Enhancing Luciferase Activity and Stability through Generative Modeling of Natural Enzyme Sequences

Wen Jun Xie [1,2,†,*], Dangliang Liu [3,†], Xiaoya Wang [3,†], Aoxuan Zhang [1], Qijia Wei [3], Ashim Nandi [1], Suwei Dong [3,*], Arieh Warshel [1,*]

[1]Department of Chemistry, University of Southern California, Los Angeles, CA, USA;
[2]Departmet of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development (CNPD3), Genetics Institute, University of Florida, Gainesville, FL, USA;
[3]State Key Laboratory of Natural and Biomimetic Drugs, Chemical Biology Center, and School of Pharmaceutical Sciences, Peking University, Beijing, China

* Correspondence: warshel@usc.edu, dongsw@pku.edu.cn, wenjunxie@ufl.edu
† Co-first authors

Abstract
The availability of natural protein sequences synergized with generative artificial intelligence (AI) provides new paradigms to create enzymes. Although active enzyme variants with numerous mutations have been produced using generative models, their performance often falls short compared to their wild-type (WT) counterparts. Additionally, in real-world applications, choosing fewer mutations that can rival the efficacy of extensive sequence alterations is usually more advantageous. Pinpointing beneficial single mutations continues to be a formidable task. In this study, using the generative maximum entropy model to analyze *Renilla* luciferase homologs, and in conjunction with biochemistry experiments, we demonstrated that natural evolutionary information could be used to predictively improve enzyme activity and stability by engineering the active center and protein scaffold, respectively. The success rate of designed single mutants is ~50% to improve either luciferase activity or stability. This finding highlights nature's ingenious approach to evolving proficient enzymes, wherein diverse evolutionary pressures are preferentially applied to distinct regions of the enzyme, ultimately culminating in an overall high performance. Our research also reveals an evolutionary preference in *Renilla* luciferase towards emitting blue light that holds advantages in terms of water penetration compared to other light spectrum. Taken together, our approach facilitates navigation through enzyme sequence space and offers effective strategies for computer-aided rational enzyme engineering.

Introduction

Nature has evolved enzymes as exceptional catalysts with fast turnover rates and precise specificity. Improving enzyme performance, including activity and stability, is highly desirable in various biological, medical, and industrial applications.[1] Rational enzyme engineering provides a solution

by generating variants of enzymes with improved properties, which requires a thorough understanding of enzymes and their vast variants.[2 3 4] Computational methods have been developed to study the performance of enzyme variants,[4 5] but accurately modeling the impact of mutations on enzyme catalytic power remains a significant challenge.[6] Therefore, further computational tools are necessary for rational enzyme engineering.

Machine learning offers novel ways of designing enzymes.[7 8 9] One such innovative approach involves the use of generative models to analyze evolutionarily-related protein sequences.[10 11 12 13 14] Generative models learn a probabilistic distribution of protein sequences evolved in nature, and the probability of a specific variant correlates with its readout in deep mutation scan experiments.[10 11] Such correlations suggest that generative models are capable of capturing the fitness of mutations during protein evolution. Furthermore, generative models hold great promise in exploring the functional sequence space of enzymes. For instance, such models have been applied to design functional variants of different enzymes.[15 16 17 18 19] In the majority of these enzyme engineering efforts, the tested sequences do not outperform the WT or natural enzymes, and even in instances where improvements were seen, a significant number of mutations have to be introduced. While these highly diverse variants with natural-like functions showcase the generative capacity of machine learning models, their practical application may face certain limitations. The introduction of too many mutations increases the cost, complexity, as well as causing poor solubility in some cases. In a clinical context, such extensively varied variants may elicit unforeseen immune responses and present ethical concerns, such as undermining genetic integrity.[20] A critical challenge in computational enzyme engineering remains: how can we improve the enzyme to a certain performance with a limited number of mutations? [21]

Beyond the realm of enzyme engineering, delving into natural sequences may offer insights into the intricate world of enzymes. The newfound knowledge can further refine our approaches to enzyme engineering. It is well-acknowledged that both structural and functional constraints can influence evolutionary rates across protein sites.[21] Enzymes grapple with a myriad of evolutionary pressures during their evolutionary journey. These pressures include enhancing their activity and ensuring their stability. Such pressures play a pivotal role in shaping enzyme evolution. For instance, mutations that enhance enzyme efficacy may be evolutionarily favored, finding their place in contemporary enzyme homologs. By harnessing generative AI models to extract evolutionary metrics from natural sequences, we can establish links between these metrics and the physicochemical properties of enzymes.

Through data mining of enzyme homologs with a generative maximum entropy (MaxEnt) model, we have recently illuminated the intricate evolution-catalysis relationship. Specifically, we have shed light on how enzyme sequences evolving in nature are shaped by their physicochemical characteristics, such as activity and stability, both of which are deeply intertwined with enzyme catalysis.[22] The MaxEnt model takes into account pairwise epistasis and assigns statistical energy

for each protein or mutant sequence in the multiple sequence alignment (MSA) as the evolutionary metric. For enzymes that catalyze diverse chemical reaction classes, the statistical energy corresponds with enzyme activity when the mutation is in the active center, and with stability when the mutation is situated within the scaffold. Thus, our investigations suggested key regions within enzymes that are primarily influenced by distinct evolutionary pressures.[22] Evolution seems to prioritize the enzyme's active center, the hub of chemical activity, and its scaffold for enhancing activity and ensuring stability, respectively. Nonetheless, the practical applicability of these theoretical insights in engineering high-performance enzymes still remains an area of exploration. Furthermore, the degree of success such an approach might yield is yet to be ascertained.

In this study, Renilla luciferase (RLuc) was selected as our testbed to further scrutinize the evolution-catalysis relationship, and to assess its utility in enzyme engineering. Luciferases represent a class of enzymes that catalyze the oxidation of luciferin, subsequently resulting in bioluminescent emission (Fig 1A-B).[23 24 25] Distinctly more complex than other enzymes we have previously investigated,[22 26] luciferases are subject to a diverse array of evolutionary pressures, encompassing factors such as catalytic efficiency, protein stability, and bioluminescence. This complexity transforms the engineering of luciferase into a multifaceted and intricate endeavor. Notably, our findings indicate that harnessing the evolution-catalysis relationship can inform rational luciferase engineering, leading to a success rate of approximately 50% in enhancing either activity or stability via the introduction of a single mutation. Beyond successful enzyme engineering, our study also revealed that RLuc has evolved to emit blue light, which has superior water penetration compared to other regions of the light spectrum. Collectively, this research underscores the potential of evolutionary data for effectively enhancing biological activity.

Results

Generative modeling of luciferase evolution

Generative models analyze evolutionarily-related protein sequences and provide a probabilistic model to capture the effects of mutations in natural evolution. The probability associated with a certain sequence or variant has been connected to protein fitness.[10 11] Among these models, the MaxEnt model, which is based on information theory, has demonstrated a strong generative capacity.[27] The statistical energy of sequence ($\boldsymbol{S}$) as derived from the model is expressed as

$$E_{\mathrm{MaxEnt}}(\boldsymbol{S}) = \sum_i h_i S_i + \sum_{i>j} J_{ij} S_i S_j, \qquad \text{(Equation 1)}$$

where h_i and J_{ij} are parameters, $S_{i/j}$ is amino acid at residue position i/j in the sequence. The $E_{\mathrm{MaxEnt}}(\boldsymbol{S})$ follows the Boltzmann distribution

$$P(\boldsymbol{S}) = e^{-E_{\mathrm{MaxEnt}}(\boldsymbol{s})}/Z, \qquad \text{(Equation 2)}$$

where $Z = \sum_{\boldsymbol{S}} e^{-E_{\mathrm{MaxEnt}}(\boldsymbol{s})}$ is the normalization constant. The derivation and parameterization details of MaxEnt for enzymes can also be found in Ref [22]. A lower E_{MaxEnt} value for a variant suggests a higher probability of survival in the natural environment, indicating certain evolutionary advantages. In this context, the MaxEnt model learned from MSA describes an energy landscape as shown in the schematic picture (Fig 1C).

The complex nature of enzymes, with intertwined evolutionary pressures, means that a direct correlation between E_{MaxEnt} and individual parameters is not assured. This intricacy stems from the interplay of various physicochemical properties that govern enzyme function, and the multifaceted relationship between specific enzyme activities and protein thermostability.[28–30] Such a convoluted landscape suggests that relying solely on E_{MaxEnt} as a predictor for enzymes might be oversimplistic. However, drawing connections from our investigations on multiple enzymes, we have discerned that E_{MaxEnt} (or $P(\boldsymbol{S})$) for variants with mutations in the active center and enzyme scaffold tend to correlate with enzyme activity and stability, respectively.[22] This connection between evolution and catalysis suggested a structured approach to enhance various enzyme properties by fine-tuning E_{MaxEnt}, particularly targeting distinct enzyme regions (Fig 1C).

Here we applied the MaxEnt model to luciferase for enzyme engineering. We used RLuc from *Renilla reniformis* (UniProt ID: P27652) as the target sequence and searched the UniProt database (release 2021_03) to construct the MSA. A length-normalized bit score threshold of 0.7 was applied. We obtained 1775 homologous sequences, which is sufficient to parameterize the model. Parameters were optimized using statistics obtained from the MSA, including the probability of a specific amino acid at a single residue and the probability of a specific pair of amino acids between two residues. Once parameterized, the statistical energy E_{MaxEnt} for each sequence or variant was calculated. The statistical energy was shifted so that the WT has a value of zero, which does not affect any analysis here. We then correlated E_{MaxEnt} with available biochemical properties of luciferase. This correlation aimed to elucidate the association between evolution and catalysis, which could potentially shed light on nature's strategy for generating luciferase.

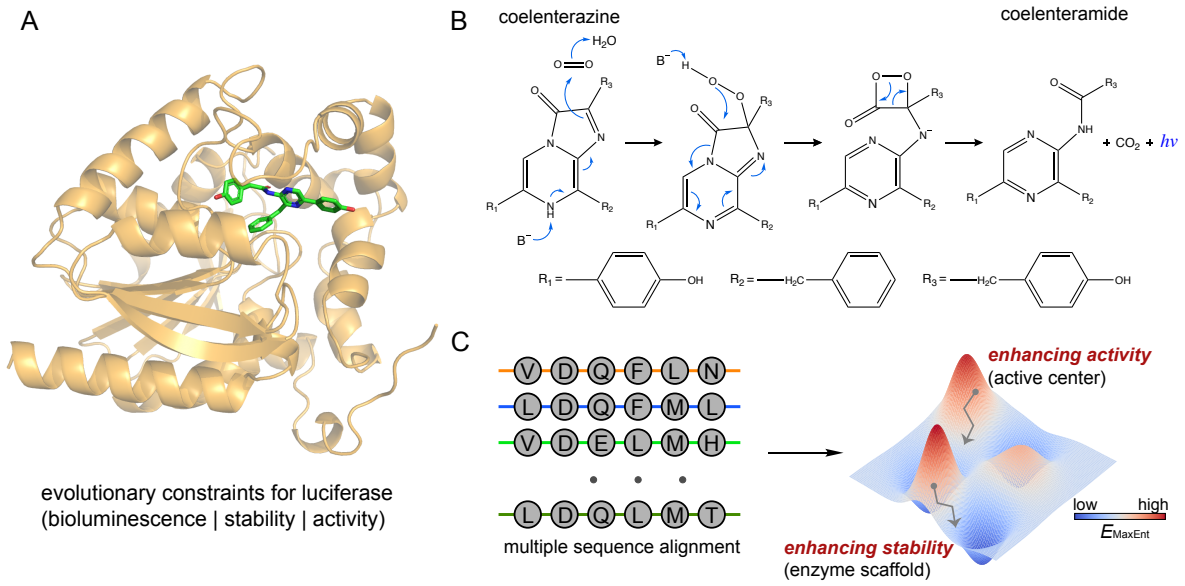

Fig 1. The generative maximum entropy model for luciferase. (A) Structure of RLuc with coelenterazine as luciferin. PDB ID used in rendering the structure is 6YN2 with substrate docked. (B) Scheme of the chemical reaction catalyzed by RLuc. Coelenterazine undergoes oxidation to coelenteramide with oxygen acting as the oxidant, and a photon of blue light is released as a result. (C) Rationally enhancing enzyme activity/stability guided by the generative MaxEnt model. The statistical energy E_{MaxEnt} as an evolutionary metric reflects evolutionary pressures. For variants with mutated residues in the active center and enzyme scaffold, E_{MaxEnt} highly correlates with enzyme activity and stability, respectively.[22] Thus, enzyme performance can be improved by optimization on the evolutionary landscape.

Correlating luciferase activity with natural evolutionary information

We first examined the correlation between luciferase activity and E_{MaxEnt}. RLuc has been extensively engineered as a crucial reporter in live-cell imaging, with 26 variants identified through consensus design.[31 32] These variants primarily consist of double mutations, with C124A being present in all of them (Table S1). The average distance between the substrate and the mutated residues was determined for each variant, and the variants were categorized as belonging to the active center or enzyme scaffold based on an 8.5 Å cutoff.[22] For variants classified as the active center, their activity has a significant correlation with E_{MaxEnt}, as determined by a Pearson correlation value of -0.69 (p-value = 0.057), as seen from Fig 2A. However, this correlation noticeably declines for variants with mutations on the enzyme scaffold, with a Pearson correlation of -0.25 (p-value = 0.29). It should be noted that the luminometers used in the measurement were calibrated to absolute units (photons/s),[31 32] which implies that the correlation established pertains to enzyme activity and not bioluminescence.

These findings indicate that enzyme activity considerably shapes the evolution of active center compared to enzyme scaffold. The 8.5 Å cutoff value approximately distinguishes two residue shells from the substrate and places them within the active center. The first-shell residues around substrate are directly involved in substrate positioning and catalyzing reactions. There is also an increasing body of evidence to suggest that second-shell residues play a key role in regulating enzyme activity.[33] Thus, residues within the active center have a direct impact on enzyme catalysis and are expected to be mainly shaped by biological activity during natural evolution.[22]

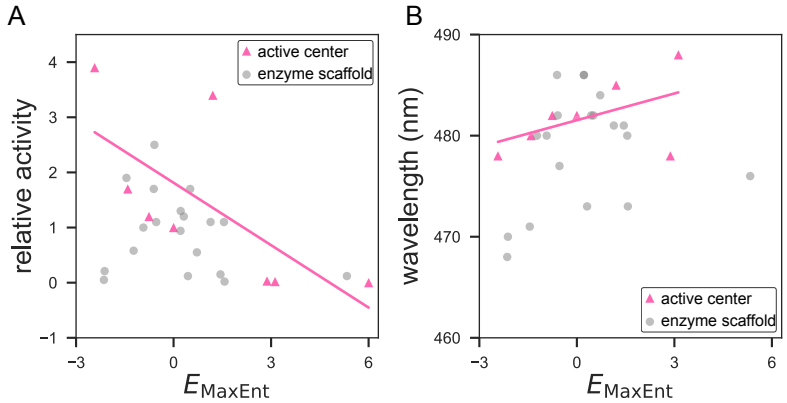

Fig 2. Correlations between the statistical energy E_{MaxEnt} and biological activity for luciferase. (A) Luciferase activity relative to WT. (B) Emission peak of luciferase bioluminescence.

Intrigued by the patterns observed in catalytic efficiency, we delved deeper to discern if natural evolutionary data could also shed light on the bioluminescent attributes of luciferase. This phenomenon of bioluminescence is not just a standalone feature; it is an outcome of evolutionary pressures acting on the luciferase-mediated reactions. Parameters like emission spectrum, inactivation time, among others, play pivotal roles in shaping bioluminescence (Table S1). Our analysis revealed a notable correlation of 0.51 (p-value = 0.24) between E_{MaxEnt} and the peak emission spectrum for mutations localized in the active center, with a more modest correlation of 0.22 (p-value = 0.38) in the enzyme scaffold region, as seen from Fig 2. Tackling the prediction of bioluminescence spectra is notoriously tricky in computational chemistry, given it demands insight into excited states. To our knowledge, no study till date has proposed a correlation metric adept at linking bioluminescence spectra across diverse enzyme variants. This observed positive correlation hints at a natural evolutionary tendency in RLuc to favor blue light emission, possibly because of blue light's superior water penetration capabilities compared to other wavelengths (Fig S1).[34 35] In a related vein, nature seems to exhibit a bias towards enzymes exhibiting prolonged inactivation phases (Fig S2). Hence, E_{MaxEnt}, in its role as an evolutionary barometer, provides a potential means to link with an array of evolutionary pressures, targeting specific protein domains.

Besides the native reaction that converts coelenterazine, we also investigated how sensitive this approach is to variations in the reacting substrate. To optimize the spectrum for live-cell imaging, the substrate has been slightly modified. The correlation between E_{MaxEnt} and enzyme activity is much weaker for some modified substrates (Fig S3). Thus, the natural sequences evolved for converting coelenterazine are sensitive to the specific substrates.

Furthermore, we extended our analysis to mutations made to RLuc8, which is an engineered variant of RLuc to have eight-point mutations (Table S2). The findings underscored that the evolutionary metric, E_{MaxEnt}, adeptly differentiates active enzymes from inactive ones, as

demonstrated in Fig S4. In a more detailed assessment, we analyzed saturation mutagenesis data at the active site residue I223 of RLuc8 (Table S3). A notable correlation of -0.71 (p-value < 0.001) between E_{MaxEnt} and enzyme activity emerged, aligning with observations from Fig 2A, as seen in Fig S5. Additionally, we observed that the mutations at this residue tend to have similar effects on converting substrate analogs (Fig S6).

Low-dimensional embedding of generated variants

In the context of enzyme design, investigating the E_{MaxEnt} energy landscape enables the identification of low-energy sequences that have the potential to enhance enzyme performance, as indicated from our previous analysis.[22] Instead of sampling the energy landscape, we systematically catalog all possible single and double mutants situated in the enzyme's active center and scaffold. By employing this straightforward strategy, we guarantee the comprehensive consideration of all lower-order mutations, which forms the central focus of the subsequent proof-of-concept biochemistry characterization. The sequences with E_{MaxEnt} lower than the WT were collected. We obtained 220 such variants with mutated residues within 8.5 Å of the substrate and 394 variants with mutated residues more than 15.0 Å from the substrate (Table S4-S5). The generative MaxEnt model provides statistical energy for any given sequence, which has been shown to be correlated with luciferase biological activity depicted in the above section.

We used variational autoencoder (VAE) to perform the low-dimensional embedding to characterize the relation between generated variants and natural luciferase sequences (SI Methods). We embedded the sequences in a two-dimensional space using a VAE trained with natural sequences.[36] As illustrated in Fig 3A, the natural sequences are organized into different spikes in the latent space, which is also found in previous studies, potentially related to phylogeny in evolution.[37] The variants created through redesigning the active center or enzyme scaffold are confined to a specific local region in the latent space due to the focus on single or double mutants that are similar to the WT sequence (Fig 3B and 3C). These variants for the active center and enzyme scaffold occupy different regions in the latent space, close to different spikes in the latent space.

In the subsequent experiment, we selected eight single variants with the lowest E_{MaxEnt} values in the active center for experimental characterization. These variants are C124V, M185L, A143M, K189S, C124A, K189G, K189A, and C124G, listed in order of increasing suggestions. These mutations span four residues, all of which were successfully expressed (Fig S7). Additionally, we selected another set of eight single variants on the enzyme scaffold that also had the lowest E_{MaxEnt} values: I34M, I75A, E195T, F33K, V64I, N35S, Y298A, and V212S, again listed by increasing suggestions. However, Y298A and E195T mutants exhibited notably lower protein yield, and thus their activity was not measured, as indicated in Fig S7. The mutated residues in the active center and on the enzyme scaffold, which were characterized in the experiment, are highlighted in Figs 3D and 3E, respectively.

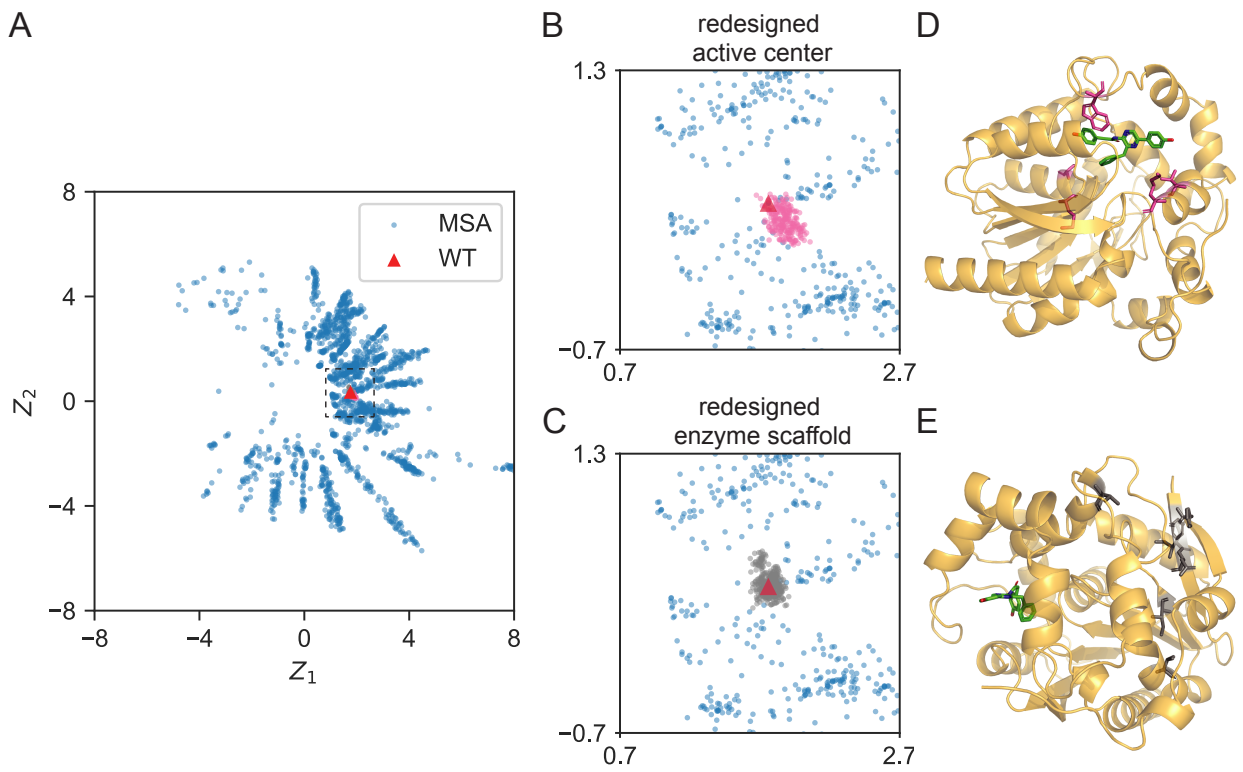

Fig 3. Organization of luciferase sequences in latent space. (A) Embedding MSA and designed sequences in a two-dimensional latent space with VAE. (B, C) Sequences featuring a redesigned active center (B) are shown in pink, while those with a redesigned enzyme scaffold (C) are depicted in gray. (D, E) Variants that underwent testing in the experiment have their mutated residues emphasized: those in the active center (D) are colored in pink, and those on the enzyme scaffold (E) are presented in gray.

Enhancing Luciferase Activity using Natural Evolutionary Information Validated in Experiments

For mutants in the active center, the relative activity compared to the WT enzyme was shown in Fig 4A. The activity of RLuc variants was determined with luminometers (as described in the SI Methods). Our results indicate that mutations in the active center have a significant impact on catalysis. Of the eight designs, half showed lowered performance compared to the WT activity. Notably, the A143M mutant displayed a relative activity of 1.02, and M185L, C124A, and C124G, showed notable increases in activity, with 2.01, 1.63, and 1.72 times relative activity to WT, respectively. Overall, about 50% of our attempts to design beneficial mutations in the active center works.

It is noteworthy that the C124A mutation was previously identified as a potential target for mutation due to concerns that the cysteine residue may be prone to oxidation and impair enzyme activity.[38] This mutation was subsequently used in different luciferase engineering projects.[31, 32] Our approach also identifies C124G as a beneficial mutation. However, it would be valuable to

further investigate the chemical mechanisms underlying the observed effects, as C124V, another mutation on the same residue results in an inactive luciferase. Our method also suggested that the M185L mutation could improve catalysis. The methionine residue at residue 185, located close to the substrate, may be able to adjust its orientation to enhance activity. This is in line with previous findings, which identified a beneficial valine mutation in this location.[31] However, we found that three proposed K189 mutations were not effective in practice. Enzyme design is indeed a challenging field, and it is extremely hard to understand why certain designs fail.[15 16 17 18 19] Additional investigations are necessary to comprehend the molecular intricacies underlying results present here, which extends beyond the scope of the generative AI studies conducted in this work.

For mutations on the enzyme scaffold, all six expressed variants show slightly decreased activity (Fig 4B). The discovery is fascinating and aligns with our earlier conclusion that the primary evolutionary driving force for the enzyme scaffold is not centered on improving enzyme efficiency.[22] Changes to the enzyme scaffold have a less pronounced effect on activity than changes in the active center, further supporting the idea that enzyme activity drives the evolution of residues near the substrate. Together, our computational and experimental results provide strong support for using natural evolutionary information to improve activity by focusing on the active center rather than other regions.

Furthermore, we also analyzed the emission spectra for each design (Fig S8). The shift in the spectrum caused by mutations to the active center is more evident compared to those on the enzyme scaffold. Given that the spectrum is associated with the enzyme-catalyzed reaction rather than protein stability, the results are consistent and logical.

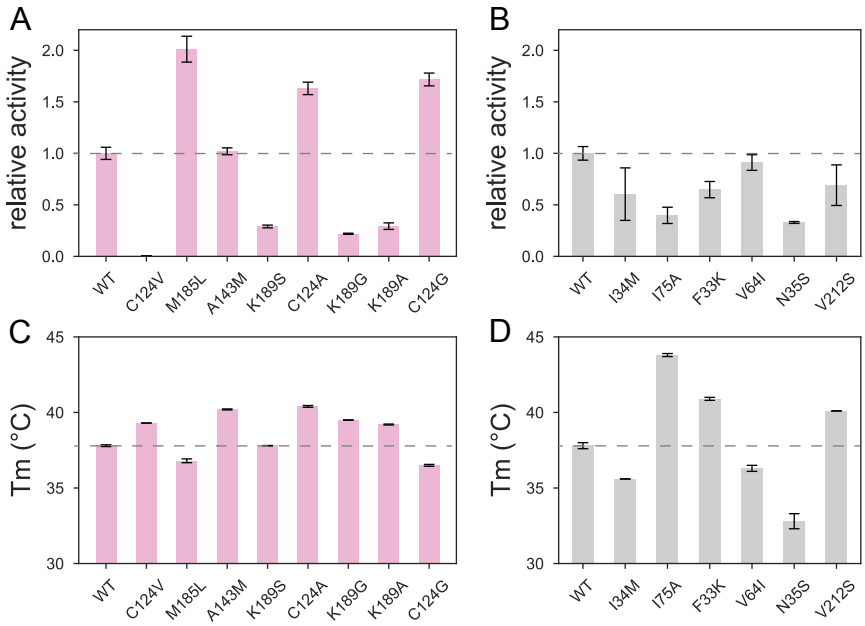

Fig 4. Characterization of designed variants. (A, B) Relative activity for variants mutated in the active center (A) and on the enzyme scaffold (B). (C, D) Melting temperature for variants mutated in the active center (C) and on the enzyme scaffold (D). Values of relative activity and melting temperature used in the plot were given in Table S6 and S7, respectively.

Enhancing Luciferase Stability using Natural Evolutionary Information Validated in Experiments

The stability of an enzyme is a crucial property. We have previously observed that the enzyme scaffold is primarily shaped by evolutionary pressure to maintain stability for some enzymes.[22] Unfortunately, we do not possess stability data for luciferase to determine its correlation with E_{MaxEnt} across different variants. Thus, it will be a crucial test of the evolution-stability correlation by measuring the melting temperature of variants designed using natural evolutionary information.

The mutations in the active center suggested by our method are expected to mainly improve biological activity rather than stability, as the active center is primarily shaped by chemical reactions. Therefore, it is not expected that these mutations will have a significant effect on stability. To verify this, we measured the stability of eight variants in the active center using differential scanning fluorimetry assay (Fig S9). In line with our expections, compared to the WT enzyme, they all have similar melting temperatures (T_m) (Fig 4C) and the change in T_m ranges from -1.3°C to 2.4°C.

Conversely, mutations on the scaffold of the enzyme can significantly affect the stability of RLuc, causing T_m to vary between -5.0°C and 6.0°C (Fig 4D). Our method successfully identified three out of six mutations that led to an increase in stability, resulting in a 50% success rate. Additionally, mutations on the enzyme scaffold showed a wider range of T_m compared to those at the active center. These stability results suggest that natural evolutionary information can be utilized to design stabilizing mutations away from the active regions.

Comparison of Enzyme Engineering Efforts Utilizing Generative Models

Generative models, trained on natural homologous sequences, have shown potential in generating functional sequences akin to those found in nature. [15] [16] [17] [18] Despite these advancements, engineering enzymes to exceed the performance of their WT counterparts remains a considerable challenge. As outlined in Table 1, recent efforts have employed various generative models capable of producing functional enzyme sequences. However, these experimentally tested variants do not surpass the catalytic activity of WT enzymes in most studies. Here we succeeded in designing variants with enhanced enzyme activity, with a success rate close to 50%.

On a different note, in Ref [19], another generative model was employed which facilitated the design of variants with enhanced activity through the incorporation of multiple mutations, realizing a maximum fold increase of 2.5. Notably, our approach enabled us to introduce a single mutation, leading to a fold increase of 2.0, a result that is on par with those achieved by the introduction of

multiple mutations in Ref [19]. In practical scenarios, fewer mutations are favorable as they minimize experimental effort and potential ethical issues.[20] Our approach thus marks an improvement as it identifies single mutations to enhance enzyme efficiency in comparison to WT. It can be attributed to our establishment of the evolution-catalysis relationship using generative models, which provides clues to effectively harness evolutionary information for rational enzyme engineering.

Table 1. Comparison of generative enzyme engineering studies

Enzyme	Generative Model	Mutant Type	Functional Assay Comment	Activity Assay Comment
Luciferase RLuc	MaxEnt (this study)	Single		4 out of 8 designs show enhanced activity compared to WT; ***maximum fold increase: 2.0***
Chorismate Mutase	MaxEnt [15]	Multiple	481 out of 1618 are functional	5 tested designs exhibit comparable activity to natural enzymes
Malate Dehydrogenase	GAN [16]	Multiple	13 out of 55 are functional	
Luciferase LuxA	VAE [17]	Multiple	9 out of 11 are functional	
Lysozyme	Language [18]	Multiple		6 tested designs exhibit comparable activity compared to WT
Ornithine Transcarbamylase	VAE [19]	Multiple		58 out of 87 designs show enhanced activity compared to WT; ***maximum fold increase: 2.5***

*GAN: Generative Adversarial Network; VAE: Variational Autoencoder; Language: Language Model

Discussions

This study reinforces our previously proposed connection between evolution and catalysis. Our suggested mutations, derived from evolutionary data, successfully enhance enzyme activity in the active center and bolster stability within the protein scaffold. While mutations in the active center introduce minor changes to enzyme stability, all the recommended scaffold mutations here resulted in reduced enzyme activity.

The engineering success underscores the potential of generative models to enhance enzyme performance. Nonetheless, we did not achieve a flawless success rate in designing beneficial single variants. This could be due to limitations in classifying residues solely based on their distance to the substrate, an incomplete categorization of the enzyme architecture, a current lack of understanding of the complexities of luciferase evolution, etc. Beyond single variants, it is essential to examine the maximal fold increase by assessing higher-order variants experimentally.

Combining advantageous single mutations with other neutral ones could potentially produce many improved diverse variants, potentially adding complexity to pinpointing the optimal higher-order ones. It remains to be seen whether this applies to Ref [19], especially since its fold increase mirrors that of a single mutant.

The MaxEnt model is based on the construction of protein MSA. Meanwhile, the protein language model, trained on millions of natural sequences, might discern common protein patterns. Like the MaxEnt model, this language model is generative. A question then arises: how well would a language model elucidate the evolution-catalysis relationship for luciferase, compared to the MaxEnt model? To this end, we utilized ESM-1v,[39] trained on 98 million natural sequences, to rank luciferase variants with mutations proximate to the substrate. The correlation derived was somewhat inferior to the MaxEnt model that leverages an MSA (Fig S10). Several explanations are plausible. Firstly, the variants are predominantly double mutants; the scoring function in ESM-1v treats mutated positions in isolation, overlooking epistasis. Secondly, the myriad of enzyme-catalyzed reactions suggests diverse mechanisms. Extracting universally applicable rules might be challenging. Hence, interspersing information across different protein classes might not notably bolster the prediction of enzyme activity. However, to definitively ascertain whether a protein language model surpasses the MSA-based generative model's efficacy, comprehensive evaluations across larger datasets are imperative.

We now turn our attention to the connection between designs derived from generative models and consensus design. In consensus design, the most frequent residue type within a protein family is selected.[40,41] We showcased the sequence logo for the mutated positions in our experiment in Fig S11, emphasizing the conservation of sequences at these positions. Some single variants proposed by the MaxEnt model align with the consensus residue at specific positions. For instance, for mutations within the active center, A143M selects the consensus methionine. Yet, the consensus residues at positions 124, 185, and 189 are not the best choices as per the MaxEnt model. Concerning mutations on the scaffold, five designs, namely I34M, I75A, V64I, Y298A, and V212S, align with consensus design. This consistency is anticipated since a MaxEnt model, when not accounting for epistasis, independently evaluates each position using the statistical energy $E_{\mathrm{MaxEnt}}(\boldsymbol{S}) = \sum_i h_i S_i$ and tends to favor the consensus residue. Yet, the evolution-catalysis relationship provides rationale for why consensus design might be effective.

The brighter luciferase could have applications in single-cell imaging, and evolutionary constraints have driven the evolution of blue light in RLuc, whereas red-shifted light is optimal for deep tissue penetration. It is notable that modifications to the substrate, rather than enzyme engineering, have largely been adopted to tune RLuc's spectrum. Additionally, luciferase has evolved independently multiple times on Earth, such as in fireflies that produce fluorescence with red light which is different from RLuc. The specific reasons for these different evolutionary trends remains a mystery, and further effort is needed to better understand the factors that drive the evolution of bioluminescence and luciferase activity in different organisms.

The remarkable catalytic properties of enzymes evolved by nature serve as a valuable source of inspiration for protein engineering. However, there is still much to be learned about the underlying principles of natural evolution that can inform rational enzyme engineering. Our study contributes to the advancement of this understanding and provides new insights into rational enzyme design.

Acknowledgment
A.W. is supported by the National Institutes of Health R35 GM122472 and the National Science Foundation Grant MCB 1707167. S.D. is supported by the National Natural Science Foundation of China (Nos. 22177004 and 92153301). We thank the University of Southern California High Performance Computing and Communication Center for computational resources. We thank Tianmin Fu, Zhenghan Liao, and Xiaoyu Chen for their insightful discussions.

Competing interests
None.

Additional information
Supplementary information and dataset are available for this paper.

Data Availability
The multiple sequence alignments are uploaded as supporting files. The calculated statistical energies for mutants are reported in the supplementary information.

References
1. Goldsmith, M. & Tawfik, D. S. Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr Opin Struct Biol* **47**, 140–150 (2017).
2. Steiner, K. & Schwab, H. Recent advances in rational approaches for enzyme engineering. *Comput Struct Biotechnol J* **2**, e201209010 (2012).
3. Cedrone, F., Ménez, A. & Quéméneur, E. Tailoring new enzyme functions by rational redesign. *Curr Opin Struct Biol* **10**, 405–410 (2000).
4. Ferreira, P., Fernandes, P. A. & Ramos, M. J. Modern computational methods for rational enzyme engineering. *Chem Catalysis* **2**, 2481–2498 (2022).
5. Warshel, A. Multiscale modeling of biological functions: From enzymes to molecular machines (nobel lecture). *Angew Chem Int Ed* **53**, 10020–10031 (2014).
6. Jindal, G. *et al.* Exploring the challenges of computational enzyme design by rebuilding the active site of a dehalogenase. *Proc Natl Acad Sci U S A* **116**, 389–394 (2019).
7. Feehan, R., Montezano, D. & Slusky, J. S. G. Machine learning for enzyme engineering, selection and design. *Protein Eng Des Sel* **34**, 1–10 (2021).
8. Mazurenko, S., Prokop, Z. & Damborsky, J. Machine Learning in Enzyme Engineering. *ACS Catal* **10**, 1210–1223 (2020).
9. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat Methods* **16**, 687–694 (2019).

10. Hopf, T. A. *et al.* Mutation effects predicted from sequence co-variation. *Nat Biotechnol* **35**, 128–135 (2017).
11. Figliuzzi, M., Jacquier, H., Schug, A., Tenaillon, O. & Weigt, M. Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase tem-1. *Mol Biol Evol* **33**, 268–280 (2016).
12. Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proc Natl Acad Sci U S A* **110**, 15674–15679 (2013).
13. Marks, D. S. *et al.* Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, e28766 (2011).
14. Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* **108**, E1293–E1301 (2011).
15. Russ, W. P. *et al.* An evolution-based model for designing chorismate mutase enzymes. *Science* **369**, 440–445 (2020).
16. Repecka, D. *et al.* Expanding functional protein sequence spaces using generative adversarial networks. *Nat Mach Intell* **3**, 324–333 (2021).
17. Hawkins-Hooker, A. *et al.* Generating functional protein variants with variational autoencoders. *PLoS Comput Biol* **17**, e1008736 (2021).
18. Madani, A. *et al.* Large language models generate functional protein sequences across diverse families. *Nat Biotechnol* **41**, 1099–1106 (2023).
19. Giessel, A. *et al.* Therapeutic enzyme engineering using a generative neural network. *Sci Rep* **12**, 1–17 (2022).
20. Brooks, D. A., Kakavanos, R. & Hopwood, J. J. Significance of immune response to enzyme-replacement therapy for patients with a lysosomal storage disorder. *Trends Mol Med* **9**, 450–453 (2003).
21. Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nat Rev Genet* **17**, 109–121 (2016).
22. Xie, W. J., Asadi, M. & Warshel, A. Enhancing computational enzyme design by a maximum entropy strategy. *Proc Natl Acad Sci U S A* **119**, e2122355119 (2022).
23. Haddock, S. H. D., Moline, M. A. & Case, J. F. Bioluminescence in the sea. *Ann Rev Mar Sci* **2**, 443–493 (2010).
24. Schenkmayerova, A. *et al.* Catalytic mechanism for Renilla-type luciferases. *Nat Catal* **6**, 23–38 (2023).
25. Paley, M. A. & Prescher, J. A. Bioluminescence: A versatile technique for imaging cellular and molecular features. *Medchemcomm* **5**, 255–267 (2014).
26. Xie, W. J. & Warshel, A. Natural evolution provides strong hints about laboratory evolution of designer enzymes. *Proc Natl Acad Sci U S A* **119**, e2207904119 (2022).
27. McGee, F. *et al.* The generative capacity of probabilistic protein sequence models. *Nat Commun* **12**, (2021).
28. Miller, S. R. An appraisal of the enzyme stability-activity trade-off. *Evolution* **71**, 1876–1887 (2017).
29. Beadle, B. M. & Shoichet, B. K. Structural bases of stability-function tradeoffs in enzymes. *J Mol Biol* **321**, 285–296 (2002).
30. Siddiqui, K. S. Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability. *Crit Rev Biotechnol* **37**, 309–322 (2017).

31. Loening, A. M., Fenn, T. D., Wu, A. M. & Gambhir, S. S. Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output. *Protein Eng Des Sel* **19**, 391–400 (2006).
32. Loening, A. M., Wu, A. M. & Gambhir, S. S. Red-shifted Renilla reniformis luciferase variants for imaging in living subjects. *Nat Methods* **4**, 641–643 (2007).
33. Tomatis, P. E., Rasia, R. M., Segovia, L. & Vila, A. J. Mimicking natural evolution in metallo-β-lactamases through second-shell ligand mutations. *Proc Natl Acad Sci U S A* **102**, 13761–13766 (2005).
34. Trzesniak, D., Kunz, A. P. E. & Van Gunsteren, W. F. *Perception and Performance Under Water*. *ChemPhysChem* (John Wiley & Sons, 1974).
35. Widder, E. A. Bioluminescence in the ocean: Origins of biological, chemical, and ecological diversity. *Science* **328**, 704–708 (2010).
36. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *2nd International Conference on Learning Representations* http://arxiv.org/abs/1312.6114 (2014).
37. Riesselman, A. J., Ingraham, J. B. & Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat Methods* **15**, 816–822 (2018).
38. Liu, J. & Escher, A. Improved assay sensitivity of an engineered secreted Renilla luciferase. *Gene* **237**, 153–159 (1999).
39. Meier, J. *et al.* Language models enable zero-shot prediction of the effects of mutations on protein function. *Adv Neural Inf Process Syst* **34**, 29287–29303 (2021).
40. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng Des Sel* **29**, 245–251 (2016).
41. Sternke, M., Tripp, K. W. & Barrick, D. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proc Natl Acad Sci U S A* **166**, 11275–11284 (2019).