Aligning Protein Language Models to Stability Preferences using 1M+ Experimental Mutant Effects

Anonymous Author(s)
Affiliation
Address
email

Abstract

Protein language models (pLMs) demonstrate clear scaling laws for structure prediction but exhibit deteriorating performance on mutation effect prediction as model size increases. We present StableESM, applying Direct Preference Optimization (DPO) to ESM2 using over 1 million experimental stability measurements across 1,000 protein domains. Similar to the GPT-3 to ChatGPT transition that aligned language models to be more helpful and aligned with human preferences, we show that preference alignment enables larger model sizes to continue improving on mutation effect prediction while maintaining structure prediction capabilities. StableESM demonstrates improved zero-shot generalization to unseen protein domains and families, and to higher-order mutational effects. In a computational protein design campaign to engineer more stable variants of multicopper oxidase which was unseen during preference alignment, StableESM identified promising designs that were both novel compared to the natural protein and different from what the original model would suggest. Moreover, experimental testing validated StableESM predictions, with designed mutants showing performance equal to or exceeding wild type across normal, thermophilic, and extremophilic temperature conditions. This study shows that preference optimization for protein language models not only improves the base model in mutational effects, but also improves on unseen protein domains and families, and even changes the fitness/stability landscape for a completely unseen/distant protein from the preference alignment dataset.

1 Introduction

Large-scale pretraining has enabled remarkable scaling laws in both language models and protein language models (pLMs), with performance improving predictably with model size for tasks like structure prediction [1, 2]. However, pLMs exhibit a fundamental limitation: while larger models excel at structure-related tasks, their zero-shot mutation effect prediction performance tends to saturate or even deteriorate with scale [3, 4]. This scaling failure suggests that raw parameter scaling is insufficient for capturing sequence-function relationships.

Recent work indicates this limitation stems from larger models overfitting to phylogenetic noise rather than functional constraints [5]. The transformation from GPT-3 to ChatGPT demonstrated that post-training alignment using reinforcement learning from human feedback (RLHF) can unlock model capabilities with minimal additional data [6] (Figure S1a). The explosive growth of experimental protein variant data—millions of measurements linking sequence changes to functional outcomes—presents an analogous opportunity for protein language model alignment.

We hypothesize that preference optimization using experimental stability data can enable protein language models to learn generalizable behaviors and features that transfer beyond the training dataset

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

to unseen protein domains, families, and related tasks (Figure S1b). Protein stability represents a relatively context-agnostic property suitable for large-scale alignment, unlike activity measurements that may be beneficial in one context while deleterious in another. This raises the key question of whether stability-based preference alignment teaches models fundamental principles of protein physics that generalize broadly, rather than simply memorizing specific sequence-stability relationships from the training data.

2 Methods

2.1 Large-Scale Preference Dataset Curation

We curated a comprehensive preference dataset by combining three major experimental databases: MegaScale [7], Human Domainome 1 [8], and FireProtDB [9]. Our curation yielded over 1 million mutations across 1,154 protein domains, generating 143.6M preference pairs for training and 1.8M for validation (Figure S1c). Each preference pair consists of two mutations from the same protein domain where one is experimentally more stable, providing direct ranking signals for optimization.

2.2 StableESM: Direct Preference Optimization of a Protein Language Model

We implemented StableESM using Low-Rank Adaptation (LoRA) applied to pretrained ESM2 models (8M to 3B parameters), enabling efficient fine-tuning while preserving evolutionary representations. We applied Direct Preference Optimization (DPO) [10] to train models to assign higher likelihood to experimentally stable mutations compared to unstable ones from identical protein contexts (Figure S1d).

Our approach leverages log-likelihood ratios computed by masking mutation positions and calculating $\pi(y|x) = \frac{p(x_i=\text{mut}|x_{-i})}{p(x_i=\text{wt}|x_{-i})}$, where x_{-i} represents the sequence with position i masked. Each preference pair consists of stable (y_w) and unstable (y_l) mutations from the same protein domain, but not necessarily the same sequence position.

The DPO objective fundamentally shifts training from masked language modeling to preference-based optimization:

$$\mathcal{L}_{\text{DPO}} = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\right)\right] \tag{1}$$

where π_{ref} is the reference ESM2 model and π_θ represents the aligned model's log-likelihood ratios computed via the masking procedure.

2.3 Experimental Validation of Designed Variants

To validate StableESM predictions, we experimentally tested designed mutations in laccase (CotA) from *Bacillus subtilis*. We generated target mutants and assessed their thermostability through enzyme activity assays at multiple temperatures (25°C, 65°C, and 85°C). All experiments were performed in quadruplicate. Detailed experimental protocols are provided in Supplementary Materials B.1.

3 Results

3.1 Training Dynamics and Preference Learning

DPO training successfully learned to distinguish stable from unstable mutations across all model sizes. Training and validation losses decreased smoothly while the probability of preferring stable over unstable mutations increased from no preference (0.5) to $\sim$0.70, demonstrating successful preference learning without overfitting (Figure S2a).

3.2 StableESM Demonstrates Improved Performance and Generalization Across Scales and Protein Families

StableESM provides consistent improvements across all model scales while preventing the performance deterioration typically observed in larger protein language models. While ESM2 performance saturates or declines from 650M to 3B parameters on out-of-distribution benchmarks [4], StableESM demonstrates continued improvement with scale, suggesting that preference alignment enables larger models to effectively utilize their increased capacity for mutation effect prediction. The model demonstrates robust generalization across multiple dimensions, including unseen homolog domains, double mutation effects despite training only on single mutations, and most notably to proteins substantially longer than training sequences. On FireProt domains ranging from 254-1022 amino acids (compared to maximum 254 during training), 94% of domains show improved correlations, with dramatic improvements such as Tyrosine-protein kinase Fyn correlation increasing from 0.543 (ESM2 3B) to 0.886 (StableESM 3B) (Supplementary Materials B.2, Figures S2, S3).

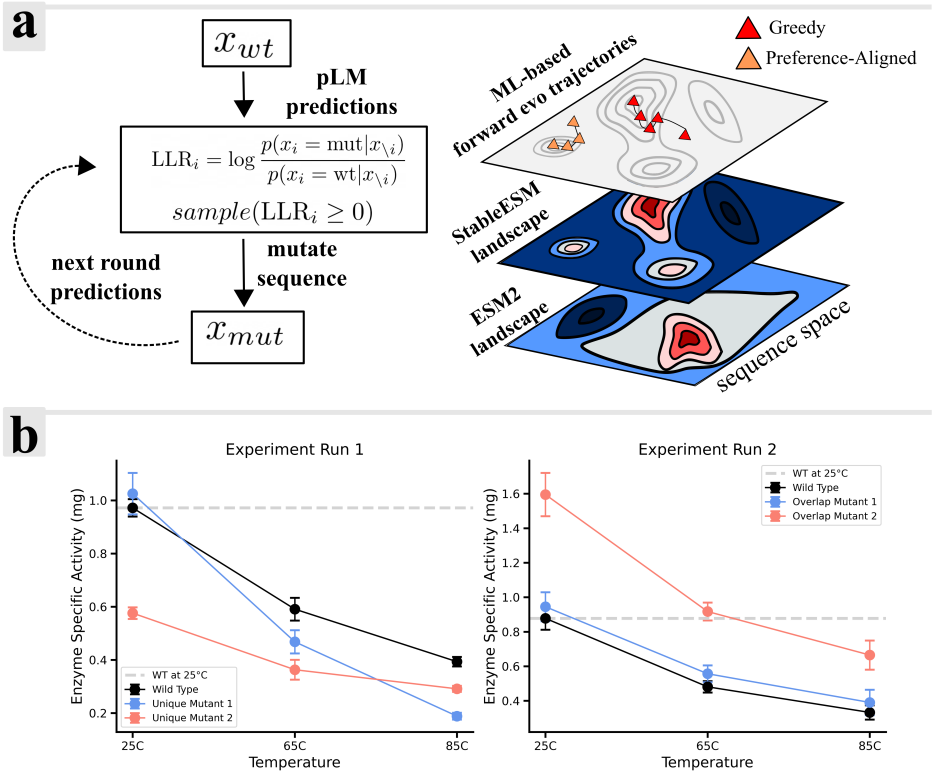

Figure 1: **Experimental validation of StableESM mutant predictions.** **(a)** Machine learning-directed evolution framework using StableESM predictions to guide exploration of fitness landscapes. **(b)** Enzyme activity measurements for multicopper oxidase variants at elevated temperatures. Overlap mutant 2 (T261F) in experiment run 2, computationally predicted as stabilizing by both models, demonstrates enhanced thermostability with maintained activity at 65°C and 2-fold improvement over wild-type at 85°C, validating the accuracy of stability predictions.

3.3 Experimental Validation in Design of Thermostable Multicopper Oxidases

To validate StableESM's predictive accuracy, we experimentally tested designed mutations in laccase (CotA) from Bacillus subtilis, a protein family not represented in the preference alignment dataset. We selected mutations based on StableESM predictions, including variants where both StableESM and ESM2 agreed on stabilizing effects (overlap mutants) and StableESM-specific predictions.

Experimental enzyme activity assays confirmed StableESM's predictions, with designed mutants demonstrating enhanced thermostability across temperature conditions (Figure 1). Notably, overlap mutant 2 (T261F), predicted as stabilizing by both models (LLR > 0), maintained superior activity relative to wild-type at 65°C and retained approximately 2-fold higher activity than wild-type even at the extreme condition of 85°C. These results validate that StableESM learns transferable stability principles that generalize to completely unseen protein families, enabling accurate prediction of functional improvements in distantly related proteins.

Additionally, we demonstrate StableESM's utility in large-scale protein design through machine learning-directed evolution on multicopper oxidase (PDB: 1GSK), achieving substantial predicted thermostability improvements with high-confidence predictions (Supplementary Materials B.3, Figure S4). The experimental validation of single mutations provides direct evidence that preference alignment enables protein language models to make accurate stability predictions across diverse protein families.

4 Discussion and Implications

StableESM establishes that preference optimization with over 1 million experimental measurements enables protein language models to learn generalizable stability principles that transfer robustly to unseen protein domains and families. The approach addresses multiple critical challenges simultaneously: consistent improvements across all model sizes, enhanced generalization to completely novel proteins, and fundamental alteration of fitness landscape exploration that benefits protein design applications.

A key finding is that preference alignment provides universal benefits regardless of model size, while enabling larger models to achieve superior performance on unseen domains. This suggests that larger protein language models possess latent capabilities for understanding protein stability that can be unlocked through targeted alignment, rather than being fundamentally limited by their architecture. The consistent improvements across the 8M to 3B parameter range demonstrate that preference optimization teaches transferable principles rather than memorizing dataset-specific patterns.

Most significantly, experimental validation with single mutants provides direct evidence that StableESM predictions translate to measurable functional improvements in proteins completely absent from the training data. Designed variants outperformed wild-type across all tested temperatures (25°C, 65°C, 85°C), with mutant activity at 65°C exceeding wild-type performance at 25°C and maintaining approximately 2-fold higher activity than wild-type even at the extreme condition of 85°C. This experimental confirmation demonstrates that preference alignment fundamentally alters how models explore sequence space, enabling identification of beneficial mutations that translate to real-world protein improvements. Additionally, our computational protein design validation reveals that the 86.2% similarity between StableESM and ESM2 greedy designs (71 amino acid differences) despite comparable edit numbers demonstrates that aligned models navigate different regions of the fitness landscape. This landscape alteration has profound implications for protein engineering: preference-aligned models can identify beneficial mutations and design strategies that would be entirely missed by traditional approaches and can effectively guide experimental thermostability engineering campaigns.

The framework established here opens new directions for developing more capable protein foundation models. Rather than accepting that protein language models cannot effectively utilize increased parameters for functional prediction, targeted post-training optimization can unlock these capabilities while requiring orders of magnitude less data than pretraining. The $\sim$1M preference mutants used here represent a tiny fraction compared to pretraining data, yet fundamentally alter model behavior and design capabilities.

Future work should explore extending this landscape-altering approach to other experimental properties beyond stability, investigating multi-objective alignment procedures, and developing preference learning objectives that capture complex epistatic interactions. As experimental datasets continue to grow, preference-based alignment may become essential for developing protein foundation models that not only understand evolutionary patterns but can effectively navigate fitness landscapes for practical protein design applications or viral escape prediction.

References

[1] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021.

[2] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[3] Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alex Rives. Language models enable zero-shot prediction of the effects of mutations on protein function. *Advances in neural information processing systems*, 34:29287–29303, 2021.

[4] Chao Hou, Di Liu, Aziz Zafar, and Yufeng Shen. Understanding protein language model scaling on mutation effect prediction. *bioRxiv*, pages 2025–04, 2025.

[5] Eli Weinstein, Alan Amin, Jonathan Frazer, and Debora Marks. Non-identifiability and the blessings of misspecification in models of molecular fitness. *Advances in neural information processing systems*, 35:5484–5497, 2022.

[6] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[7] 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, 2023.

[8] Antoni Beltran, Xiang'er Jiang, Yue Shen, and Ben Lehner. Site-saturation mutagenesis of 500 human protein domains. *Nature*, 637(8047):885–894, 2025.

[9] Jan Stourac, Juraj Dubrava, Milos Musil, Jana Horackova, Jiri Damborsky, Stanislav Mazurenko, and David Bednar. Fireprotdb: database of manually curated protein stability data. *Nucleic acids research*, 49(D1):D319–D324, 2021.

[10] 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. *Advances in neural information processing systems*, 36:53728–53741, 2023.

[11] Chiara Rodella, Symela Lazaridi, and Thomas Lemmin. Temberture: advancing protein thermostability prediction with deep learning and attention mechanisms. *Bioinformatics Advances*, 4(1):vbae103, 2024.

A Additional Algorithms, Figures, and Tables

Algorithm 1 Machine Learning-Based Directed Evolution with StableESM

Require: Wild-type sequence x_{wt}, StableESM model π_θ, base ESM2 model π_{ref}, number of edits N, sampling strategy S

Ensure: Designed protein sequence x_{final}

$x_{current} \leftarrow x_{wt}$
for $round = 1$ **to** N **do**
  for $i = 1$ **to** $|x_{current}|$ **do**
    Mask position i in $x_{current}$ to get x_{-i}
    for each amino acid $a \neq x_{current}[i]$ **do**
      $\text{LLR}_{i,a}^{\text{StableESM}} \leftarrow \log \frac{p_\theta(a|x_{-i})}{p_\theta(x_{current}[i]|x_{-i})}$
      $\text{LLR}_{i,a}^{\text{ESM2}} \leftarrow \log \frac{p_{ref}(a|x_{-i})}{p_{ref}(x_{current}[i]|x_{-i})}$
    end for
    $\text{LLR}_{i}^{\text{StableESM}} \leftarrow \max_a \text{LLR}_{i,a}^{\text{StableESM}}$
    $\text{LLR}_{i}^{\text{ESM2}} \leftarrow \max_a \text{LLR}_{i,a}^{\text{ESM2}}$
  end for
  if $S =$ greedy **then**
    $i^* \leftarrow \arg\max_i \text{LLR}_{i}^{\text{StableESM}}$ where $\text{LLR}_{i}^{\text{StableESM}} \geq 0$
  else if $S =$ sample **then**
    $i^* \leftarrow$ sample from $\{i : \text{LLR}_{i}^{\text{StableESM}} \geq 0\}$
  else if $S =$ unique **then**
    $i^* \leftarrow$ sample from $\{i : \text{LLR}_{i}^{\text{StableESM}} \geq 0 \text{ and } \text{LLR}_{i}^{\text{ESM2}} < 0\}$
  end if
  $a^* \leftarrow \arg\max_a \text{LLR}_{i^*,a}^{\text{StableESM}}$
  $x_{current}[i^*] \leftarrow a^*$
end for
return $x_{current}$

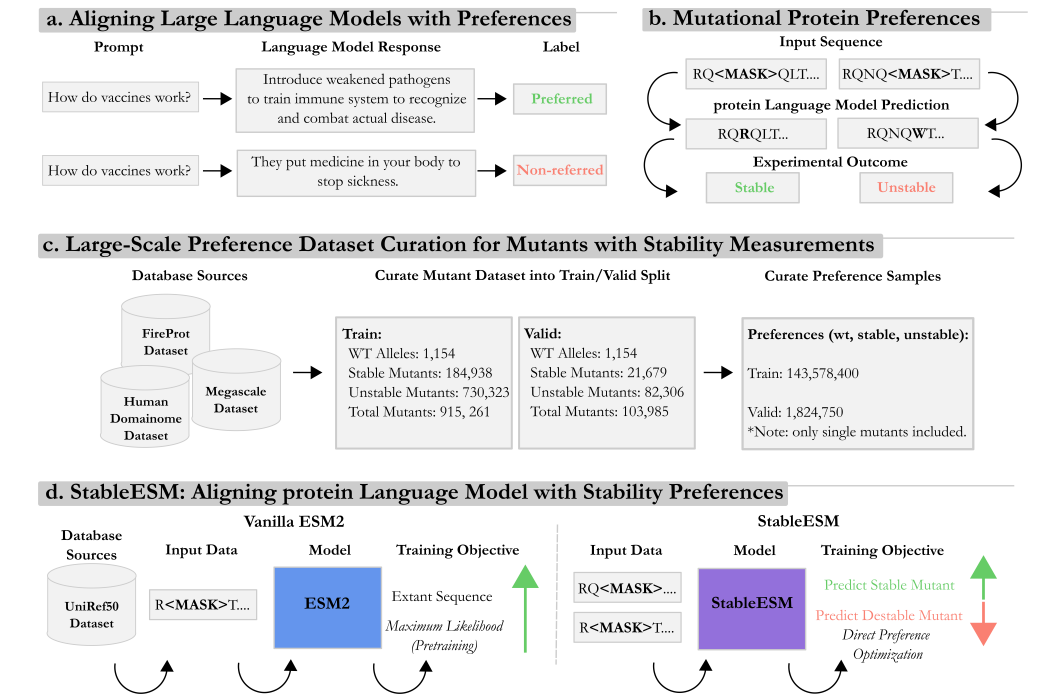

Figure S1: **StableESM: Aligning protein language models with experimental stability preferences using Direct Preference Optimization. (a)** Conceptual framework adapting preference-based alignment from large language models to protein language models. **(b)** Mutational protein preference learning paradigm where models learn to distinguish experimentally stable from unstable variants. **(c)** Large-scale preference dataset curation combining MegaScale, Human Domainome 1, and FireProtDB databases, yielding 143,578,400 preference pairs for training. **(d)** Training paradigm comparison between vanilla ESM2 and StableESM, showing the shift from masked language modeling to preference-based optimization.

B Supplementary Materials

B.1 Multicopper Oxidase Experimental Protocol

B.1.1 Cloning and Preparing DNA Constructs

Laccase (CotA) from *Bacillus subtilis* (UniProtKB accession P07788) was codon-optimized to *Escherichia coli*, synthesized as a gBlock (Twist Bioscience, South San Francisco, CA, USA), and cloned into pY71 vector using Gibson Assembly. Site-directed mutations (A296G, G270C, I368V, and T261F) were introduced using the Q5® Site-Directed Mutagenesis Kit (NEB, E0554S) with pY71-Laccase as the template. All constructs were verified by DNA sequencing.

B.1.2 Cell-Free Expression and Protein Quantification

Cell-free expression (CFE) reactions were performed using the PURExfrex2.0 system at 30°C for 20 hours, and the concentration of the expressed protein was determined by Bradford assay (ThermoFisher). Protein standards (0, 0.25, 0.5, 1, 1.5, 2 mg/mL) were prepared and used as the standard curve. The absorbance at 595 nm was measured after mixing 4 μL of protein standards with 200 μL of Bradford reagent and allowing it to react at room temperature for 20 min. The protein concentration in each CFE reaction was determined from the standard curve, using no DNA as a blank control. Each sample was measured in four replicates.

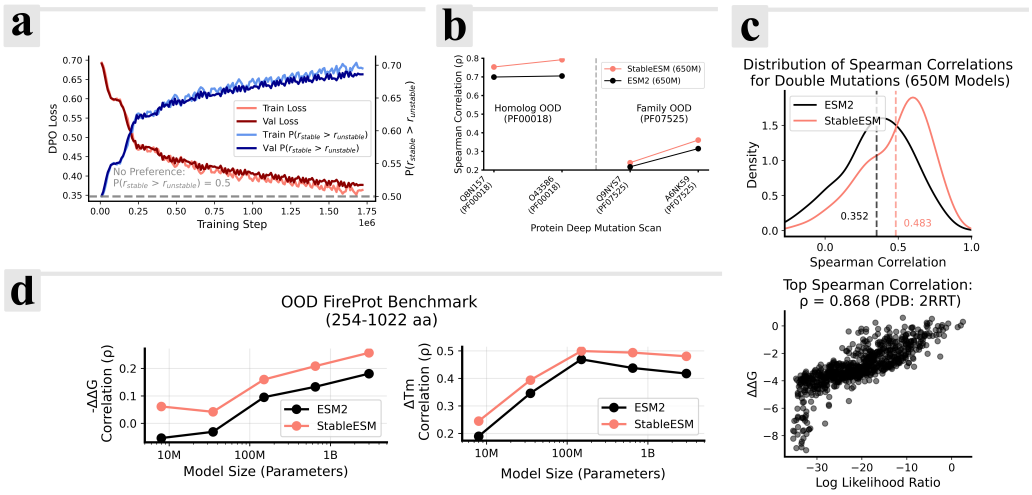

Figure S2: **StableESM training dynamics and performance validation across scales and tasks.** **(a)** Direct preference optimization (DPO) training curves showing successful preference learning without overfitting. **(b)** Zero-shot generalization to held-out protein domains across different families. **(c)** Transfer to higher-order effects: distribution of Spearman correlations for double mutations showing improved median performance. **(d)** Restored scaling laws on out-of-distribution FirePrebt benchmark (254-1022 amino acids) where StableESM continues improving with scale while ESM2 plateaus.

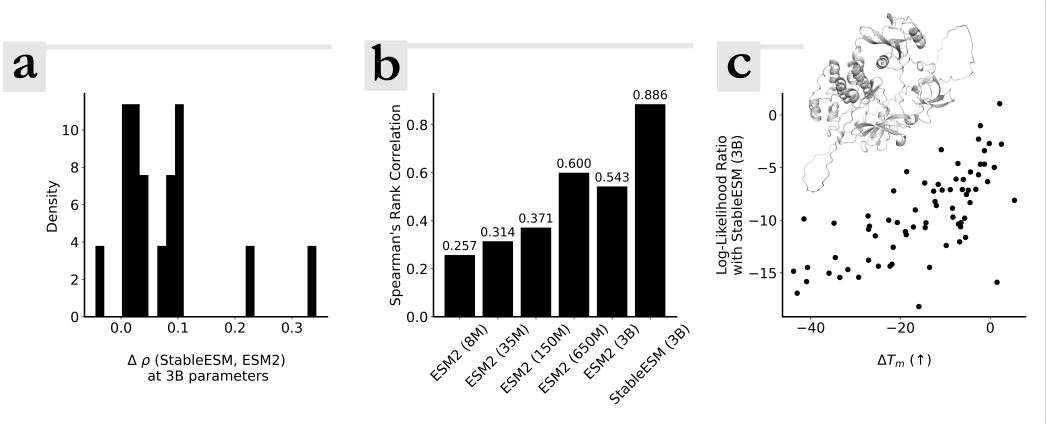

Figure S3: **Preference alignment enables robust generalization to out-of-distribution protein domains.** **(a)** Distribution of of Spearman correlation coefficient improvements of StableESM relative to ESM2 $\Delta\rho\big(\text{StableESM, ESM2}\big)$ at 3B parameters for FirePrebt domains, with 94% showing improvements. **(b)** Model scaling analysis for Tyrosine-protein kinase Fyn (537 amino acids) demonstrates a 48% elevation of the Spearman correlation coefficient to $\rho = 0.886$ for StableESM relative to $\rho = 0.600$ the top performing 650M parameter ESM2 model. **(c)** In silico validation with AlphaFold protein structure (UniProt: P06241), demonstrating generalization beyond training length distribution.

B.1.3 Copper Activation and Heat Treatment

Since laccase is activated by copper, we added copper sulfate ($CuSO_4$) into the CFE reaction after 20 hours of incubation. Specifically, 5 mM of $CuSO_4$ was added to the CFE system to reach a final concentration of 400 μM and incubated at room temperature (25°C) for 2 hours in a pH 5.5 buffer environment (50 mM PIPES, 20 mM NaCl). After copper activation, each laccase mutant

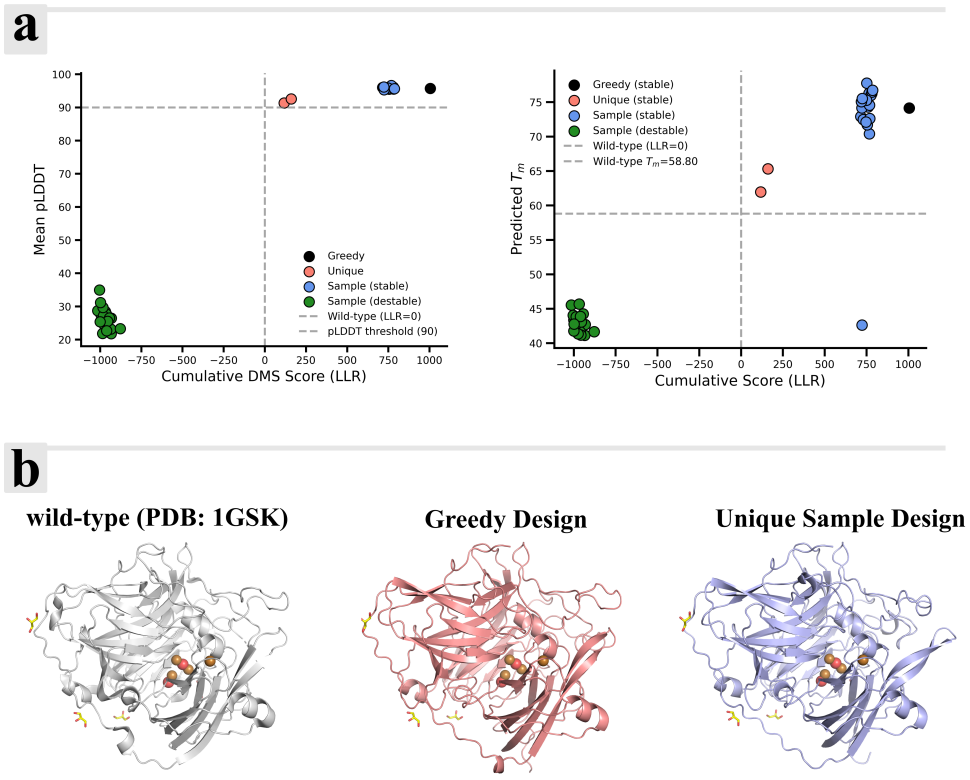

Figure S4: **In silico validation of StableESM through thermostable protein design. (a)** In silico results comparing greedy vs. preference-aligned design strategies, showing StableESM identifies beneficial mutations achieving higher predicted melting temperatures and stability using TempBERTura [11]. **(b)** Designed protein structures computationally predicted to possess improved thermostability across multiple variants, validating practical utility for experimental protein engineering.

was subjected to heat treatment at three temperatures (85°C, 65°C, 25°C) for 10 minutes to assess thermostability.

B.1.4 Multicopper Oxidase Activity Assay

The MCO activity was measured using a colorimetric assay of 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) oxidation, which turns blue upon oxidation. Colorimetric assays were performed in a pH 5.5 buffer environment (50 mM PIPES, 20 mM NaCl) containing 4 mM ABTS. The absorbance at 420 nm was measured once per minute for 30 minutes. Each sample was measured in four replicates.

B.2 Detailed Performance Analysis and Generalization Studies

StableESM demonstrates comprehensive improvements across multiple evaluation dimensions. Using Spearman correlation (ρ) between experimental values and log-likelihood ratio (LLR), StableESM generalizes effectively to unseen homolog domains and marginally improves on held-out protein families (Figure S2b). For both $\Delta\Delta G$ and ΔT_m correlations, StableESM shows continued improvement with scale while ESM2 plateaus (Figure S2d).

The model exhibits improved performance on double mutation effect prediction despite training only on single mutations (Figure S2c), with mean performance improving from 0.352 (ESM2) to 0.483 (StableESM) at 650M parameters and best-case performance reaching exceptional correlation

($\rho = 0.868$). This transfer to higher-order mutational effects demonstrates that preference alignment teaches generalizable principles rather than memorizing specific mutation patterns.

Case studies reveal dramatic improvements on long proteins: Tyrosine-protein kinase Fyn (537 amino acids) correlation increases from 0.543 (ESM2 3B) to 0.886 (StableESM 3B) (Figure S3b,c). This proves that preference alignment enables effective utilization of larger model capacity while achieving robust generalization beyond the training distribution to completely unseen protein families and lengths.

B.3 Large-Scale Protein Design Validation through Machine Learning-Directed Evolution

To demonstrate StableESM's utility for large-scale protein engineering, we applied the model to design thermostable multicopper oxidase variants using machine learning-directed evolution (Algorithm 1) on an unseen 513 amino acid protein (PDB: 1GSK). This protein was not represented in any form within the preference alignment dataset, providing a stringent test of generalization capabilities.

We evaluated multiple design strategies to assess StableESM's design space exploration: (1) StableESM greedy sampling, selecting positions with highest predicted stabilizing effects; (2) StableESM unique sampling, targeting positions where StableESM predicts stabilizing mutations but ESM2 does not ($\text{LLR}_{\text{StableESM}} \geq 0$ and $\text{LLR}_{\text{ESM2}} < 0$); and (3) ESM2 greedy sampling as baseline. As a negative control, designs incorporating destabilizing mutations (LLR < 0) resulted in poor structural predictions and low predicted thermal stability, confirming the importance of stabilizing mutation selection.

Using TempBERTura [11] for thermal stability evaluation, all stabilizing design strategies demonstrated substantial improvements over the wild-type protein. The wild-type sequence was classified as non-thermophilic with high confidence (0.993; $p = 0.007$) and predicted T_m of $58.8°$C. In contrast, all designed variants achieved thermophilic classification with dramatically improved thermal stability predictions (Figure S4a).

StableESM greedy design achieved the highest predicted T_m of $74.2°$C with high confidence (0.978), representing a $15.4°$C improvement over wild-type. The StableESM unique sampling strategy achieved a predicted T_m of $61.9°$C with even higher confidence (0.981), demonstrating that StableESM identifies stabilizing mutations missed by the base model. ESM2 greedy design reached a similar predicted T_m of $73.7°$C but with substantially lower confidence (0.686), suggesting less reliable thermal stability predictions and highlighting the value of preference alignment for confident predictions.

All designs demonstrated substantial novelty from the wild-type sequence, with StableESM greedy (62.2% similarity, 194 edits) and unique (62.4% similarity, 193 edits) variants showing comparable divergence to ESM2 designs (62.0% similarity, 195 edits). Critically, StableESM and ESM2 greedy designs differed significantly from each other (86.2% similarity, 71 mismatches) despite making similar numbers of edits ($N = 400$), demonstrating that preference alignment fundamentally alters sequence space exploration even for completely unseen proteins distant from the training data.

Structural quality assessment using ColabFold confirmed that all stabilizing designs maintained high predicted structural integrity with mean pLDDT scores above 91 (Figure S4b). This validates that StableESM learns generalizable stability principles that transfer to distant protein families not represented in the preference alignment dataset, enabling the design of novel thermostable variants through systematic alteration of protein fitness landscape exploration.