs NHE by manufacturer) connected to the WaveNow potentiostat (Pine Research Instrumentation). The redox titration was performed using 1.2 ml of working solution containing ~ 20 μM protein and 100 μM phenazine methosulfate (MilliporeSigma) as a redox mediator in 20 mM HEPES, pH 7.0 or 20 mM Tris, pH 8.0. A highly positive potential of +100 mV vs Ag/AgCl was initially applied to ensure complete oxidation of protein. Next, the potential was applied with an increment of 25 mV from +25 mV to −500 mV (vs Ag/AgCl) with the equilibration time of 5 min at 20 °C for each step. The absorbance spectra of the protein containing mixtures at each potential were recorded using UV-vis spectrophotometer (Agilent 8453). The absorbance at 434 nm (A_{434}, corrected for the contribution of the mediator) was normalized to 800 nm and used to determine the fraction of reduced protein in the sample using the following equation: Fraction reduced = $(A_{434} - A_{434min})/(A_{434max} - A_{434min}$, where A_{434max} and A_{434min} are maximum and minimum values of absorbance at 434 nm, representing reduced and oxidized species, respectively). The fraction reduced was plotted as a function of the applied potential and the midpoint reduction potential (E_m) was determined using the following equation: Fraction reduced = $(A + g_1 \times E_{app}) + (B + g_2 \times E_{app})/(1 + \exp((E_m - E_{app})/b))$, where A represents the upper asymptote, $A + B$ represents the lower asymptote, g_1 and g_2 represent the slopes of the

upper and the lower asymptotes, respectively, b is the growth rate, E_{app} is applied potential, and E_m is midpoint reduction potential. The midpoint reduction potentials obtained from the fits (Extended Data Fig. 4) are summarized in Supplementary Table 5.

Determination of AlleyCats and 6-NBT dissociation constants

The thermodynamic parameters of 6-NBT binding to AlleyCat proteins were measured using a MicroCal PEAQ-ITC instrument (Malvern Panalytical). Proteins were dialysed against buffer (20 mM HEPES, 100 mM NaCl, 10 mM $CaCl_2$, pH 7.0) containing 2% acetonitrile, filtered using 0.22 µm low protein binding PES filter (Santa Cruz Biotechnology) and the concentration of each sample was determined by UV-vis spectroscopy using extinction coefficients of 2,980 M^{-1} cm^{-1} at 280 nm. Solid 6-NBT was dissolved in the protein dialysis buffer to obtain 1 mM solution. All titrations were performed at 25 °C in the high feedback mode with 750 rpm stir speed and an appropriate equilibration time between injections (150 s). The protein sample (~100 µM) was placed in the calorimeter cell and 1 mM solution of 6-NBT was added to the protein in 18 2 µl aliquots. As a control, 1 mM 6-NBT solution was titrated into the dialysis buffer.

The analysis, including baseline correction, peak integration and correction for heat of dilution observed at the protein saturation with inhibitor, was performed using the MicroCal PEAQ-ITC analysis software provided by the manufacturer (version 1.21). To obtain binding parameters for each reaction, the data were fitted to the one set of sites model (Supplementary Fig. 10). Each titration was repeated at least two times. The thermodynamic parameters are summarized in Supplementary Table 6.

General data processing and analysis

Unless explicitly described above, data were processed and analysed using Microsoft Excel (version 16.59), KaleidaGraph (version 4.5.2, Synergy Software) and Igor Pro (version 8.4, Wavemetrics).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The crystallographic data and refinement statistics were deposited in the Protein Data Bank (PDB) with the entry code 7VUC (FerrElCat), 7VUR (AlleyCat9), 7VUS (AlleyCat9 with 6-NBT), 7VUT (AlleyCat10) and 7VUU (AlleyCat10 with 6-NBT). The NMR chemical shift data are provided as part of the Supplementary Information.

36. Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. Physical organic chemistry of benzisoxazoles. I. Mechanism of base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38**, 2294–2301 (1973).
37. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, hemes b, and hemes c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).
38. Barik, S. in *PCR Cloning Protocols* (eds. Chen, B.-Y. & Janes H.) 189–196 (Humana Press, 2002).
39. Delaglio, F. et al. NMRPipe—a multidimensional spectral processing system based on UNIX Pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
40. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
41. Kannt, A., Young, S. & Bendall, D. S. The role of acidic residues of plastocyanin in its interaction with cytochrome f. *Biochim. Biophys. Acta* **1277**, 115–126 (1996).
42. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60**, 432–438 (2004).
43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
44. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
45. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
46. 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).
47. Makhlynets, O. V. & Korendovych, I. V. Minimalist design of allosterically regulated protein catalysts. *Meth. Enzymol.* **580**, 191–202 (2016).

Acknowledgements This work was supported by the National Institutes of Health grant GM119634, the Japan Society for the Promotion of Science and the Alexander von Humboldt Foundation. The authors thank R. 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, and 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.G.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 from all the authors.

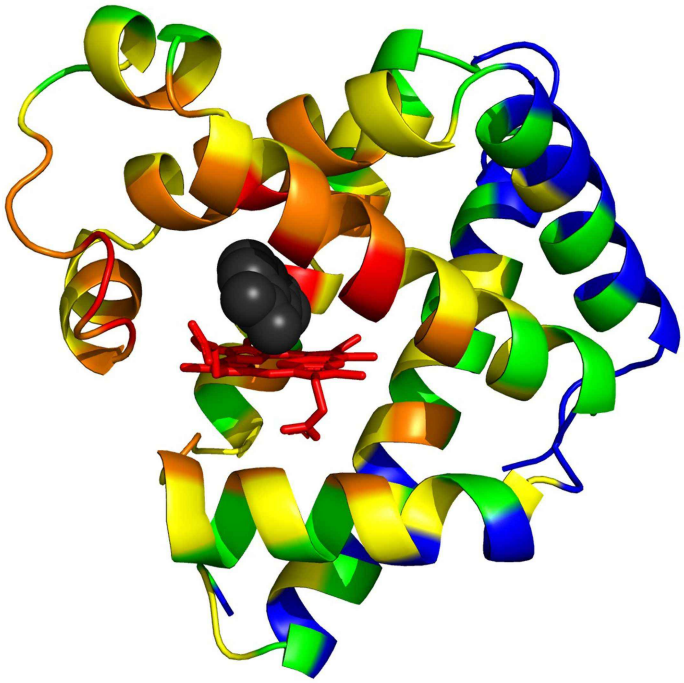
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05278-9.

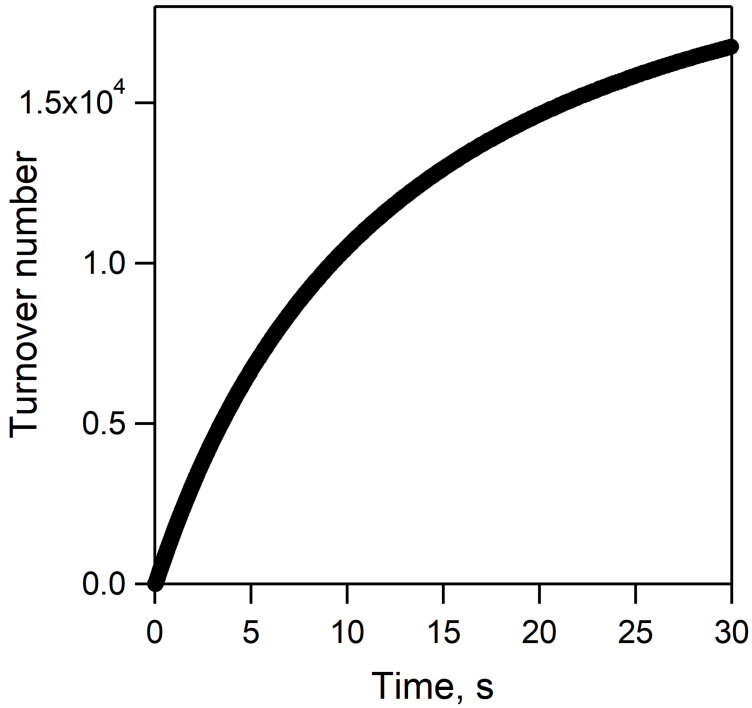
Correspondence and requests for materials should be addressed to Alexander N. Volkov, Olga V. Makhlynets or Ivan V. Korendovych.

Peer review information *Nature* thanks Giovanna Ghirlanda, Anthony Mittermaier and Jose Sanchez-Ruiz for their contribution to the peer review of this work.

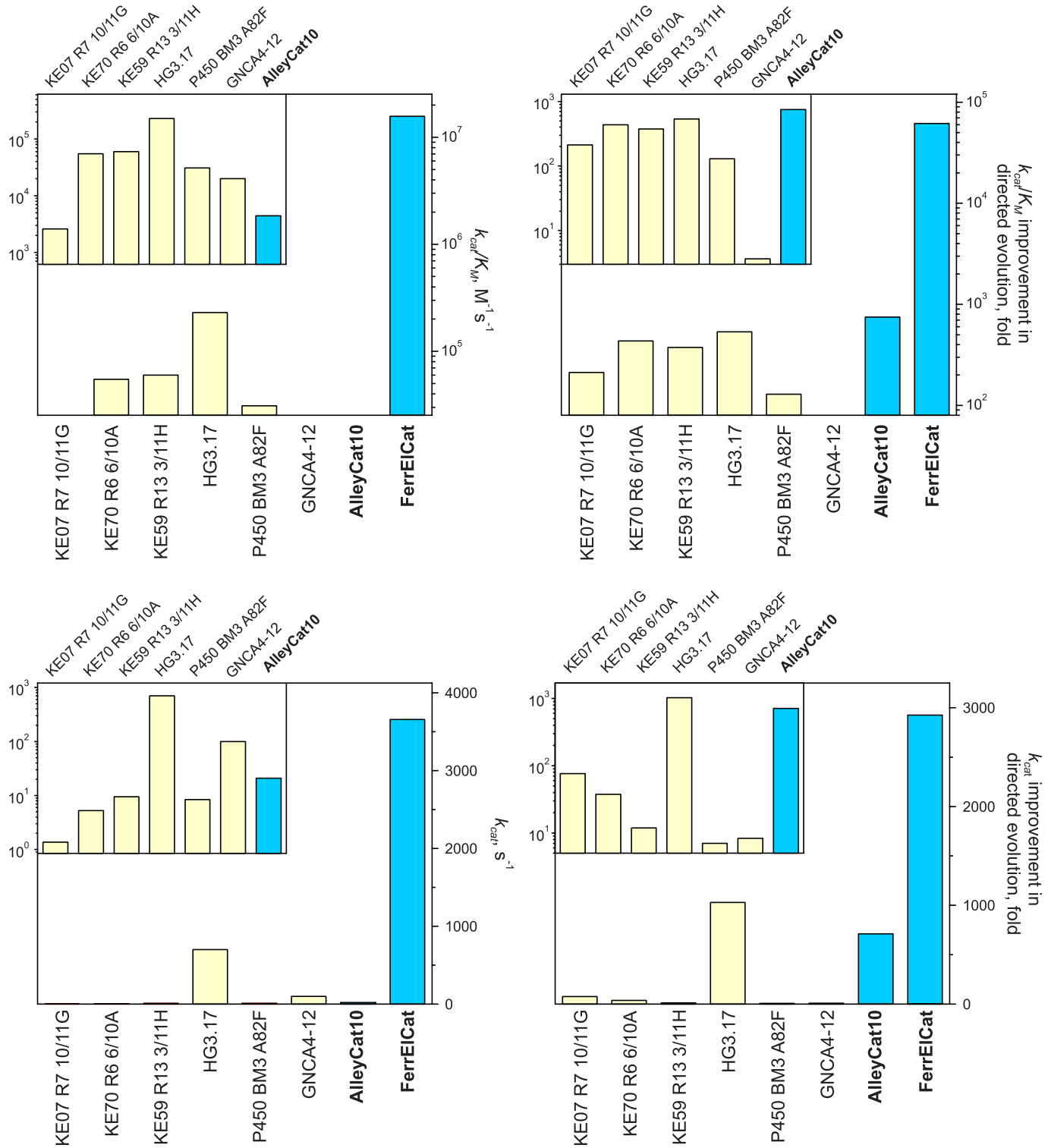


Bin 1	Bin 2	Bin 3		Bin 4		Bin 5
29	25	14	75	7	80	1
43	26	21	76	10	90	2
45	28	22	82	11	91	3
46	30	23	86	17	94	4
61	32	24	88	18	95	5
64	33	27	92	20	102	6
67	39	31	96	51	105	8
68	40	34	98	53	113	9
107	42	35	100	57	116	12
heme	44	36	101	73	117	13
	47	37	103	74	118	15
	48	38	106	77	119	16
	49	41	108	78	123	19
	55	50	109	85	128	25
	58	52	112	79	130	81
	60	54	114		132	83
	63	56	115		134	84
	65	59	131		136	87
	66	62	135		137	120
	71	69	139		141	121
	72	70	142		143	122
	89	73	146		145	124
	93		151		149	125
	97				150	126
	99				152	127
	104					129
	110					133
	111					140
	138					144
						147
						148
						153

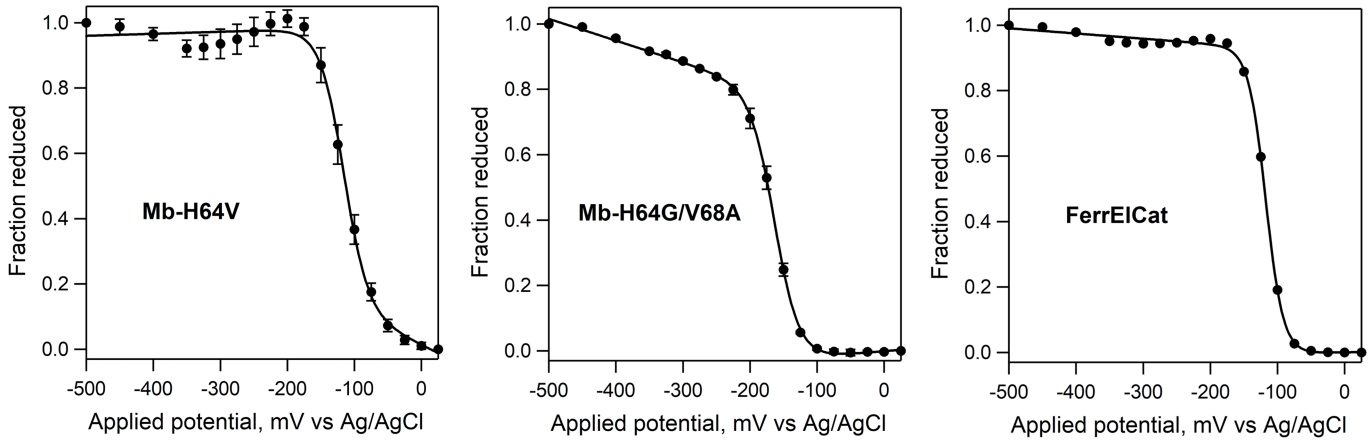
Extended Data Fig. 1 | Selection of negative controls for screening. Left. All residues in myoglobin were sorted in bins based on their distance to the docked inhibitor (black) in FerrElCat. The residues in the van der Waals contact with the docked inhibitor were placed in bin 1 (red), the residues in direct contact with the residues in bin 1 were placed in bin 2 (orange), etc. A total of five bins were devised: red, orange, yellow, green and blue. **Right**. The list of the residues sorted in the five bins. Residues showing large backbone CSP and their immediate neighbors are highlighted in red and yellow, respectively. Unassigned positions and residues immediately next to unassigned stretches are shown in dark grey and light grey, respectively. Prolines are highlighted in blue. Residues showing small CSP that were selected as controls are either highlighted in green or labeled in green font (when located next to unassigned residues).



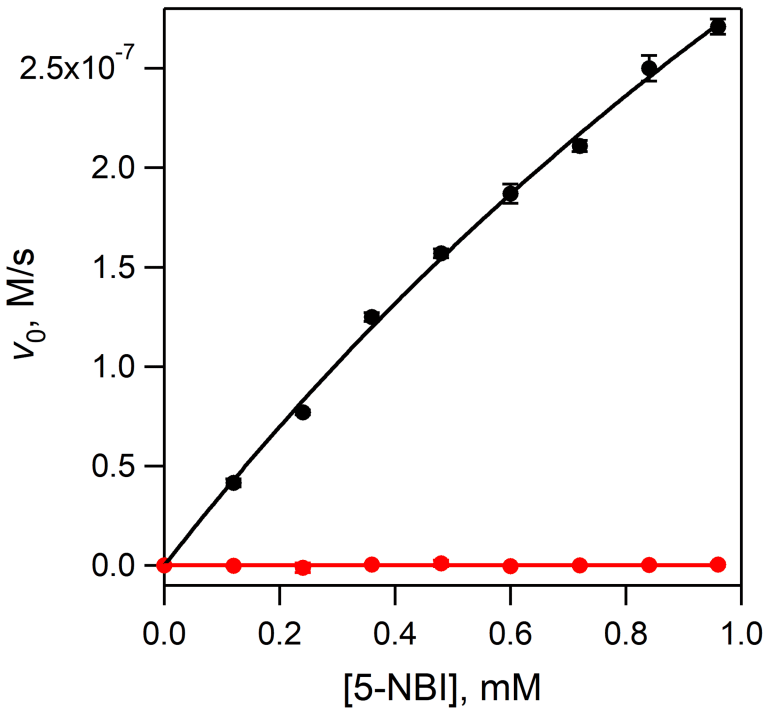
Extended Data Fig. 2 | Substrate turnover by reduced Mb(L29I/H64G/V68A) (FerrElCat). Reaction was monitored using stopped-flow at pH 8.0 for 30 s at 25 °C with 140 μM of 5-NBI and 5 nM of FerrElCat.



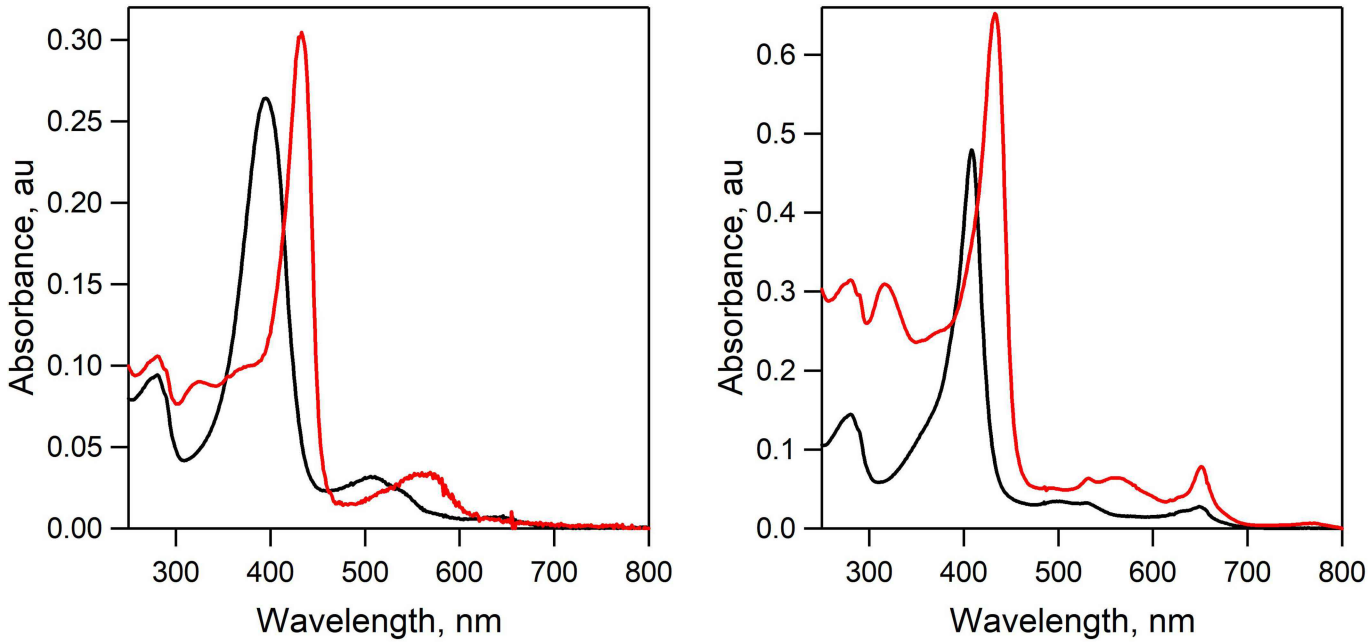
Extended Data Fig. 3 | Catalytic parameters of Kemp eliminases evolved using directed evolution. k_{cat}/K_M and k_{cat} values for the evolved enzymes (top and bottom left, respectively) and improvement in k_{cat}/K_M and k_{cat} achieved by directed evolution (top and bottom right, respectively). In cases where only k_{cat}/K_M was reported, we used 5 mM for K_M, to obtain low estimate of the k_{cat}.



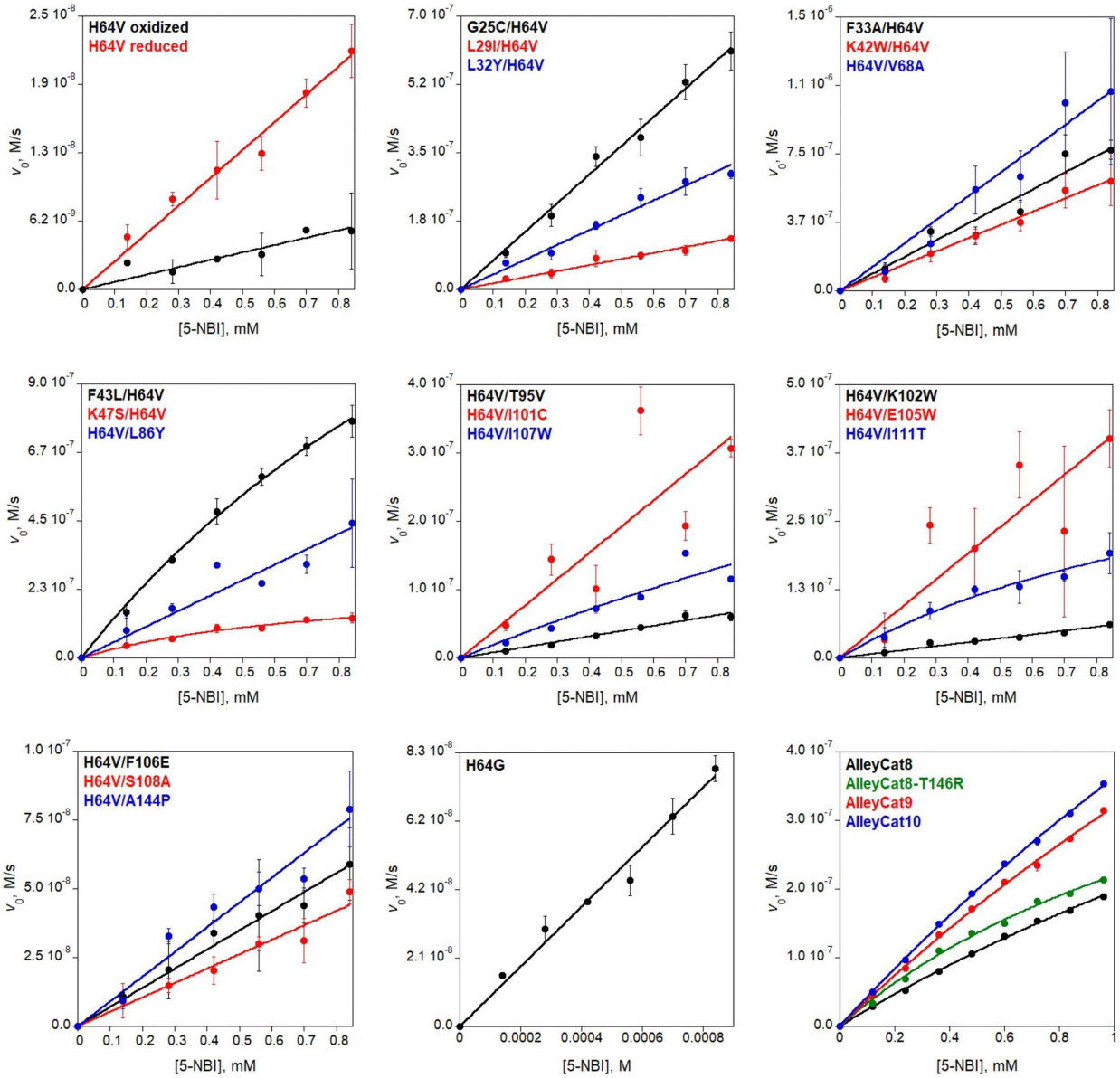
Extended Data Fig. 4 | Spectroelectrochemical determination of redox potentials of selected myoglobin mutants. Mb(H64V) (left), Mb(H64G/V68A) (middle) and FerrElCat (right). The proteins were analyzed in 20 mM Tris-HCl, pH 8.0 at 20 °C in presence of the mediator (100 μM phenazine sulfate). The redox potentials (vs Ag/AgCl) are summarized in Supplementary Table 5. Data are presented as the mean values and the error bars represent standard deviations obtained from three independent measurements for Mb(H64V) and Mb(H64G/V68A). For FerrElCat, the experiment was repeated twice with essentially identical results, one set of data is shown.



Extended Data Fig. 5 | Kemp elimination catalyzed by AlleyCat10 in presence of Ca^{2+} (black) and in absence of Ca^{2+} (red). The activity of 0.1 μM AlleyCat10 was tested with 0.12-0.96 mM substrate in 20 mM HEPES, 100 mM NaCl, pH 7.0 at 22 °C with 10 mM $CaCl_2$ (black) or 100 mM EDTA (red). Data are presented as the mean values and the error bars (most are small compared to the symbol size) represent standard deviations obtained from three independent measurements.



Extended Data Fig. 6 | Absorbance spectra of Mb(H64V) (left) and Mb(L29I/H64G/V68A) (FerrElCat) (right) in the oxidized (black) and reduced (red) forms. Spectra for the proteins (1.6 μM for Mb(H64V), 3.1 μM for FerrElCat) were collected in 20 mM HEPES, pH 7.0 in an anaerobic cuvette with a 1 cm pathlength at room temperature.



Extended Data Fig. 7 | Michaelis-Menten plots of Kemp elimination catalyzed by reduced (unless otherwise stated) myoglobin mutants and by AlleyCat proteins. Final reaction mixtures for myoglobin mutant analyses contained 1 mM L-ascorbic acid, 0.1 μM SOD, 20 nM catalase, 140-840 μM substrate, 1.5% acetonitrile in 20 mM Tris (pH 8.0). The protein concentration was 1 μM for Mb(H64V), 0.1 or 0.25 μM for Mb(H64V)-based double variants, 5 nM for Mb(H64G) or Mb(H64G)-based double or triple variants. For the AlleyCat proteins reaction mixtures contained 0.1 μM proteins with 0.12-0.96 mM substrate in 1.5% acetonitrile, 20 mM Tris, pH 8.0, 10 mM $CaCl_2$, 100 mM NaCl. Kinetic parameters are summarized in Table 1 and Extended Data Table 1. Data are presented as the mean values and the error bars represent standard deviations obtained from at least three independent measurements (six for Mb(H64G) and reduced Mb(H64V).

Extended Data Table 1 | Kinetic parameters for the Kemp elimination reaction catalyzed by myoglobin variants in the reduced state (unless otherwise stated) at pH 8.0

Protein	k_{cat}, s^{-1}	K_M, mM	k_{cat}/K_M, $M^{-1}s^{-1}$
H64V	-	-	255 ± 8
H64V (oxidized)	-	-	7 ± 1
G25C/H64V	-	-	7,354 ± 143
L29I/H64V	-	-	1,550 ± 55
L32Y/H64V	-	-	3,795 ± 122
F33A/H64V	-	-	9,270 ± 523
K42W/H64V	-	-	7,237 ± 192
F43L/H64V	26.10 ± 3.85	1.94 ± 0.38	13,458 ± 670
K47S/H64V	2.55 ± 0.39	0.79 ± 0.21	3,240 ± 994
H64V/V68A	-	-	12,939 ± 622
H64V/L86Y	-	-	5,097 ± 355
H64V/T95V	-	-	785 ± 32
H64V/I101C	-	-	3,844± 551
H64V/K102W	-	-	710 ± 28
H64V/E105W	-	-	4,786 ± 538
H64V/F106E	-	-	697 ± 22
H64V/I107W	-	-	1,680 ± 135
H64V/S108A	-	-	524 ± 24
H64V/I111T	-	-	2,356 ± 114
H64V/A144P	-	-	900 ± 43
H64G	-	-	18,152 ± 519
H64G/V68A	2,557 ± 372	1.28 ± 0.28	1,992,300 ± 143,420
L29I/H64G/V68A	3,656 ± 667	0.23 ± 0.13	15,721,000 ± 6,035,800

The k_{cat}, K_M and k_{cat}/K_M values were obtained from fitting the data presented in Extended Data Fig. 7 to the Michaelis-Menten equation.

Extended Data Table 2 | Sequences of the proteins used in this study

Protein	Sequence
Mb-H64V	MVLSEGEWQL VLHVWAKVEA DVAGHGQDIL IRLFKSHPET LEKFDRFKHL KTEAEMKASE DLKKVGVTVL TALGAILKKK GHHEAELKPL AQSHATKHKI PIKYLEFISE AIIHVLHSRH PGNFGADAQG AMNKALELFR KDIAAKYKEL GYQG
FerrElCat	MVLSEGEWQL VLHVWAKVEA DVAGHGQDII IRLFKSHPET LEKFDRFKHL KTEAEMKASE DLKKGGVTAL TALGAILKKK GHHEAELKPL AQSHATKHKI PIKYLEFISE AIIHVLHSRH PGNFGADAQG AMNKALELFR KDIAAKYKEL GYQG
AlleyCat	MKDTDSEEEI REAFRVEDKD GNGYISAAEL RHVMTNLGEK LTDEEVDEMI READIDGDGQ VNYEEFVQMM TAK
AlleyCat7	MKDTDSEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELI RETDIDGDGQ VNYEEFVQRM TAK
AlleyCat8	MKDTDSEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELH RETDIDGDGQ VNYEEFVQRM TAK
AlleyCat8-T146R	MKDTDSEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEK LTDEEVDELH RETDIDGDGQ VNYEEFVQRM RAK
AlleyCat9	MKDTDSEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEP LTDEEVDELH RETDIDGDGQ VNYEEFVQRM TAK
AlleyCat10	MKDTDSEEEL REQFRVEDKD GNGYISAAEL RIVMTNRGEP LTDEEVDELH RETDIDGDGQ VNYEEFVQRM RAK

Residues introduced into the native sequences in the course of design and directed evolution are shown in blue. The catalytic base in the AlleyCat family of proteins is shown in red.

Extended Data Table 3 | Crystallographic data collection and refinement statistics

	FerrElCat	AlleyCat9	AlleyCat9 with 6-NBT	AlleyCat10	AlleyCat10 with 6-NBT
Data collection					
Space group	P6	P212121	P43212	P212121	P43212
Cell dimensions					
a, *b*, *c* (Å)	90.02, 90.02, 45.37	27.38, 58.71, 87.51	82.97, 82.97, 104.07	27.31, 50.50, 88.25	82.92, 82.92, 104.95
a, b, c (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	22.68-1.40 (1.42-1.40) *	12.75-1.70 (1.73-1.70) *	12.98-1.70 (1.73-1.70) *	12.50-1.70 (1.73-1.70) *	13.19-1.95 (2.02-1.95) *
R_{sym} or R_{merge}		0.072	0.095	0.081	0.153
I / sI	15.5	13.6	15.8	10.0	11.2
Completeness (%)	99.8	99.3	99.7	99.5	99.7
Redundancy	6.9	4.9	9.8	4.5	10.6
Refinement					
Resolution (Å)	22.68-1.40	12.75-1.70	12.98-1.70	12.75-1.70	13.19-1.95
No. reflections	286602	16079	40502	13978	27277
R_{work} / R_{free}	17.7 / 19.5	17.9 / 21.4	21.4 / 25.2	19.7 / 24.5	24.9 / 27.8
No. atoms	1501	1181	4288	1221	2261
Protein	1218	1068	4030	1084	2114
Ligand/ion		5	64/15	6	48/17
Water	231	92	135	131	82
B-factors	12.0	12.0	20.3	12.0	16.0
Protein	10.7				
Ligand/ion					
Water	22.0				
R.m.s. deviations					
Bond lengths (Å)	0.005	0.006	0.009	0.006	0.006
Bond angles (°)	0.772	0.819	0.819	0.741	0.763

*Values in parentheses are for highest-resolution shell.

nature portfolio

Corresponding author(s): Korendovych, Makhlynets, Volkov

Last updated by author(s): Aug 3, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The NMR Data were acquired using Bruker TopSpin (version 3.6). Computational docking was performed using AutoDock Vina (version 1.1.2). Stopped-flow data were acquired and analyzed using Pro-Data Application (Applied Photophysics, version 2.5.1852.0). ITC data were obrained and analyzed using MicroCal PEAQ-ITC Software (version 1.21). Kinetic platereader data were acquired and analyzed using BioTek Gen5 Software (version 2.05.5)
Data analysis	The NMR data were processed in NMRPipe (version 5.97) and analyzed in CCPNMR (version 2.4.2). Diffraction data were processed with the CrysAlisPro software suite (version 1.171.39.46e, Rigaku). All structures were determined by Molecular replacement for crystallographic studies was done using PHASER (version 2.8.2). Refinement was performed by COOT (version 0.8.9.1) and PHENIX (version 1.17.1-3660). Unless explicitly described above, data were processed and analyzed using Microsoft Excel (version 16.59), KaleidaGraph (version 4.5.2, Synergy Software) and Igor Pro (version 8.4, Wavemetrics Inc.).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The crystallographic data and refinement statistics were deposited in the Worldwide Protein Data Bank (wwPDB) with the entry code 7vuc (FerrElCat), 7vur (AlleyCat9), 7vus (AlleyCat9 with 6-NBT), 7vut (AlleyCat10), 7vuu (AlleyCat10 with 6-NBT). The NMR chemical shift data are provided as part of Supplementary Information.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All members of the high CSP group were characterized as well as the residues in the negative control group that are close to the active site (as the probability of finding productive mutations is the highest there. For AlleyCat - nearly ALL positions in the negative cohort were characterized (excluding the residues in the EF-hands deemed to be critical for folding). For Mb-H64V, we created the negative size cohort to be slightly higher than the positive control one. Considering, that NONE of the residues in the negative control cohort (covering more than 15% of all available positions) produced improved mutants, the probability of it being representative is high (although calculations were not performed as classical probabilistic approximations may be not justified in this system).
Data exclusions	No data exclusion was done.
Replication	All kinetic data were run in triplicate on at least two different batches of proteins for all key proteins (Mb-H64V, Mb-H64G, Mb-H64G/H68A, FerrElCat) as well as representative proteins in the H64V series - Mb-64V/105W and Mb-64V/L86Y. All key proteins chosen for in depth characterization were studied by at least two different researchers. For the rest of the proteins the kinetic data were collected at least in triplicate on the same batch of protein. In all case all attempts to replicate data were successful. All proteins were independently expressed at least twice, showing the same purity and electrophoretic signatures. Spectroelectrochemical experiments were performed at least twice in every case, with excellent repeatability.
Randomization	The only group comparisons were between the residues with high CSP values (Z of ~1 and higher) and the rest. All members of the high CSP group were characterized as well as the residues in the other group that are close to the active site. For the rest the allocation was random.
Blinding	No blinding was done as for all of the experiments the researchers needed to know the identity of the protein they are working on to ensure protein quality, validity of data, etc.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging