EFFICIENT, FEW-SHOT DIRECTED EVOLUTION WITH ENERGY RANK ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Directed evolution is a powerful and widely used technique for protein engineering, and reducing the cost of iterated experimental observations has become a major priority for practitioners. A number of recent efforts to use machine-learning-based predictors to improve sequence selection have led to remarkable improvements in efficiency, but the sparse data at each experimental iteration restricts these approaches to extremely simple models. Adapting large-scale pre-trained protein language models using experimental data offers an alternative that we show productively leverages the strong inductive biases of the natural distribution of protein sequences to navigate high-dimensional, combinatorially large fitness landscapes. Our approach uses a general-purpose "post-training" algorithm grounded in statistical physics that employs quantitative experimental rankings to directly produce a sampler for diverse, high fitness sequences with fewer data points than competing methods.

1 INTRODUCTION

The advent of highly accurate structure prediction and sequence design tools has refocused protein engineering on structural features that strongly correlate with desired function Abramson et al. (2024); Watson et al. (2023); Hsu et al. (2022); Dauparas et al. (2022), but these methods do not directly leverage relevant experimental data when generating predictions. Experimental design via directed evolution Wang et al. (2021) (DE) stands somewhat in contrast to this trend; in DE, an experimental observable guides the sequential perturbation of a parent sequence. There is a productive and relatively unexplored interface between these two approaches, and a variety of recent methods have sought to augment DE with data-driven machine learning Wittmann et al. (2021a); Yang et al. (2024). Choosing candidate sequences in DE is complicated by a combinatorially large fitness landscape and epistatic effects that are challenging to anticipate *a priori* Starr & Thornton (2016); Johnson et al. (2023). Integrating DE workflows with the powerful inductive biases learned by large-scale pre-trained statistical models for sequences, such as protein language models (PLMs) Lin et al. (2021); Hayes et al. (2025), offers a compelling route to navigating rugged, high-dimensional fitness landscapes.

Adapting PLMs with small amounts of experimental data closely resembles a widely studied problem in the machine learning literature known as "post-training", in which a pre-trained model is optimized to yield a desired distributional shift. While many post-training methods exist Schulman et al. (2017); Ouyang et al. (2022), preference optimization Rafailov et al. (2023); Widatalla et al. (2024) is uniquely well-suited to the comparative task of ranking sequences using quantitative experimental outputs. Preference optimization seeks to adapt the model probabilities to impose a relative ranking among outputs that reflects an observed ranking; we recently developed a method called energy rank alignment (ERA) that ensures that relative probabilities among quantitatively ordered samples are preserved Chennakesavalu et al. (2025a;b). ERA is an explicit gradient-based algorithm which can be implemented efficiently to post-train even very large PLMs.

Here we show that predicted sequences using post-trained PLMs efficiently optimize fitness across a variety of tasks: maximizing enzymatic activity, antibiotic resistance, and protein-protein binding. ERA-adapted PLMs sample sequences that identify the highest fitness sequences faster and more robustly than alternative techniques that combine simple models and active learning, including EVOLVEpro, Active Learning-assisted Directed Evolution, Machine Learning Directed Evolu-

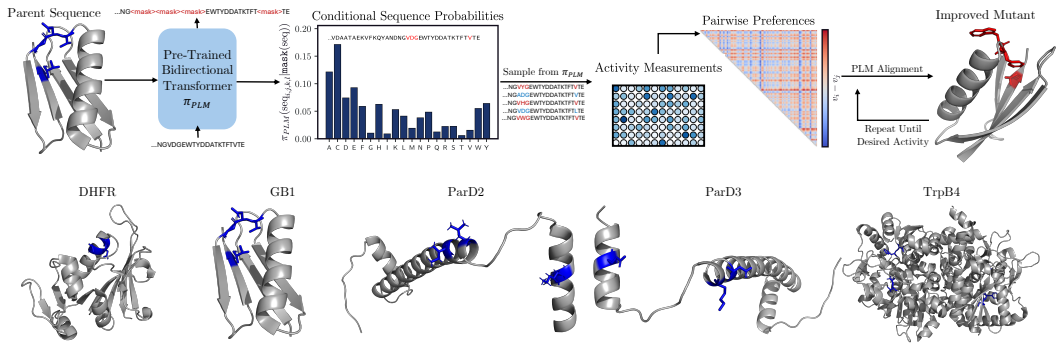

Figure 1: **Preference alignment of a protein language model for efficient protein optimization in a directed evolution campaign.** Using the conditional sequence distribution over residues, an initial set of protein sequences is sampled and evaluated. These sequences are compiled into a dataset of pairwise preferences labeled by their experimental fitness. This dataset is used to fine-tune the model with the energy rank alignment objective, targeting high-fitness sequences with tuneable entropy and controllable regularization. This process can be repeated iteratively in an active learning framework until some desired level of sequence fitness is achieved.

tion, and Direct Preference Optimization Jiang et al. (2025); Yang et al. (2025); Wu et al. (2019); Wittmann et al. (2021b); Rafailov et al. (2023). Interestingly, augmenting pre-trained PLMs with 3D structural information or thermodynamic stability data does not outperform a predictor that purely relies on protein sequence. We exhaustively benchmark our approach using previously collected combinatorially complete experimental data sets with notable epistatic effects. We consistently see that ERA shifts the sequence distribution towards high activity but importantly maintains diversity even after post-training optimization. While the relative computational effort required to optimize large-scale machine learning models like state-of-the-art PLMs is high compared to methods that use simple architectures, the conceptual and practical advantages make a strong case for expending the additional effort, given the even greater experimental costs.

2 Approach

Sampling mutant sequences requires conditionally evaluating amino acid likelihoods at residue positions of interest. Bidirectional transformers, also called masked language models Devlin et al. (2018), are well suited to this task because we can mask positions of interest and resample them. Throughout, we employ ESM3-1.4B Hayes et al. (2025), a state-of-the-art PLM that uses a bidirectional transformer architecture. Furthermore, the multi-modal nature of the model allows for additional conditioning beyond amino acid sequences, such as backbone coordinates, secondary structure elements, and even functional annotations. Given that the positions of interest for this study were largely chosen according to some structural intuition, we hypothesized that incorporating the coordinates of known crystal structures into the sequence design process would enhance predictive power.

To sample sequences from ESM3-1.4B, we mask the residue indices of interest within the protein target. These positions were then unmasked *simultaneously*, rather than permuting all possible unmasking orders between multiple sites. This enables us to treat the unmasking of token $x_i, i = 1, \ldots, n$ at each of the n positions as statistically independent events and compute generation pseudo-likelihoods from the ESM3-1.4B policy π_{ESM3} using this unmasking scheme as:

$$\pi_{\text{ESM3}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \pi_{\text{ESM3}}(x_i). \tag{1}$$

Exact likelihood calculations for arbitrary unmasking order in masked language models remains a topic of significant research interest Salazar et al. (2020); Torroba Hennigen & Kim (2023).

Adaptation with Energy Rank Alignment To fine-tune ESM3-1.4B in a manner that leverages experimental fitness readouts of generated sequences, we choose to apply preference optimization due to its contrastive signal. With a batch of N sequences sampled from π_{ESM3}, where $N \ll 1000$ (never exceeds 384 in this work), we construct the full dataset of $\frac{N(N-1)}{2}$ preference pairs labeled with their experimental readouts, as illustrated in Fig. 1. Our goal is to minimize the divergence between the parametric preference distribution p_θ:

$$p_\theta(y \succ y' \mid x) = \frac{\pi_\theta(y \mid x)}{\pi_\theta(y \mid x) + \pi_\theta(y' \mid x)} = \sigma\left(\log \frac{\pi_\theta(y \mid x)}{\pi_\theta(y' \mid x)}\right), \tag{2}$$

and the entropy regularized preference distribution p_γ:

$$p_\gamma = \sigma\left(\frac{\beta}{1+\gamma}\left[(U(x,y') - U(x,y)) + \beta^{-1}\gamma \log \frac{\pi_{\text{ref}}(y \mid x)}{\pi_{\text{ref}}(y' \mid x)}\right]\right). \tag{3}$$

Here β tunes the entropy of the Gibbs-Boltzmann target distribution, γ is a regularization hyperparameter, and $U(x, y)$ is an energy function (negative reward) to be minimized. The notation $y \succ y'$ indicates that y is preferred to y'. We optimize the following KL-divergence, the minimizer of which coincides with equation 3,

$$\mathcal{L}_{\text{ERA}}(\pi_\theta) = \mathbb{E}_{x\sim\mathcal{D},\, y,y'\sim\pi_{\text{ref}}(\cdot|x)}[D_{\text{KL}}(p_\gamma(y \succ y') \,\|\, p_\theta)]\,. \tag{4}$$

3 RESULTS

Method	DHFR	GB1	ParD2	ParD3	TrpB4
MLDE	0.96	0.67	0.99	0.9	0.78
ALDE	0.96	0.80	0.99	0.99	0.89
EVOLVEpro (ESM2)	0.95	0.81	**1.0**	0.99	0.88
EVOLVEpro (ESM3)	**1.0**	0.77	**1.0**	0.99	0.79
DPO	**1.0**	0.78	**1.0**	0.99	0.81
ERA	**1.0 ± 0.00**	**0.83 ± 0.04**	**1.0 ± 0.00**	**1.0 ± 0.00**	**0.91 ± 0.02**

Table 1: Average maximum fitness achieved across multiple replicate runs with standard errors included.

Method	DHFR	GB1	ParD2	ParD3	TrpB4
MLDE	0.60	0.06	0.98	0.12	0.02
ALDE	0.62	0.18	0.96	0.80	**0.46**
EVOLVEpro (ESM2)	0.00	0.10	0.80	0.70	0.40
EVOLVEpro (ESM3)	0.90	0.00	0.60	0.10	0.00
DPO	**1.0**	0.10	0.90	0.00	0.00
ERA	**1.0**	**0.20**	**1.0**	**0.90**	0.30

Table 2: Fraction of replicate runs reaching the maximum fitness sequence in the landscape.

Comparison to Related Methods In order to benchmark the suitability of ERA for DE campaigns, we used previously collected experimental DE datasets that were nearly combinatorially complete. Li et al. collected a set of diverse protein activity landscapes across five different protein targets which were well-suited for this task Li et al. (2025). These data feature a range of protein sizes, families, and activities, ranging from antibody binding to enzymatic efficiency. Three of the datasets—dihydrofolate reductase (DHFR) trimethoprim binding Papkou et al. (2023) and the ParD2/3-ParE2/3 bacterial antitoxin-toxin binding landscapes Lite et al. (2020) —involve mutating three positions of interest combinatorially for a total of 8,000 possible sequences. Two other datasets—the immunoglobulin binding of the B1 domain of protein G (GB1) Wu et al. (2016) and the tryptophan production rate of the β-subunit of tryptophan synthase (TrpB4) Johnston et al. (2024)—involve mutating four positions, making for 160,000 possible sequences.

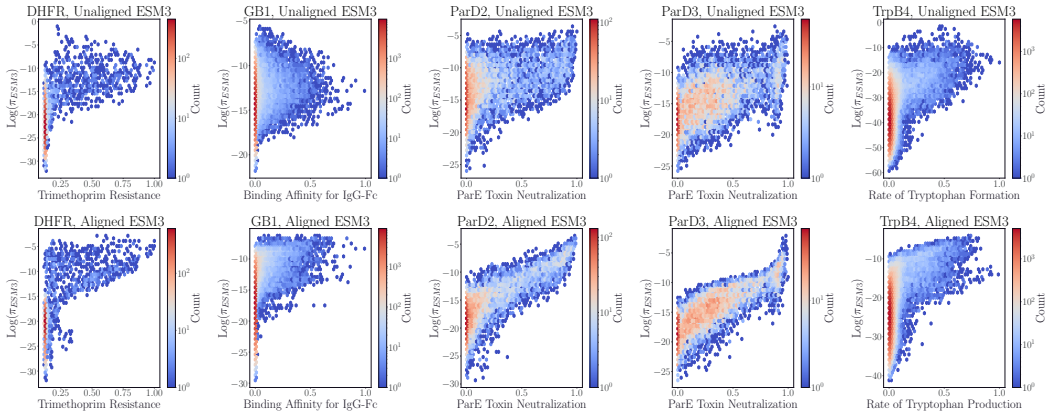

Figure 2: Evolution of the relationship between $\log \pi_{\text{ESM3}}$ and experimental activity from before (top row) to after (bottom row) alignment. The aligned models used to compute $\log \pi_{\text{ESM3}}$ in the bottom row are the final checkpoints in a random iterative alignment experiment.

Across all landscapes, four rounds of training with 96 sequences each were conducted followed by a final 96 sequences, for a total of 480 samples. The only exception to this was MLDE, for which only a single round of training on 384 sequences was conducted followed by sampling of 96 sequences as in Li et al. (2025). The sampling process across all methods and replicates allowed only for unique sequences that had not been previously trained on, except for DPO. Significant mode collapse in the DPO-trained ESM3-1.4B model required us to allow for repeated samples in order to make the benchmark tractable. Each of these experiments were repeated for 10 replicates, except for ALDE and MLDE, which were repeated for 50 replicates in Li et al. (2025).

Table 1 and Table 2 demonstrate that ERA attains state-of-the-art performance across a variety of diverse combinatorial fitness landscapes. The superior performance across the board compared to MLDE demonstrates the advantages of an iterative, active learning-type approach where the reference policy is updated more frequently. The inclusion of continuous labels in the training process leads to clearly superior performance compared to DPO. The mode collapse experienced by ESM3-1.4B when trained with DPO was so severe that 9 of the 10 replicate experiments converged to the exact same local optimum for TrpB4: V183; F184I; V227M; S228G. The comparison to ALDE and the ESM2-150M version of EVOLVEpro does show ERA falling short in terms of full optimization of the TrpB4 sequence. Nevertheless, ERA marginally outperforms these methods in terms of the average maximum fitness achieved. This suggests that even in cases where ERA is not able to fully optimize the sequence in the allowed 480-sample budget, many of the other high-activity sequences are still recovered.

Effects of Alignment Figure 2 demonstrates that fine-tuning ESM3-1.4B with ERA clearly suppresses the likelihoods of many inactive sequences and promotes the likelihoods of highly active sequences by several orders of magnitude. These correlations between $\log \pi_{\text{ESM3}}$ and the experimental activities are particularly impressive for the ParD2 and ParD3 landscapes, and the improvement for DHFR and GB1 is notable as well, despite the continued presence of some false-positives, i.e. sequences with high likelihoods and low activities. Though the overall correlation for TrpB4 improves, the bimodal behavior of the $\log \pi_{\text{ESM3}}$ distribution in the aligned plot seems to suggest the presence of the local optima in the fitness landscape being located by training with ERA. Despite this, the inherent stochasticity of sampling directly from π_{ESM3} along with the sample diversity promoted by ERA still allows for the possibility of escaping these local optima when sampling sequences.

The probability mass assigned to high-activity sequences under the post-ERA policy π_{ESM3} increases markedly. Figure 3 displays the fold-improvement (on a log-scale) of the probability mass assigned to sequences above a specified lower bound in activity after fine-tuning with ERA. More probability mass is assigned to high-activity sequences for every landscape, with the gains for DHFR, GB1, and TrpB4 appearing to be particularly substantial. Even for ParD2 and ParD3, whose fold-increases in probability mass look less impressive for their most active sequences, the relatively high probability mass assigned to those regions initially (0.25 above fitness of 0.9 for ParD2, 0.40 above fitness of 0.9

for ParD3) means these are still considerable absolute gains in probability mass for the high-activity region of sequence space.

Additional Conditioning Beyond Primary Sequence The multi-modal nature of ESM3 enabled us to investigate whether additional conditioning beyond a protein's primary sequence would help or harm the performance of ERA in optimizing for protein activity. Thermostability is a well-understood prerequisite for any arbitrary protein function, so we aimed to test whether learning more general information about thermostability prior to fine-tuning on a specific experimental dataset would provide ESM3-1.4B with a helpful prior for mutational effects. Connections between sequence likelihoods and thermostability were originally described in Lapedes et al. (2012); Notin et al. (2023) and have been recently employed to predict changes in affinity conditioned on structure predictions Deng et al. (2025). To this end, we pre-trained ESM3-1.4B with the full Megascale Tsuboyama et al. (2023) mutational thermostability dataset using the ERA objective. The dataset contains single and double mutations to over 400 parent sequences, so the number of total comparisons that could be formed was impractical for full fine-tuning. To alleviate this issue, we trained for a single epoch only, and only utilized pairs consisting of one stabilizing and one destabilizing mutation for each parent sequence.

Furthermore, the intrinsic link between 3D protein structure and function suggests that conditioning on backbone coordinates within the ESM3-1.4B model would provide useful information for sequence optimization. The choice of backbone coordinates we made were the ground-truth crystal structures from the PDB Berman et al. (2000) for each of the parent sequences in the landscapes of interest, because all the activities being optimized are closely tied to the natural function of each protein. Lastly, though ESM3-1.4B was not trained explicitly with bound complex structures, we also attempted to condition on the backbone coordinates of the bound complex for the three binding affinity landscapes (GB1, ParD2, and ParD3).

Method	DHFR	GB1	ParD2	ParD3	TrpB4
ERA	**1.0 ± 0.00**	**0.83 ± 0.04**	**1.0 ± 0.00**	**1.0 ± 0.00**	**0.91 ± 0.02**
ERA (w/ struct. cond.)	-	0.76 ± 0.04	-	-	0.89 ± 0.02
ERA (w/ complex)	-	0.74 ± 0.04	**1.0 ± 0.00**	0.99 ± 0.00	-
ERA (w/ stab. pretrain.)	**1.0 ± 0.00**	0.78 ± 0.04	**1.0 ± 0.00**	0.99 ± 0.00	0.85 ± 0.02

Table 3: Average maximum fitness achieved across multiple replicate runs with standard errors included for different variations of ERA on ESM3-1.4B.

Method	DHFR	GB1	ParD2	ParD3	TrpB4
ERA	**1.0**	**0.20**	**1.0**	**0.90**	**0.30**
ERA (w/ struct. cond.)	-	0.10	-	-	0.10
ERA (w/ complex)	-	0.10	**1.0**	0.70	-
ERA (w/ stab. pretrain.)	-	**0.20**	**1.0**	0.50	0.10

Table 4: Fraction of replicate runs reaching the maximum fitness sequence in the landscape for different variations of ERA on ESM3-1.4B.

Interestingly, neither structure conditioning nor thermostability pre-training substantively improves performance over the pure sequence model, despite signifciant changes to the reference π_{ESM3} policy as shown by Figures 4 - 7. Tables 3 and 4 suggest that none of these strategies led to an improved ability for ESM3-1.4B to navigate these multi-site, combinatorial landscapes. The shortcomings of the Megascale pretraining strategy can be rationalized by considering that the single- and double-mutant nature of the dataset provides no opportunity to consider higher-order epistatic effects, which are crucial toward effectively navigating rugged landscapes. Regarding the backbone conditioning strategies, those structural constraints are likely too strict for full sequence optimization and make local optima in the fitness landscape harder to escape. This is consistent with previous studies of structure-informed protein language models benefitting less from this conditioning on near-native optimization tasks Ruffolo et al. (2024), as in the five landscapes we explore here.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 2024.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

Shriram Chennakesavalu, Frank Hu, Sebastian Ibarraran, and Grant M. Rotskoff. Aligning chemical and protein language models with continuous feedback using energy rank alignment. In *Workshop on Generative and Experimental Perspectives for Biomolecular Design at ICLR 2025*, volume 2, 2025a.

Shriram Chennakesavalu, Frank Hu, Sebastian Ibarraran, and Grant M. Rotskoff. Aligning chemical and protein language models with continuous feedback via energy rank alignment. In *39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, volume 39, 2025b.

J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, October 2022. ISSN 1095-9203. doi: 10.1126/science.add2187. URL http://dx.doi.org/10.1126/science.add2187.

Arthur Deng, Karsten D. Householder, Fang Wu, K. Christopher Garcia, and Brian L. Trippe. Predicting mutational effects on protein binding from folding energy. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 13129–13151. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/deng25d.html.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018. URL https://arxiv.org/abs/1810.04805.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q. Tran, Jonathan Deaton, Marius Wiggert, Rohil Badkundri, Irhum Shafkat, Jun Gong, Alexander Derry, Raul S. Molina, Neil Thomas, Yousuf Khan, Chetan Mishra, Carolyn Kim, Liam J. Bartie, Matthew Nemeth, Patrick D. Hsu, Tom Sercu, Salvatore Candido, and Alexander Rives. Simulating 500 million years of evolution with a language model. *Science*, 387 (6736), 2025.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. April 2022. doi: 10.1101/2022.04.10.487779. URL http://dx.doi.org/10.1101/2022.04.10.487779.

Kaiyi Jiang, Zhaoqing Yan, Matteo Di Bernardo, Samantha R. Sgrizzi, Lukas Villiger, Alisan Kayabolen, B. J. Kim, Josephine K. Carscadden, Masahiro Hiraizumi, Hiroshi Nishimasu, Jonathan S. Gootenberg, and Omar O. Abudayyeh. Rapid in silico directed evolution by a protein language model with evolvepro. *Science*, 387(377), 2025. doi: 10.1126/science.adk8946. URL https://www.science.org/doi/full/10.1126/science.adr6006.

Milo S. Johnson, Gautam Reddy, and Michael M. Desai. Epistasis and evolution: recent advances and an outlook for prediction. *BMC Biology*, 21(1):120, 2023. doi: 10.1186/s12915-023-01584-4.

Kadina E. Johnston, Patrick J. Almhjell, Ella J. Watkins-Dulaney, Grace Liua, Nicholas J. Porterb, Jason Yang, and Frances H. Arnold. A combinatorially complete epistatic fitness landscape in an enzyme active site. *Proceedings of the National Academy of Sciences*, 121(32), 2024.

Alan Lapedes, Bertrand Giraud, and Christopher Jarzynski. Using sequence alignments to predict protein structure and stability with high accuracy, 2012. URL https://arxiv.org/abs/1207.2484.

Francesca-Zhoufan Li, Jason Yang, Kadina E. Johnston, Emre Gürsoy, Yisong Yue, and Frances H. Arnold. Evaluation of machine learning-assisted directed evolution across diverse combinatorial landscapes. *Cell Systems*, 16(9):101387, 2025. ISSN 2405-4712. doi: https://doi.org/10.1016/j.cels.2025.101387. URL https://www.sciencedirect.com/science/article/pii/S2405471225002200.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil1, Ori Kabeli1, Yaniv Shmueli1, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), 2021.

Thuy-Lan V. Lite, Robert A Grant, Isabel Nocedal, Megan L. Littlehale, Monica S. Guo, and Michael T Laub. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife*, 9(e60924), 2020.

Pascal Notin, Aaron W. Kollasch, Daniel Ritter, Lood van Niekerk, Steffanie Paul, Hansen Spinner, Nathan Rollins, Ada Shaw, Ruben Weitzman, Jonathan Frazer, Mafalda Dias, Dinko Franceschi, Rose Orenbuch, Yarin Gal, and Debora S. Marks. Proteingym: Large-scale benchmarks for protein design and fitness prediction. December 2023. doi: 10.1101/2023.12.07.570727. URL http://dx.doi.org/10.1101/2023.12.07.570727.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 27730–27744. Curran Associates, Inc., 2022.

Andrei Papkou, Lucia Garcia-Pastor, José Antonio Escudero, and Andreas Wagner. A rugged yet easily navigable fitness landscape. *Science*, 382(6673), 2023.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741. Curran Associates, Inc., 2023.

Jeffrey A. Ruffolo, Aadyot Bhatnagar, Joel Beazer, Stephen Nayfach, Jordan Russ, Emily Hill, Riffat Hussain, Joseph Gallagher, and Ali Madani. Adapting protein language models for structure-conditioned design. August 2024. doi: 10.1101/2024.08.03.606485. URL http://dx.doi.org/10.1101/2024.08.03.606485.

Julian Salazar, Davis Liang, Toan Q. Nguyen, and Katrin Kirchhoff. Masked language model scoring. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2699–2712. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.acl-main.240. URL http://dx.doi.org/10.18653/v1/2020.acl-main.240.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms, August 2017.

Tyler N Starr and Joseph W Thornton. Epistasis in protein evolution. *Protein Science*, 25(7):1204–1218, 2016. doi: 10.1002/pro.2897.

Lucas Torroba Hennigen and Yoon Kim. Deriving language models from masked language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 1149–1159. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.acl-short.99. URL http://dx.doi.org/10.18653/v1/2023.acl-short.99.

Kotaro Tsuboyama, Justas Dauparas, Jonathan Chen, Elodie Laine, Yasser Mohseni Behbahani, Jonathan J. Weinstein, Niall M. Mangan, Sergey Ovchinnikov, and Gabriel J. Rocklin. Mega-scale experimental analysis of protein folding stability in biology and design. *Nature*, 620(7973): 434–444, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06328-6. URL http://dx.doi.org/10.1038/s41586-023-06328-6.

Yajie Wang, Pu Xue, Mingfeng Cao, Tianhao Yu, Stephan T. Lane, and Huimin Zhao. Adaptation in protein fitness landscapes is facilitated by indirect paths. *Chemical Reviews*, 121(20), 2021.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyoung Baek, and David Baker. De novo design of protein structure and function with rfdiffusion. *Nature*, 620:1089–1100, 2023.

Talal Widatalla, Rafael Rafailov, and Brian Hie. Aligning protein generative models with experimental fitness via direct preference optimization, May 2024.

Bruce J Wittmann, Kadina E Johnston, Zachary Wu, and Frances H Arnold. Advances in machine learning for directed evolution. *Current Opinion in Structural Biology*, 69:11–18, 2021a. doi: 10.1016/j.sbi.2021.01.008.

Bruce J. Wittmann, Yisong Yue, and Frances H. Arnold. Informed training set design enables efficient machine learning-assisted directed protein evolution. *Cell Systems*, 12:1026–1045, 2021b.

Nicholas C. Wu, Lei Dai, C. Anders Olson, James O. Lloyd-Smith, and Ren Sun. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife*, 5(e16965), 2016.

Zachary Wu, S. B. Jennifer Kan, Russell D. Lewis, Bruce J. Wittmann, and Frances H. Arnold. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences*, 116(18):8852–8858, April 2019. ISSN 1091-6490. doi: 10.1073/pnas.1901979116. URL http://dx.doi.org/10.1073/pnas.1901979116.

Jason Yang, Francesca-Zhoufan Li, and Frances H. Arnold. Opportunities and challenges for machine learning-assisted enzyme engineering. *ACS Central Science*, 10(2):226–241, 2024. doi: 10.1021/acscentsci.3c01275. URL https://doi.org/10.1021/acscentsci.3c01275.

Jason Yang, Ravi G. Lal, James C. Bowden, Raul Astudillo, Mikhail A. Hameedi, Sukhvinder Kaur, Matthew Hill, Yisong Yue, and Frances H. Arnold. Active learning-assisted directed evolution. *Nature Communications*, 16(1), January 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-55987-8. URL http://dx.doi.org/10.1038/s41467-025-55987-8.

A Appendix

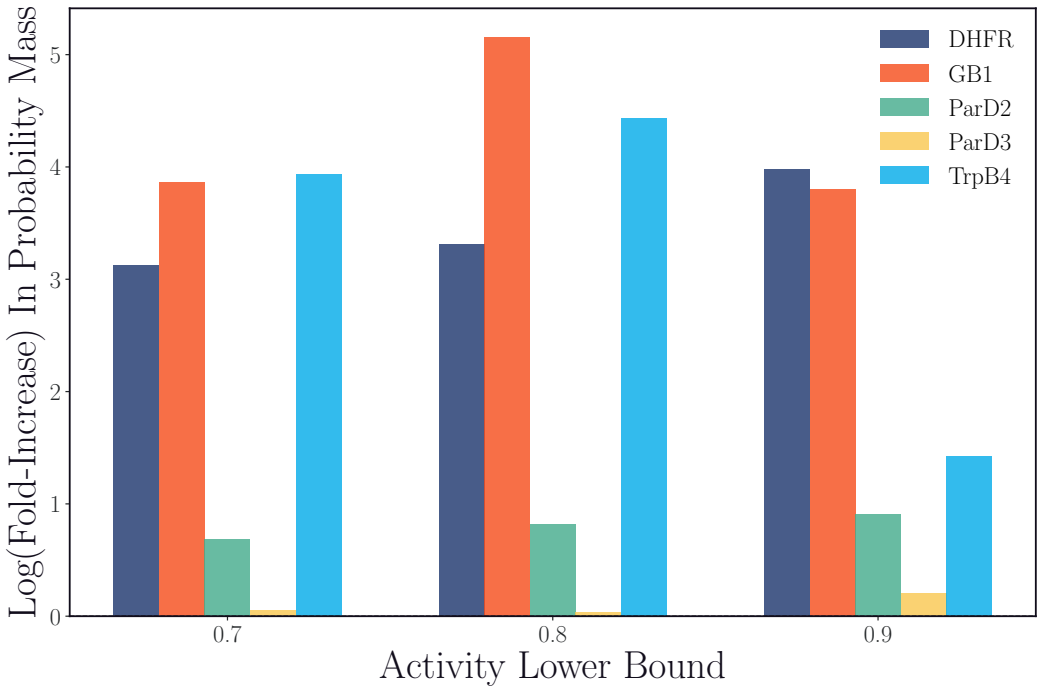

Figure 3: Fold-increase in the probability mass assigned to sequences above a certain experimental activity under π_{ESM3} before and after alignment across five multi-site protein fitness landscapes. The post-alignment π_{ESM3} is obtained from the final checkpoint of a random iterative alignment experiment.

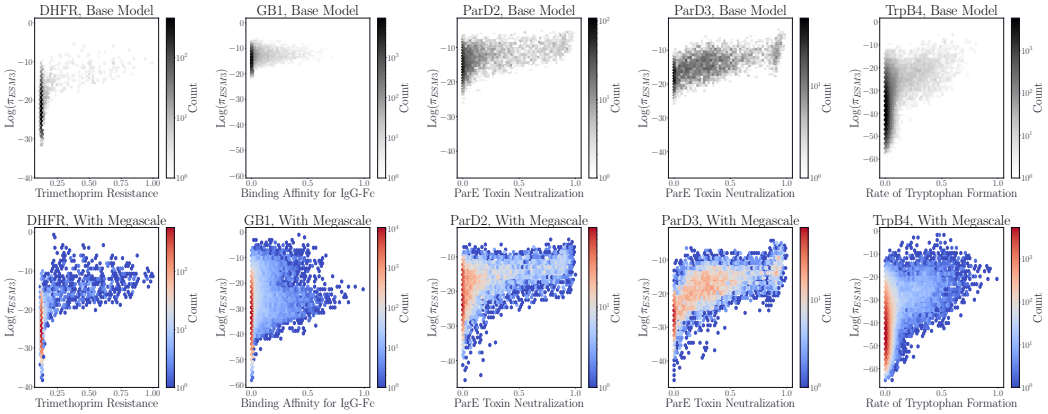

Figure 4: Evolution of the relationship between $\log \pi_{\text{ESM3}}$ and experimental activity from before (top row) to after (bottom row) ERA pre-training with the Megascale thermostability dataset Tsuboyama et al. (2023).

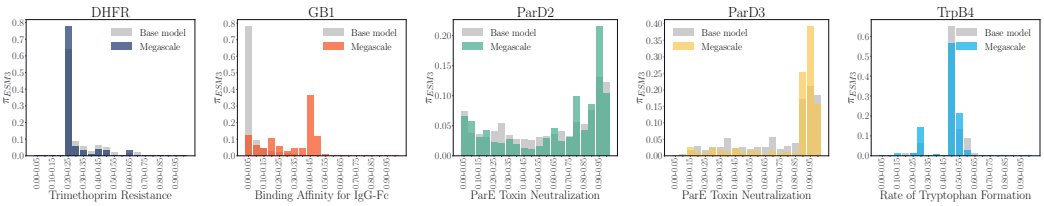

Figure 5: Evolution of probability mass from π_{ESM3} binned by sequence fitness from before (gray) to after (color) ERA pre-training on the Megascale thermostabilty dataset Tsuboyama et al. (2023).

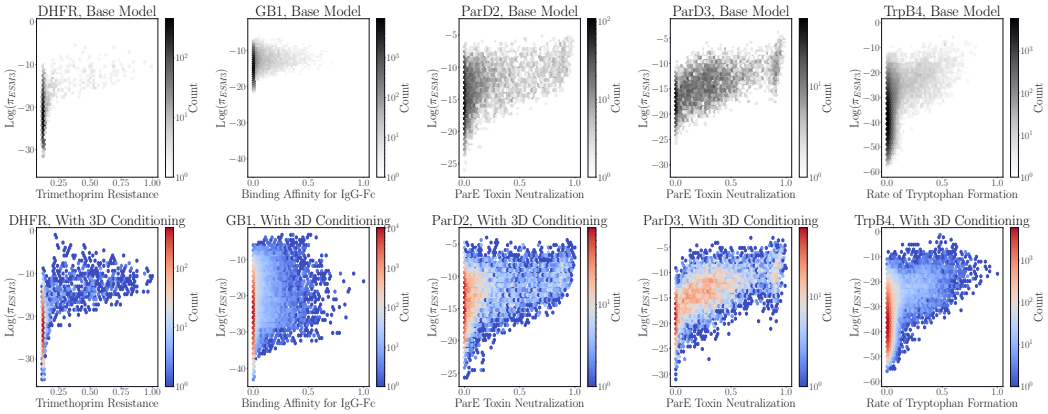

Figure 6: Evolution of the relationship between $\log \pi_{\text{ESM3}}$ and experimental activity from before (top row) to after (bottom row) conditioning ESM3-1.4B with the backbone 3D coordinates of a protein target's wild type crystal structure from the PDB.

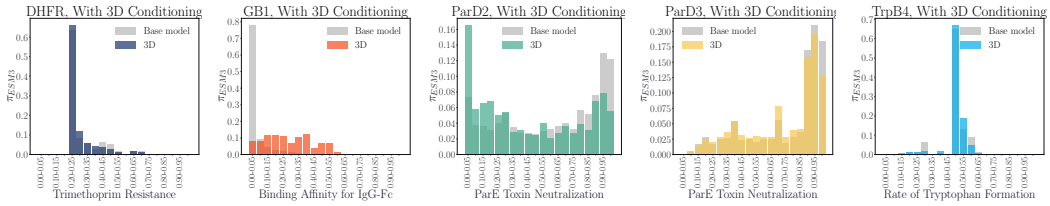

Figure 7: Evolution of probability mass landscapes with conditioning on the backbone coordinates of the wild type protein structure.