nature methods

Article

https://doi.org/10.1038/s41592-026-03137-3

Aligning protein-generative models to experimental fitness with ProteinDPO

Received: 15 July 2025

Accepted: 12 May 2026

Published online: 14 August 2026

Check for updates

Talal Widatalla[1,2], Ashir A. Borah[2,3], Samuel H. King[1,2], Claudia L. Driscoll[1,2], Rafael Rafailov[1] & Brian L. Hie[1,2]

Biological generative models can predict biological functions without task-specific training data but often under-perform specialized models. This is due to a fundamental 'alignment gap', where the rules learned during unsupervised training are not related to the function of interest. Here we demonstrate how to provide task-specific information without losing the general knowledge learned during pretraining by using direct preference optimization to align a structure-conditioned protein language model to preferentially generate stable protein sequences. Our aligned model, ProteinDPO, achieves stability prediction competitive to task-specific models and consistently outperforms unsupervised and fine-tuned versions of the model. Notably, ProteinDPO generalizes beyond its training data to enable stabilization and improved binding affinity prediction of large multichain protein complexes. When applied to stabilization of the hemagglutinin trimer, a primary component of influenza vaccines, ~80% of designs achieve increased or similar stability compared with the native hemagglutinin and up to 32 °C improvements from recently emerged mammalian strains. Our results demonstrate how to augment generative models with biophysical information and, more broadly, provide a general framework for the alignment of biological foundation models.

Machine learning models for biological structure and function prediction have seen tremendous improvements due to large-scale unsupervised training on databases such as UniProt and the Protein Data Bank (PDB)[1–7]. One such class of models are structure-informed language models or 'inverse-folding' models, which generate protein sequences that are compatible with a given structure[5,8,9]. These models are surprisingly capable of guiding improvements in binding affinity[10] and improving protein expression, solubility and stability[11] despite not observing these properties, nor any explicit biophysical information, during training.

These capabilities most likely emerge because a diverse training dataset naturally contains examples that possess these properties. For example, many of the structures in the PDB are of stable or stabilized proteins[12], so models trained on the PDB demonstrate modest zero-shot stability prediction. However, evolutionary selection does not only optimize natural proteins for stability. Consistent with this, unsupervised models are still inferior at stability scoring compared with supervised models such as ThermoMPNN[13]. Using a model trained on unlabeled data for stability-related tasks results in an 'alignment gap': a disagreement between the information learned by the unsupervised model and the specific definition of protein fitness.

A solution would be to explicitly provide the generative model with task-specific information while retaining the model's generative capabilities. A simple way of doing this is with supervised fine-tuning (SFT) in which we further train the model on a curated set of positive examples according to a property of interest[14,15]. However, SFT only maximizes the likelihood over a set of positively labeled samples, making it easy for the model to overfit to the fine-tuning dataset and lose the general information learned during unsupervised pretraining. More broadly, finding better ways to incorporate experimental

[1]Stanford University, Stanford, CA, USA. [2]Arc Institute, Palo Alto, CA, USA. [3]University of California, San Francisco, San Francisco, CA, USA.
e-mail: brianhie@stanford.edu

fitness information into protein-generative models is an important open question.

Direct preference optimization (DPO) is an algorithm for lightweight alignment of generative models[16]. DPO enables us to explicitly inform the model of examples that may appear similar at a surface level but are actually very different when considered according to a property of interest (Fig. 1a,b). This is especially useful in biological settings where single amino-acid changes can have global effects on stability and structure[17]. By providing the model with positive and negative labels of similar proteins, DPO discourages overfitting to a set of positive examples and uses the full fitness landscape to inform the model's predictions.

In this study, we demonstrate alignment of a generative model with experimental fitness data by using DPO to imbue stability information into a pretrained structure-conditioned protein language model, ESM-IF1[5], using a curated dataset from experimental measurements from the 'Megascale' stability dataset[18] containing approximately 660,000 sequence variants across 403 diverse protein domains (Fig. 1c,d; Methods). Our aligned model, which we refer to as ProteinDPO (Protein Direct Preference Optimization), substantially improves the stability scoring of vanilla (unsupervised) ESM-IF1 and a model trained via SFT on a subset of stable variants from Megascale (Fig. 1d, Extended Data Table 4 and Extended Data Fig. 6) and even performs competitively to supervised and physics-based models on single-site substitution stability prediction. Despite being trained solely on the relative changes in stability ($\Delta\Delta G$) of one or two substitutions in small protein monomers, ProteinDPO generalizes to improved binding affinity scoring of large complexes and thermal stability of highly diverse and variable length antibodies, which are regimes inaccessible to several supervised models. Beyond scoring, generating new sequences with ProteinDPO results in designs that have higher stability compared with the native protein and to generations produced by the vanilla and SFT models when evaluated with an in silico force field.

To illustrate the practical utility of this approach in addressing urgent biological challenges, we applied ProteinDPO to the stabilization of H5N1 influenza hemagglutinin (HA)—a critical need given the recent emergence of H5N1 variants across mammalian populations (including human cases) and their pandemic potential[19]. HA undergoes a conformational change to enable fusion of the virus to the host membrane, and stabilization of the prefusion conformation is important for efficacious immune response[20]. However, the prefusion state presents a formidable obstacle for vaccine development due to its propensity to be unstable, have poor expression, and fail to trimerize[20,21]. We used ProteinDPO to generate variants that improve the stability of the prefusion state and maintain antigenicity, which is a challenging problem due to the vast search space of possible HA variants. Despite this, with only 45 total designs, ProteinDPO identified 27 stabilizing variants, achieving up to a 17 °C improvement in melting temperature while preserving binding to broadly neutralizing antibodies. We also show these results can generalize to recent H5N1 strains of concern, achieving up to a 32 °C improvement in melting temperature in variants from 2024. These results demonstrate ProteinDPO's capacity for rapid response to emerging pathogens and establish DPO as an effective method for aligning a generative model with an experimental fitness landscape.

Results

In large language models of text, directly sampling from a pretrained model (such as GPT-4) is not practically useful, as it produces sequences that resemble raw unstructured data from the Internet (GPT-4's training data). In practice, these models need to be aligned with human 'preference' labels to produce useful responses to a user's prompt; for example, aligning GPT-5 to become a helpful question-answering system like ChatGPT[22]. Alignment is accomplished with feedback-based reinforcement learning (RL) or with lighter-weight alternatives such as DPO[16] (see the Methods for additional details on this procedure).

We repurposed DPO, originally designed to align the responses of natural language models to text prompts using human preference labels (Fig. 1a), to instead align the generation of protein sequences ('responses') given their native structure ('prompts') using experimental stability measurements ('preferences') (Fig. 1b). We implemented the previously described DPO reward models that utilize preference-ordered pairs and rankings as input (referred to as 'paired' and 'ranked', respectively)[16]. In this study, we also derive a third, new DPO training objective ('weighted') that directly incorporates scalar labeled data (Methods).

Megascale training and relative stability prediction performance

All ProteinDPO models were trained on the Megascale dataset with the DPO objective using native backbones and stability ordered pairs, rankings or scalar-labeled sets of sequence variants as input, depending on the reward model (Methods). To prevent data leakage between training and evaluation splits, we structurally clustered the variants by native structure with FoldSeek[23] and randomly distributed clusters across splits. To determine whether DPO and SFT training improved upon the zero-shot capability of vanilla ESM-IF1 to score changes in stability upon mutation ($\Delta\Delta G$) (Fig. 2a), we evaluated the average correlation of sequence likelihoods with $\Delta\Delta G$ across each native protein in the holdout set (Fig. 2b). Although the SFT model modestly improves (Pearson $R = 0.59$, Spearman $\rho = 0.57$) from vanilla ESM-IF1 (Pearson $R = 0.55$, Spearman $\rho = 0.53$), ProteinDPO clearly surpasses both models (Pearson $R = 0.72$–0.73, Spearman $\rho = 0.69$–0.72). We also report the area under the receiver operating characteristic curve (AUROC) similarly averaged across each protein in the holdout set, representing the models' ability to distinguish between stabilizing ($\Delta\Delta G < 0$) or destabilizing ($\Delta\Delta G > 0$) variants and observe the same trend in the vanilla (AUROC of 0.74), SFT (AUROC of 0.76) and ProteinDPO models (AUROC of 0.82–0.84).

In addition, we sought to evaluate performance specifically on Megascale variants that have multiple mutations relative to the native sequence. Along with SFT and vanilla ESM-IF1, we compared with ThermoMPNN, a state-of-the-art supervised model based on ProteinMPNN[9] also trained on Megascale. For this comparison, we retrained our models on the training split from Diekhaus et al.[13] and evaluated on the double-mutation variants belonging to their holdout proteins. Notably, ProteinDPO can be trained with double-mutation variants (although we did not for this evaluation for a fair comparison) and can also naturally evaluate the likelihood of multiple mutations. By contrast, the architecture and training procedure of ThermoMPNN restricts the model to single mutations.

We found that all models perform worse when scoring variants with multiple mutations (Pearson R and Spearman $\rho = 0.3$–0.5) (Fig. 2c) compared with their ability to score single-site substitutions (Pearson R and Spearman $\rho = 0.5$–0.8)[13], suggesting the nonlinear effects of simultaneous mutations, which can energetically influence one another[24], pose a challenge to stability prediction. When we score these variants as a linear combination of single mutations by adding likelihoods (ProteinDPO) or scores (ThermoMPNN) of only the individual mutations, ProteinDPO underperforms ThermoMPNN (ProteinDPO Pearson $R = 0.36$–0.40, ThermoMPNN Pearson $R = 0.41$) (Extended Data Table 1). However, when we use ProteinDPO to score variants by the likelihood summed across the whole sequence, all ProteinDPO models surpass ThermoMPNN (Pearson $R = 0.44$–0.48) (Fig. 2c), which is fundamentally constrained to the 'additive' regime. This suggests ProteinDPO, even when trained only on single-mutation variants, benefits from synthesizing information across the full sequence to capture nonlinear effects of multiple mutations.

https://doi.org/10.1038/s41592-026-03137-3

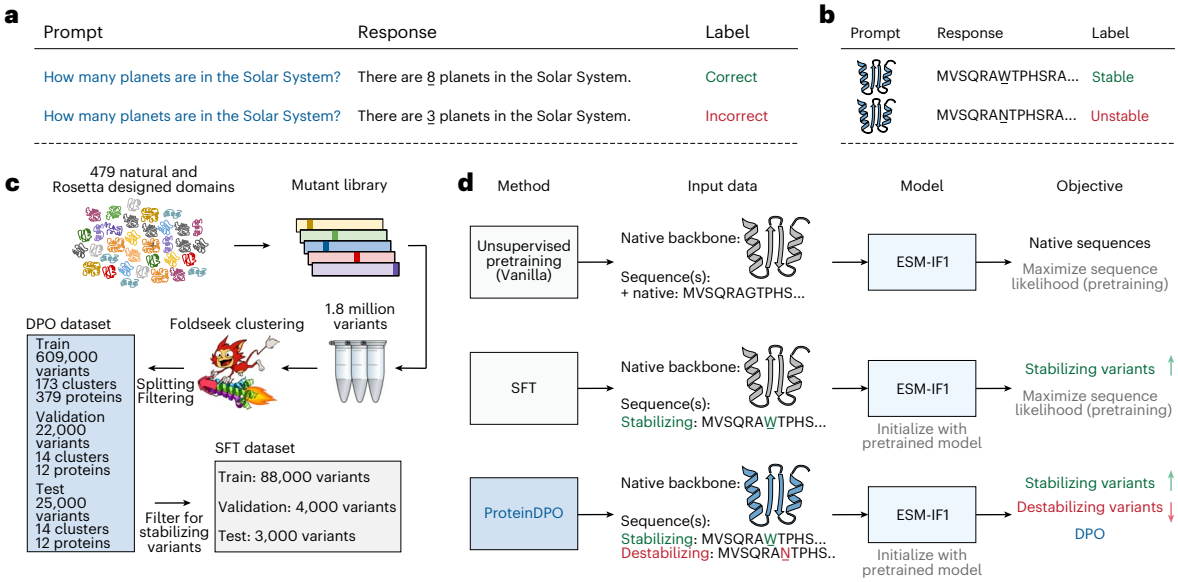


Fig. 1 | Dataset curation and model training. a, DPO was developed to align the responses of language models to labeled human preferences given a text prompt. **b**, Analogously, we align a structure-conditioned language model to perform thermostable sequence design using protein backbone conditioning as a 'prompt', sequence generation as a 'response' and experimental fitness as 'preference' data. **c**, Overview of the curation of DPO and SFT datasets from the Megascale dataset, which contains stability measurements of ~1.84 million sequence variants from 479 protein domains from the PDB and Rosetta design[18,46,47]. While filtering out unreliable $\Delta\Delta G$ measurements, we also structurally cluster the protein domains with FoldSeek at a threshold of 0.5 similarity to alleviate leakage between training splits, resulting in a final dataset with ~660,000 measurements across 403 domains. From these same splits used for DPO, variants with stability superior to the native protein ($\Delta\Delta G < 0$) were selected for SFT. **d**, Overview of the training procedure for the models compared in this study: vanilla ESM-IF1, SFT ESM-IF1 and ProteinDPO. Vanilla ESM-IF1 was developed and trained by Hsu et al.[5] with maximum likelihood training on structure-conditioned, autoregressive token prediction with protein domains from PDB (CATH)[48] and AlphaFold2-predicted[2] structures (AF2)[5]. The SFT model was trained in the same manner but initialized with the vanilla weights and trained on stabilizing variants from the curated Megascale dataset. ProteinDPO was trained on the curated Megascale on variants of all fitness with a DPO objective (Methods). Foldseek logo reproduced under GNU General Public License v3.0.

Comparison with other methods for predicting relative stability

Next, we evaluated if ProteinDPO, trained on Megascale proteins (40–72 residues), can generalize to other datasets with larger (up to ~300 residues) and unseen folds: S669 and FireProt. These datasets are compilations of experimental stability measurements of derived from several small-scale mutagenesis experiments of native proteins (Fig. 1a) and are popular benchmarks of protein stability predictors. To alleviate data leakage with S669, we retrained ProteinDPO models on a S669-Homolog-free Megascale training set from Cutrello et al.[25], and to alleviate leakage with FireProt, we evaluated on a Megascale-Homolog-free FireProt curation from Diekhaus et al.[13].

We observed that ProteinDPO ranks highly amongst supervised and physics-based $\Delta\Delta G$ predictors (Fig. 2d), some of which were also trained on Megascale. In addition, we find ProteinDPO is competitive with ThermoMPNN[13,25] on the respective S669[26] (ProteinDPO Pearson $R = 0.44–0.47$, ThermoMPNN Pearson $R = 0.43$) (Fig. 2d) and FireProt[27] (ProteinDPO Pearson $R = 0.6$, ThermoMPNN Pearson $R = 0.65$) (Fig. 2e) test sets. Interestingly, we find the SFT model, despite improving upon vanilla ESM-IF1 on Megascale holdout prediction (Pearson R increase of 0.03, Spearman ρ increase of 0.04), does slightly worse on the FireProt (Pearson R decrease of 0.05, Spearman ρ decrease of 0.04) and S669 datasets (Pearson R decrease of 0.01), while the improvement of ProteinDPO over the Vanilla model generalizes to FireProt (Pearson R and Spearman ρ increases of 0.13–0.14) and S669 (Pearson R and Spearman ρ increases of 0.13–0.16) as well, evidence that the SFT model has overfit to the training data while ProteinDPO has incorporated generalizable stability information.

Generalization to diverse design tasks

Motivated by previous work in which ESM-IF1 guided antibody affinity maturation by ranking sequence likelihoods conditioned on the respective antibody–antigen complex backbone[10], we were interested if our method, optimized for stability, could generalize to binding affinity given that they are both determined by similar entropic and enthalpic interactions. To this end, we evaluate on two datasets, SKEMPIv2[28] and AB-Bind[29], which consist of experimentally measured binding affinities for thousands of sequence variants across hundreds of protein–protein complexes, with AB-Bind containing exclusively antibody–antigen complexes (Fig. 3a). Due to SKEMPIv2 containing very few variants per complex (as low as one), we report the correlation of all model likelihoods with all affinity measurements, while for AB-Bind we report correlations averaged across each complex. Surprisingly, despite only training on the $\Delta\Delta G$ of small protein monomers, ProteinDPO has modestly increased scoring of the binding affinity of large protein–protein complexes (Pearson R increase of 0.02–0.04, Spearman ρ increase of 0.02–0.04 and AUROC increase of 0.04–0.05) and larger performance gains when scoring antibody–antigen complexes (Pearson R increase of 0.04–0.08, Spearman ρ increase of 0.03–0.10, AUROC increase of 0–0.02) (Fig. 3b). Similar to previous results (Fig. 2), the SFT model fails to generalize to binding-related tasks.

Encouraged by our finding that using whole-sequence likelihoods with our method improves upon addition of single-mutation likelihoods for scoring the relative stability ($\Delta\Delta G$) of double-mutation variants in Megascale, we sought to evaluate if ProteinDPO can use whole-sequence reasoning to rank absolute stability of sequences within a given fold, particularly antibodies. Antibodies are relevant therapeutic proteins for a wide array of diseases, for which high thermal stability is important for developability and storage[30–32]. To this end, we evaluate on a dataset of thermal melting points (T_m) of 483 clinical and human-derived antibodies from Jain et al. and Shehata et al.[33,34] (Fig. 3c). Despite only training on relative stability data of small monomers, ProteinDPO improves (Pearson R and Spearman ρ increases of 0.08–0.12) on the ability of vanilla ESM-IF1 in ranking the differences in

https://doi.org/10.1038/s41592-026-03137-3

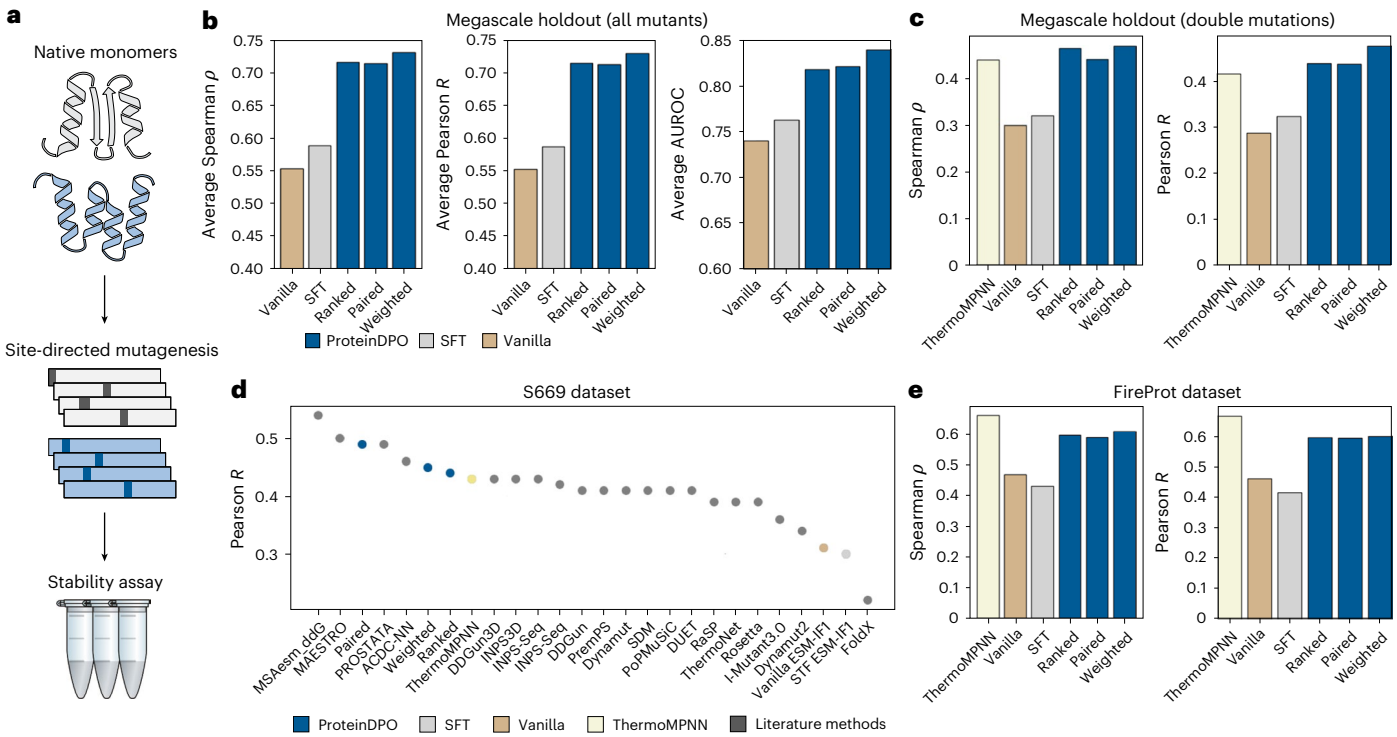


Fig. 2 | ProteinDPO achieves competitive mutational stability prediction. **a**, We benchmarked models using experimental protein stability datasets FireProt[27] and S669[26]. These dataset are compiled from small-scale mutagenesis experiments across diverse monomeric proteins, in which certain residues are mutated, variants are expressed and their stability is measured experimentally. **b**, Performance of models on predicting stability changes upon mutation ($\Delta\Delta G$) in curated Megascale holdout data, including variants with one or two substitutions. We report the correlation of sequence likelihoods with their experimentally measured $\Delta\Delta G$ averaged across each native protein in the holdout data. AUROC measures the models' ability to classify between variants which increase ($\Delta\Delta G < 0$) or decrease ($\Delta\Delta G > 0$) stability compared with the native sequence. Distributional plots of the individual values are included in Extended Data Fig. 1. **c**, Trained on the same training data as ThermoMPNN, we show the correlation between all model predictions (ThermoMPNN) or likelihoods (Vanilla, SFT and ProteinDPO) with $\Delta\Delta G$ of all double-mutation variants in ThermoMPNN holdout data[13]. As ThermoMPNN cannot score multiple mutations, we add the predicted $\Delta\Delta G$ of the two respective single-mutations as a proxy. **d**, Performance of ProteinDPO, SFT, vanilla ESM-IF1 and literature methods on the S669 dataset. Performance of literature methods are provided from Cuturello et al.[25,26]. **e**, Performance of models on predicting stability changes after single-site substitutions ($\Delta\Delta G$) with a version of the FireProt dataset[27] filtered of proteins homologous to those in Megascale[13].

thermal melting points (T_{mid}) across diverse clinical and human-derived multichain antibodies. Notably, other supervised models trained on Megascale to predict $\Delta\Delta G$ (such as ThermoMPNN) cannot evaluate proteins with multiple chains (such as most antibodies) or produce absolute stability metrics.

Beyond merely a supervised scoring function, ProteinDPO is also a generative model. We were therefore interested in testing if ProteinDPO would be useful for stabilizing sequences for given protein backbones, and whether it could do so more favorably than the original ESM-IF1 model. This is a complex task, as combining favorable single-mutations to design a sequence that is more stable than the native protein is non-trivial due to the nonlinear and higher order interactions among protein residues[24]. Here, we generate protein sequences for three diverse, native backbones (Fig. 3d): staphylococcal nuclease[35] (PDB: 1STN), PGK[36] (PDB: 1PHP), and the membrane protein claudin-15[37] (PDB: 4P79). We computed the distribution of the average Rosetta force-field energy (Rosetta energy units, REU) across replicate runs for 500 generations for each model and native backbone, as well as the energy of each native sequence, all derived by modeling the mutated sequence onto the native backbone via the Rosetta cart_ddg protocol[38]. Although Rosetta energy can be inaccurate and is not a substitute for experimental testing, it is a widely used independent evaluator of machine learning generations due to its reliance on physics-based parameters rather than shared training data with machine learning models[39,40].

Despite only training on the stabilities of small proteins (40–75 residues), sequences generated by ProteinDPO for much larger backbones (136–394 residues) are predicted to be substantially more stable (lower energy) than vanilla ESM-IF1, the SFT model and the native sequence. Surprisingly, despite having lower similarity to the native sequence than SFT and vanilla ESM-IF1, sequences generated by ProteinDPO have confident ESMFold-predicted structures (pLDDT >80) (Fig. 3d) that closely match the native structure (Extended Data Table 2). Although these sequences were not experimentally verified to express, this functions as a computational 'sanity-check' that the stability optimized ProteinDPO has not lost the rules of structure-conditioned sequence design learned from the unsupervised pretraining of ESM-IF1.

Prefusion stabilization of H5 influenza HA

Protein stabilization has particularly critical applications in vaccine development, where rapid-response capabilities and thermal stability are essential for their global deployment[41]. For influenza vaccines, maintaining the prefusion conformation of HA helps elicit protective antibodies, as the native protein undergoes dramatic conformational changes during viral entry that can expose undesirable epitopes[20]. In addition, temperature stability is vital for vaccine storage and distribution, especially in resource-limited settings where cold-chain maintenance is challenging, and vaccines may be exposed to elevated temperatures[41].

Traditional stabilization approaches rely on rational design or resource-intensive screens and, therefore, remain imperfect solutions. These manual changes to improve stability often leverage proline insertions, new disulfide bond formation and core stabilization through

https://doi.org/10.1038/s41592-026-03137-3

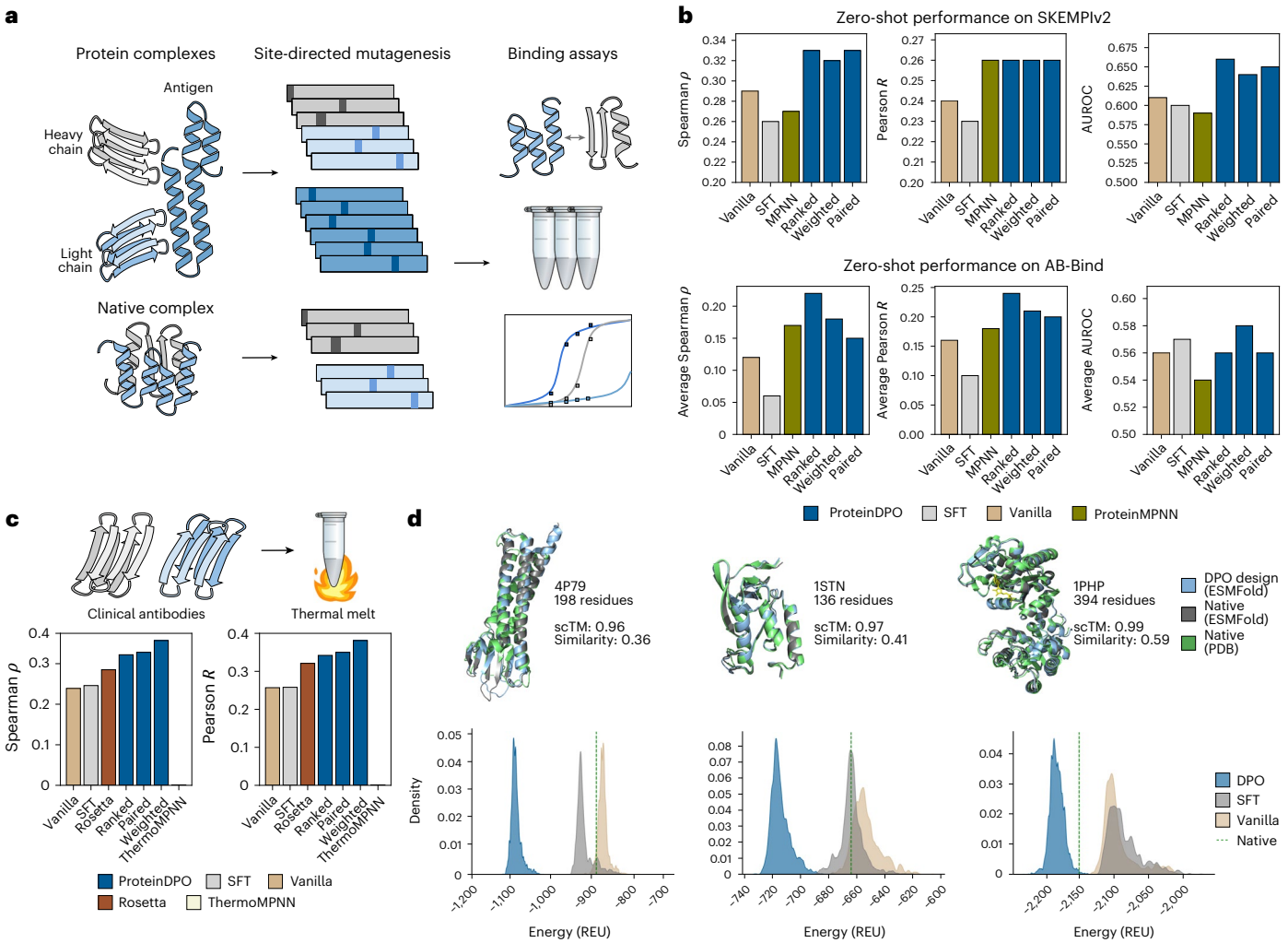


Fig. 3 | ProteinDPO generalizes beyond stability prediction to diverse tasks. **a**, We benchmarked models using binding affinity mutagenesis studies, which mutate one or more protein chains that form a bound complex and subsequently measure the resulting binding affinity of the complex. **b**, Zero-shot correlation and AUROC of model likelihoods with protein–protein binding affinity from the SKEMPIv2 dataset[28]. Zero-shot correlation and AUROC of complex-conditioned model likelihoods with $\Delta\Delta G$ of antibody–antigen binding from the AB-Bind dataset[29], averaged across each complex. AUROC measures a model's ability to classify mutations that increase or decrease affinity relative to the original complex. Distributional plots of the individual values for the AB-Bind dataset are included in Extended Data Fig. 1. **c**, We benchmarked models using datasets of thermal unfolding measurements (determination of the temperature which unfolds half the concentration of a protein) of antibodies. Correlation of model likelihoods (Vanilla, SFT, ProteinMPNN and ProteinDPO) or predicted values (ThermoMPNN) with experimentally measured thermal melting points of clinical and naturally derived mAbs curated from Shehata et al.[33] and Jain et al.[34]. **d**, We evaluate the self-consistency and confidence of Paired model ProteinDPO designs for PDB structures 1STN, 1PHP and 4P79, comparing with experimental[6] and ESMFold predicted[1] structures. We show the experimental structure of the native sequence in green, the predicted structure of the native sequence in gray and the predicted structure of the ProteinDPO design in blue. Self-consistency TM (scTM) is the structural similarity score (TM) of the native sequence predicted structure ProteinDPO sequence predicted structure; pLDDT is the average ESMFold[1] confidence (predicted error) for the structure of the generated sequence. Similarity is the percent of shared residue identity between the generation and native sequence. Average in silico free energy of protein folding (ΔG) for sequences from ProteinDPO, vanilla ESM-IF1 and SFT ESM-IF1 in REU compared with the native sequence derived from Rosetta *cart_ddg* protocol[38] with five replicate runs for each sequence. 500 samples were taken from each model across five sampling temperatures within [0.01, 1] with no additional filtering. Lower energy corresponds to higher stability.

cavity filling[42]. However, these changes can often come at the cost of reduced protein expression, altered surface properties and disrupted native folding patterns when applied naively without considering the context of the full protein structure. Structure-conditioned language models, such as ProteinDPO, allow us to sample mutations that are likely to preserve the global protein structure—a key component for vaccine design, where preservation of antigenicity is critical.

We therefore applied ProteinDPO to the task of stabilizing the prefusion state of influenza HA (Fig. 4a), a primary component of modern influenza vaccines and consequently an important target of protein stabilization efforts[20]. We specifically target the H5 subtype of HA, given the emergence of highly pathogenic H5N1 influenza A virus that has recently been found in mammalian populations, including sporadic human cases[19]. We use the structure from the HA of A/Vietnam/1203/04 (H5N1) (referred to as Vietnam 2004) as the starting point for our stabilization campaign (PDB: 6CFG), using ProteinDPO to rank all possible single-site substitutions based on their effects on ProteinDPO likelihood.

Remarkably, despite only using the H5 HA structure from Vietnam 2004, ProteinDPO recovers known stabilizing mutations to H1 and H3 subtypes from the literature. Specifically, among the top-30 substitutions tested, six (H26W, K51M, K51I, E103I, E103L and N95L)

https://doi.org/10.1038/s41592-026-03137-3

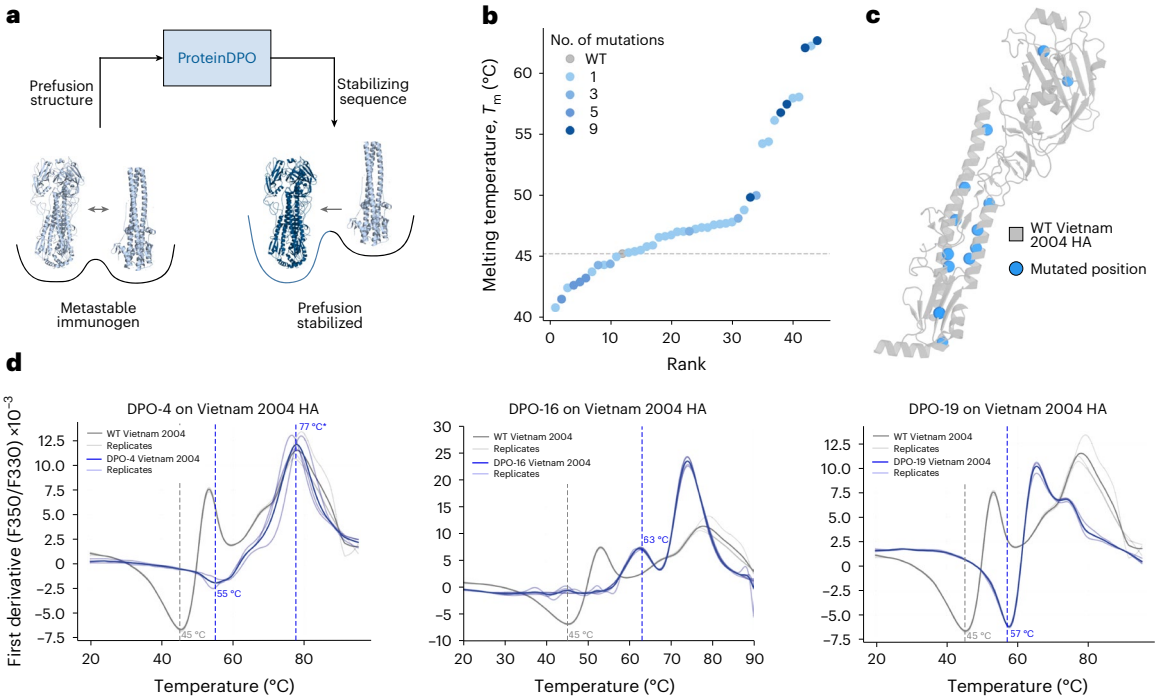


Fig. 4 | ProteinDPO enables rapid zero-shot stabilization of H5 influenza HA. **a**, Illustration of our prefusion stabilization campaign. Mutations are introduced by conditioning ProteinDPO with the prefusion structure of a fold-switching protein in order to stabilize this conformation and prevent a transition to the post-fusion state. **b**, Rank-ordered thermal stability of ProteinDPO variants applied to 2004 Vietnam H5N1 in melting temperature (T_m) as determined by differential scanning fluorimetry (DSF) and colored by the number of mutations. **c**, Mutation positions of ProteinDPO variants plotted onto an experimental structure (PDB: 6CFG) of A/Vietnam/1203/04 (H5N1) (Vietnam 2004). **d**, The first derivative of the fluorescence ratio is plotted versus temperature as determined by DSF of three stabilizing nine-substitution variants: DPO-4, DPO-16 and DPO-19. Thermal melting events correspond to positive or negative peaks of the first-derivative curve and are labeled with gray (native) and blue (variant) dashed lines. *Although we report the first thermal melting event of DPO-4 as 56 °C, as determined by NanoTemper software, the dynamic range at this temperature is much lower than the first major melting event at 77 °C. WT, wild type.

exactly match the eight stabilizing mutations (BY94D and BH106R were not identified) reported in previous studies by Milder et al. and Yassine et al., which each report four mutations respectively[20,21]; an additional 5 of these top-30 single mutations occur at identical positions as the above 6 mutations. We then compared the ability of the Vanilla and SFT models, as well as ThermoMPNN[13] and its predecessor ProteinMPNN[9], to recover these eight stabilizing mutations from literature (Extended Data Fig. 2). In brief, this is evaluated by taking the percentage of stabilizing mutations that lie within the top-*n* single-mutations, upon scoring and ranking every possible single-mutation with the given model. While other models were able to recover one mutation, ProteinDPO has a higher recovery rate across all top-*n* levels. Most strikingly, ProteinDPO successfully identified critical residues (H26, K51 and E103) which make up the 'pH-switch' hypothesized by Milder et al. to modulate the pH-sensitive transition to the post-fusion state[20]. Importantly, the model was not trained on any data from HA and identified these residues in a zero-shot manner despite the >500 residue search space of HA. The remaining mutations primarily consist of cavity-filling substitutions in the helical core involving the introduction of leucine, tryptophan, tyrosine and phenylalanine residues that enhance hydrophobic core packing[43] (Fig. 4c and Extended Data Table 3). In total, these results demonstrate that the model has learned to recognize fundamental structural principles that govern protein conformational stability.

We experimentally measured the thermal melting profiles with differential scanning fluorimetry on an initial round of the top-30 ProteinDPO-identified substitutions. Of these, 19 variants obtained greater thermal stability (T_m) than the wild-type HA (>1 °C improvement), and 7 variants obtained similar melting temperature (<1 °C difference) (Fig. 4b). We then conducted a similar analysis as done for literature mutation recovery, but instead comparing models' ability to identify 6 of the 30 ProteinDPO-identified single-substitution variants which were found to be significantly stabilizing (≥9 °C) (Extended Data Fig. 3). Only the SFT and Vanilla models were able to recover mutations in their top 200; although, it is possible other top-ranked mutants from these baseline models are also stabilizing. In addition, while all models are unable to rank the thermal stability (T_m) of these 30 variants (Extended Data Fig. 4), the SFT model and ThermoMPNN show strong ability (F1 score >0.75) to classify them as stabilizing or destabilizing (Extended Data Fig. 5), despite crucially failing to capture most highly stabilizing mutations (Extended Data Fig. 3).

Beyond single-substitutions and aiming to discover synergistic effects among beneficial mutations, we used ProteinDPO to score all possible combinations of size three, five and nine (the total number of unique single-substitution positions was nine) mutations among non-deleterious mutations from the first round. For combinations of three and five, we additionally included a variant consisting of a simple combination of the top-n stabilizing variants. In line with in silico results suggesting the ability of ProteinDPO to generalize well to multiple mutations (Figs. 2c and 3c), all nine-substitution variants achieved greater thermal stability than the wild type, with the most successful variants achieving improvements of up to 17 °C (Fig. 4b). While the best nine-substitution variant did not significantly surpass the best single-substitution variant, nine-substitution variants were 9.8 °C more stable on average than single-substitution variants. All variants and their associated melting temperatures are reported in Extended Data Table 3. We show the thermal shift profiles of three variants containing nine substitutions (DPO-4, DPO-16 and DPO-19) demonstrating clear improvement in thermal stability from the wild type, with 10 °C (with a 32 °C change in the major thermal shift transition), 17 °C and 12 °C increases in T_m, respectively (Fig. 4d).

https://doi.org/10.1038/s41592-026-03137-3

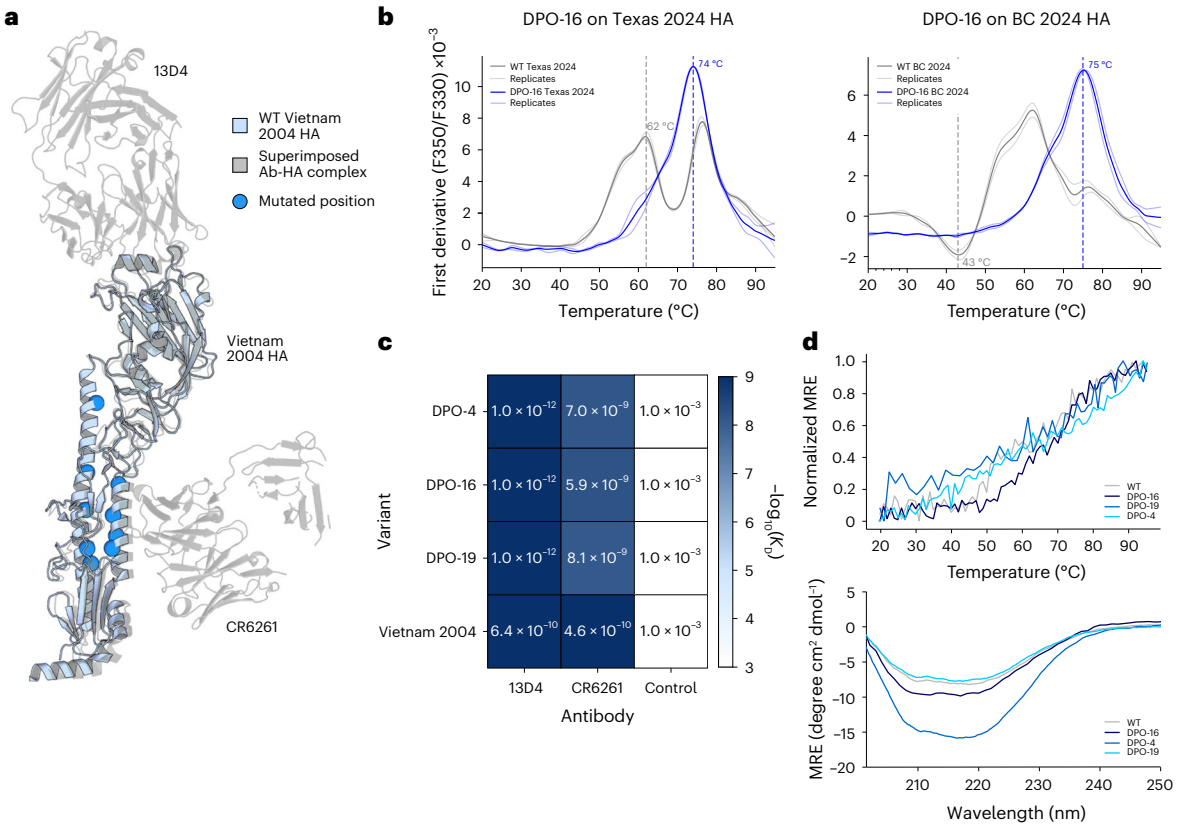


Fig. 5 | Variants generalize to recent variants of concern and retain native antigenicity. a, Structure of the Vietnam 2004 HA monomer (PDB: 6CFG, pale blue) along with mutation positions (bright blue) aligned and superimposed onto HA-bound structures of fragment antigen binding (Fab) domains of broadly neutralizing antibodies (bnAbs) CR6261 (PDB: 3GBN) and 13D4 (PDB: 6A0Z), which target the stem and head regions respectively. **b**, First derivative of the fluorescence ratio with respect to increasing temperature determined by differential scanning fluorimetry (DSF) of native Texas 2024 and British Columbia (BC) 2024 H5 HA as well as DPO-16-mutated variants. **c**, Heat map displaying antibody binding affinity in K_D of the native Vietnam 2004 H5N1 and ProteinDPO variants to bnAbs 13D4 and CR6261, with an anti-SARS-CoV-2 RBD antibody AZD3152 as a non-binding control. **d**, CD spectra showing mean residue ellipticity (MRE) at varying wavelengths of native Vietnam 2004 HA and ProteinDPO-identified variants at 20 °C (bottom) and a melting profile from 20 °C to 95 °C showing normalized MRE at 217 nm (top).

To determine whether ProteinDPO-identified mutations maintain key epitopes recognized by existing antibodies, we measured the binding affinity of HA variants with nine substitutions to anti-HA broadly neutralizing antibodies (bnAbs) targeting either the head or the stem domains of HA via biolayer interferometry. Strikingly, despite several mutations being located within or adjacent to key antibody-binding epitopes (Fig. 5a), we observed strong nanomolar or subnanomolar affinities for all antibody–antigen pairs (Fig. 5c). In addition, circular dichroism (CD) spectra at 20 °C and a melting profile up to 95 °C confirmed that variants preserved the secondary structure character of the native HA (Fig. 5d). Differences in absorbance signal magnitude are due to concentration differences between samples (Methods).

Given the recent emergence of H5N1 variants with pandemic potential, we then introduced the same ProteinDPO-identified mutations to the A/Texas/dairycattle/Texas/24009367011/2024 variant (GISAID EPI_ISL_19094620; referred to as Texas 2024) derived from dairy cattle and the A/British_Columbia/PHL-2032/2024 (GISAID EPI_ISL_19548836; referred to as BC 2024) human variant, which are two distinct genotypes (B.3.13 and D1.1, respectively) of the H5N1 clade 2.3.4.4b virus that began circulating in 2024[44]. We again saw substantial improvements in T_m despite 20 years of viral evolution: DPO-16 achieves 13 °C and 32 °C improvements in T_m compared with native HA from the Texas and British Columbia strains, respectively (Fig. 5b). The ability to stabilize a desirable conformation of a protein, in this case the prefusion state of HA, while retaining native antigenicity and generalizing to evolved variants, are broadly useful components in the development of robust vaccine candidates that can withstand the rapid evolution characteristic of influenza viruses.

Discussion

We have shown that DPO can effectively imbue an unsupervised structure-conditioned language model with biophysical information. The resulting model, ProteinDPO, outperforms the original pretrained model and a model further specialized via SFT in its ability to score mutations for stability and to generate stable protein sequences.

Strikingly, we see that the information learned from training ProteinDPO only on the relative changes in stability of small protein monomers generalizes to larger, unseen folds. This includes scoring the binding affinity of large protein–protein complexes, thermal melting points of multichain antibodies and single-step stabilization of proteins an order of magnitude larger than the size of training data examples. By contrast, SFT, which does improve scoring within the training data, achieves poor generalization to these tasks, suggesting that SFT merely overfits to the data in the training distribution. Moreover, in silico training of ProteinDPO on small, monomeric proteins has generalized toward an in vitro stabilization campaign of a large, trimeric, viral immunogen. In total, our results suggest that ProteinDPO has not 'forgotten' the general information learned during unsupervised pretraining on diverse proteins and instead learned generalizable rules of protein stability from its experimental biophysical alignment dataset.

The HA H5 stabilization campaign with ProteinDPO demonstrates the practical utility of this approach. We observed that 80% of

ProteinDPO-identified variants result in either increased or maintained stability relative to the native protein. Testing only 45 total designs, we achieved up to 17 °C improvements in melting temperature (T_m). Notably, heavily mutated variants retained binding affinity to neutralizing antibodies despite several mutations being located within or adjacent to the antibody binding epitope (Fig. 5a). When applied to genetically diverse H5N1 strains which emerged during the development of this work, a nine-substitution ProteinDPO variant (DPO-16) achieved improvements in T_m of 13 °C and 32 °C. This suggests that ProteinDPO stabilization occurs via fundamental structural principles rather than strain-specific modifications. This capability enables rapid development of broadly stable vaccine candidates that can withstand the antigenic evolution characteristic of influenza viruses, as demonstrated by mutations identified from a 2004 strain maintaining effectiveness across 20 years of viral evolution on strains from 2024.

Although the model is competitive with supervised, task-specific regression models such as ThermoMPNN[13], it has not clearly surpassed all supervised models in scoring[25] (Fig. 2b,d,e). However, using RL, unlike supervised regression, preserves essential generative capabilities. This enables ProteinDPO to both score sequences (such as regression models) and generate entirely new sequences aligned with desirable functions, which are capabilities that supervised models cannot combine. This generative capability further extends to the prediction of binding affinity and thermal melting across diverse wild-type proteins, multichain antibodies and protein complexes more accurately than the unaligned ESM-IF1 model (Fig. 3b–d) and the sampling of experimentally validated nine-substitution variants (Figs. 4 and 5). Furthermore, we find that alignment based on one specific property (such as the stability of small monomers) improves performance on these diverse tasks (such as binding affinity of large protein–protein complexes), most likely because both properties share common underlying physical principles.

Future directions of this work could explore whether alignment based on other properties could likewise enable generalizable information transfer. In addition, while our experimental efforts aimed to stabilize the prefusion state of HA, explicitly discouraging the postfusion state was not considered in variant generation but could be considered in future versions of our scoring procedure[45]. Moreover, while prefusion stabilization and retention of binding to native antibodies are key components of vaccine design, further evidence would be needed to demonstrate antigenic properties required for vaccine efficacy. Although ProteinDPO incorporates stability information into a structure-conditioned language model, this approach generalizes to any property, any generative model and any biological data modality. To this end, either for rapid protein stabilization or as a framework for alignment of biological generative models, we make ProteinDPO open-source and publicly available at https://github.com/evo-design/protein-dpo.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-026-03137-3.

References

1. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
2. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
3. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
4. Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
5. Hsu, C. et al. Learning inverse folding from millions of predicted structures. In *Proc. 39th International Conference on Machine Learning* (eds Chaudhuri, K. et al.) Vol. 162, 8946–8970 (PMLR, 2022).
6. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
7. Shuai, R. W., Ruffolo, J. A. & Gray, J. J. Iglm: infilling language modeling for antibody sequence design. *Cell Syst.* **14**, 979–989. e4 (2023).
8. Ingraham, J., Garg, V., Barzilay, R. & Jaakkola, T. in *Advances in Neural Information Processing Systems* Vol. 32 (eds Wallach, H. et al.) (NeurIPS, 2019).
9. Dauparas, J. et al. Robust deep learning-based protein sequence design using proteinmpnn. *Science* **378**, 49–56 (2022).
10. Shanker, V. R., Bruun, T. U. J., Hie, B. L. & Kim, P. S. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science* **385**, 46–53 (2024).
11. Sumida, K. H. et al. Improving protein expression, stability, and function with proteinmpnn. *J. Am. Chem. Soc.* **146**, 2054–2061 (2024).
12. Doerr, A. Widening the protein crystallization bottleneck. *Nat. Methods* **3**, 961–961 (2006).
13. Dieckhaus, H., Brocidiacono, M., Randolph, N. Z. & Kuhlman, B. Transfer learning to leverage larger datasets for improved prediction of protein stability changes. *Proc. Natl Acad. Sci. USA* **121**, e2314853121 (2024).
14. Nijkamp, E., Ruffolo, J. A., Weinstein, E. N., Naik, N. & Madani, A. Progen2: Exploring the boundaries of protein language models. *Cell Systems* **14**, 968–978.e3 (2023).
15. Schmirler, R., Heinzinger, M. and Rost, B. Fine-tuning protein language models boosts predictions across diverse tasks. *bioRxiv*, (2023).
16. Rafailov, R. et al. Direct preference optimization: Your language model is secretly a reward model. In *Thirty-seventh Conference on Neural Information Processing Systems*, (2023).
17. Shortle, D. One sequence plus one mutation equals two folds. *Proc. Natl Acad. Sci. USA* **106**, 21011–21012 (2009).
18. Tsuboyama, K. et al. Mega-scale experimental analysis of protein folding stability in biology and design. *Nature* **620**, 434–444 (2023).
19. Goodman, J. L. et al. Prepare now for a potential H5N1 pandemic. *Science* **387**, 1047–1047 (2025).
20. Milder, F. J. et al. Universal stabilization of the influenza hemagglutinin by structure-based redesign of the ph switch regions. *Proc. Natl Acad. Sci. USA* **119**, e2115379119 (2022).
21. Yassine, H. M. et al. Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nat. Medicine* **21**, 1065–1070 (2015).
22. Ouyang, L. et al. in *Advances in Neural Information Processing Systems* Vol. 35 (Koyejo, S. et al.) 27730–27744 (Curran Associates, 2022).
23. Van Kempen, M. et al. Fast and accurate protein structure search with foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
24. Gromiha, M. & Selvaraj, S. Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol.* **86**, 235–277 (2004).
25. Cuturello, F., Celoria, M., Ansuini, A. & Cazzaniga, A. Enhancing predictions of protein stability changes induced by single mutations using MSA-based language models. *Bioinformatics* **40**, btae447 (2024).
26. Pancotti, C. et al. Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. *Brief. Bioinform.* **23**, bbab555 (2022).
27. Stourac, J. et al. FireProtDB: database of manually curated protein stability data. *Nucleic Acids Re.* **49**, D319–D324 (2020).

28. Jankauskaitė, J., Jiménez-García, B., Dapkūnas, J., Fernández-Recio, J. & Moal, I. H. SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics* **35**, 462–469 (2018).
29. Sirin, S., Apgar, J. R., Bennett, E. M. & Keating, A. E. Ab-<bind: antibody binding mutational database for computational affinity predictions. *Protein Sci.* **25**, 393–409 (2016).
30. Bailly, M. et al. Predicting antibody developability profiles through early stage discovery screening. *mAbs* **12**, 1743053 (2020).
31. Rollins, Z. A., Widatalla, T., Waight, A., Cheng, A. C. & Metwally, E. Global Ablef: antibody language ensemble fusion for thermodynamically empowered property predictions. *Bioinformatics* **40**, btae268 (2024).
32. Widatalla, T., Rollins, Z., Chen, M.T., Waight, A. & Cheng, A.C. Abprop: language and graph deep learning for antibody property prediction. In *The 2023 ICML Workshop on Computational Biology* (2023).
33. Shehata, L. et al. Affinity maturation enhances antibody specificity but compromises conformational stability. *Cell Rep.* **28**, 3300-3308.e4 (2019).
34. Jain, T. et al. Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl Acad. Sci. USA* **114**, 944–949 (2017).
35. Hynes, T. R. & Fox, R. O. The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins Struct. Funct. Bioinform.* **10**, 92–105 (1991).
36. Davies, G. J. et al. Structure of the ADP complex of the 3-phosphoglycerate kinase from *Bacillus stearothermophilus* at 1.65 Å. *Acta Crystallogr. D* **50**, 202–209 (1994).
37. Suzuki, H. et al. Crystal structure of a claudin provides insight into the architecture of tight junctions. *Science* **344**, 304–307 (2014).
38. Frenz, B. et al. Prediction of protein mutational free energy: benchmark and sampling improvements increase classification accuracy. *Front. Bioeng. Biotechnol.* **8**, 558247 (2020).
39. Pacesa, M. et al. One-shot design of functional protein binders with BindCraft. *Nature* **646**, 483–492 (2025).
40. Baryshev, A. et al. Massively parallel measurement of protein-protein interactions by sequencing using mp3-seq. *Nat. Chem. Biol.* **20**, 1514–1523 (2024).
41. Pelliccia, M. et al. Additives for vaccine storage to improve thermal stability of adenoviruses from hours to months. *Nat. Commun.* **7**, 13520 (2016).
42. Byrne, P. O. & McLellan, J. S. Principles and practical applications of structure-based vaccine design. *Curr. Opin. Immunol.* **77**, 102209 (2022).
43. Hsieh, C.-L. et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).
44. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: Gisaid's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).
45. Cavanagh, P. E. et al. Computational design of conformation-biasing mutations to alter protein functions. *Science* **391**, eadv7953 (2026).
46. Anishchenko, I. et al. De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
47. Rocklin, G. J. et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).
48. Knudsen, M. & Wiuf, C. The cath database. *Hum. Genom.* **4**, 207 (2010).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

Methods

Generative model alignment

Generative AI applications are now routinely fine-tuned with feedback-based RL during post-training, which has been crucial for the success of applications in the natural language domain[49,50]. Here, we provide a brief overview of the classical model training pipeline, which consists of three stages: SFT, feedback collection and RL fine-tuning[51,52].

SFT. Here, we define SFT as continuing the original pretraining objective on high-quality samples from a downstream labeled dataset. In this work for example, for SFT on the 'Megascale' dataset, high-quality samples are sequences which achieve superior stability to the native sequence. During the SFT stage we use a dataset of input-output pairs, $\mathcal{D} = \{\mathbf{x}, \mathbf{y}\}_{i=1}^{N}$ and train a model p_θ with next-token maximum likelihood loss on high-quality pairs, where y_i is the ith token of the output $\mathbf{y}$

$$\mathcal{L}_{p_\theta} = -\mathbb{E}_{\mathbf{x},\mathbf{y}\sim\mathcal{D}}\left[\sum_{i=1}^{|\mathbf{y}|}\log p_\theta(y_i|\mathbf{x}, \mathbf{y}_{<i})\right]. \tag{1}$$

Feedback stage. During the feedback stage we use the base model or the SFT model p_{SFT} and a set of inputs to sample a number $K > 0$ of candidate completions $\mathbf{y}^{(i)}, \ldots, \mathbf{y}^{(K)}$ per prompt $\mathbf{x}$ from a dataset of preferences $\mathcal{D}_{\text{pref}}$. Feedback can be provided in several ways:

(1) Prior RLHF work has used ranking type of comparisons in which the candidate outputs are ranked $\mathbf{y}^{(i_1)} \succ \mathbf{y}^{(i_2)} \succ \ldots \succ \mathbf{y}^{(i_K)}$, with pairwise rankings being a popular choice. To use this data from the RL stage, we need to extract an evaluation score from it. We usually assume a parameterization of the preference distribution, such as the Plackett-Luce[53,54] choice model (or the Bradley–Terry[55] for pairwise data). In the most general case, given the permutation $\tau : \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(K)} \to \mathbf{y}^{(i_1)}, \ldots, \mathbf{y}^{(i_K)}$, we have

$$P(\tau|\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(K)}, \mathbf{x}) = \prod_{k=1}^{K}\frac{\exp(r(\mathbf{x}, \mathbf{y}^{(i_k)}))}{\sum_{j=k}^{K}\exp(r(\mathbf{x}, \mathbf{y}^{(j)}))}. \tag{2}$$

We can then form ulate this as a maximum likelihood loss over the reward model

$$\mathcal{L}_r = -\mathbb{E}_{\mathbf{x},\mathbf{y}^{(i_1)},\ldots,\mathbf{y}^{(i_K)}\sim\mathcal{D}_{\text{pref}}}\sum_{k=1}^{K}\left[r\left(\mathbf{x}, \mathbf{y}^{(i_k)}\right) - \log\sum_{j=k}^{K}\exp\left(r\left(\mathbf{x}, \mathbf{y}^{(j)}\right)\right)\right]. \tag{3}$$

(2) In the case of pairwise rankings $\mathbf{y}^w \succ \mathbf{y}^l$, this just reduces to the standard binary logistic regression loss,

$$\mathcal{L}_r = -\mathbb{E}_{\mathbf{x},\mathbf{y}^w,\mathbf{y}^l\sim\mathcal{D}_{\text{pref}}}\left[\log\sigma(r(\mathbf{x}, \mathbf{y}^w) - r(\mathbf{x}, \mathbf{y}^l))\right], \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid function. Notice that if we are provided with a complete ranking over K candidates then we can construct $\binom{K}{2}$ pairwise rankings, which was used in prior works[22]. These can still be easily optimized together since we still require a single forward pass through the reward model.

(3) Alternatively a numerical score, r_i (such as used in a Likert scale) could be provided per candidate input-output pair $(\mathbf{x}, \mathbf{y}^{(i)})$.

RL stage. During the RL stage we use a reference model p_{ref}, sometimes initialized as p_{SFT} from the SFT stage, and optimize the following RL problem:

$$\max_{p_\theta}\ \underbrace{\mathbb{E}_{\mathbf{x}\sim\mathcal{D},\mathbf{y}\sim p_\theta(\mathbf{y}|\mathbf{x})}[r(\mathbf{x}, \mathbf{y})]}_{\text{maximize score of generations}} - \underbrace{\beta\mathbb{D}_{\text{KL}}[p_\theta(\mathbf{y}|\mathbf{x})||p_{\text{ref}}(\mathbf{y}|\mathbf{x})]}_{\text{distributional penalty}}. \tag{5}$$

We want to optimize a model p_θ to maximize the score, while also maintaining a distribution close to the reference model. The second terms aims to keep the model close to the distribution, which was used to collect the feedback data and prevent out-of-distribution issues[56].

Optimization algorithms

Two main algorithms have emerged for the optimization of the objective in equation (5): (1) classical RL algorithms, such as PPO[57], which use on-policy rollouts from the model; and (2) direct alignment algorithms, such as DPO[16], which train fully-offline using only feedback data. Although successful in large scale-systems, such as GPT-4, the PPO pipeline requires a very accurate score model that can be queried in a continuous manner. Prior work has directly trained a separate score network $r_\theta(\mathbf{x}, \mathbf{y})$ using the feedback data with the objective in equation (4) and directly applied PPO to the problem in equation (5) using the trained score model. However, such a pipeline can be quite complex, unstable and computationally expensive[58,59]. DPO[16] is an alternative to this algorithmic pipeline, which we present below.

DPO

The DPO algorithm is based on the observation that the objective in equation (5) has a closed form solution of the type

$$p^*(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})}p_{\text{ref}}(\mathbf{y}|\mathbf{x})\mathrm{e}^{\frac{1}{\beta}r(\mathbf{x},\mathbf{y})}, \tag{6}$$

where

$$Z(\mathbf{x}) = \sum_{\mathbf{y}}p_{\text{ref}}(\mathbf{y}|\mathbf{x})\mathrm{e}^{\frac{1}{\beta}r(\mathbf{x},\mathbf{y})} \tag{7}$$

is an intractable partition function. The key insight of the DPO algorithm is the model-reward equivalency, that is,

$$r(\mathbf{x}, \mathbf{y}) = \beta\log\frac{p^*(\mathbf{y}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}|\mathbf{x})} + \beta\log Z(\mathbf{x}). \tag{8}$$

Now, if instead of training a score model r_θ, we directly parameterize the generative model p_θ, we can use the preference data to directly optimize p_θ, essentially marginalizing the RL loop.

For the ranked objective, if we substitute the above expression into the Plackett–Luce reward model in equation (3), we obtain the following objective:

$$\begin{aligned}\mathcal{L}_{p_\theta} = -\mathbb{E}_{\mathbf{x},\mathbf{y}^{(i_1)},\ldots,\mathbf{y}^{(i_K)}\sim\mathcal{D}_{\text{pref}}}\sum_{k=1}^{K}\Bigg[\beta\log\frac{p_\theta(\mathbf{y}^{i_k}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{i_k}|\mathbf{x})} \\ - \log\sum_{j=k}^{K}\exp\left(\beta\log\frac{p_\theta(\mathbf{y}^{(i\,j)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(i\,j)}|\mathbf{x})}\right)\Bigg].\end{aligned} \tag{9}$$

For the paired objective, If we instead use the pairwise objective in equation (4), we obtain the canonical form of the DPO objective

$$\mathcal{L}_r = -\mathbb{E}_{\mathbf{x},\mathbf{y}^w,\mathbf{y}^l\sim\mathcal{D}_{\text{pref}}}\log\sigma\left(\beta\log\frac{p_\theta(\mathbf{y}^w|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^w|\mathbf{x})} - \beta\log\frac{p_\theta(\mathbf{y}^l|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^l|\mathbf{x})}\right). \tag{10}$$

Here, due to the homogeneity of both the Plackett–Luce and Bradley–Terry model (that is, comparing several options for the same context), the untractable partition term $Z(\mathbf{x})$ cancels out and we are left with tractable, fully offline objectives.

Weighted DPO algorithm

Prior methods, including DPO[16], have all focused on extracting a per-sample score $r(\mathbf{x}, \mathbf{y})$ from rankings. However, if we have some external evaluation score, that is, data $\{\mathbf{x}, \mathbf{y}_1, r(\mathbf{x}, \mathbf{y}_1)\}_{i=1}^{K}$, then this step is redundant. Given numerical scores, we could train a score model, using a regression-based framework and then use the standard PPO pipeline,

but this approach inherits all the shortcomings of the standard RLHF pipeline. We could use the numerical scores to induce a ranking over candidates $\mathbf{y}_i$ and the apply the Plackett–Luce or Bradley–Terry DPO objectives in equations (9) and (10), respectively; however that faces two drawbacks:

1. Plackett–Luce and Bradley–Terry both assume a particular parametric form of the ranking likelihood, that is, the rankings are inherently probabilistic, which would induce additional noise in the optimization procedure if we generate rankings based on those choice models.
2. We can also generate rankings based on raw scores similarly to prior works[60]. However, note that any monotone transformation would still induce the same relative rankings. That is, there might be actual relative information in the score which could be lost in such a scenario.

Weighted regression-based methods[61–63] are a good fit for this problem; however, they have been suboptimal in the large generative model setting and do not yield significant improvement[64]. Instead, we will aim to develop a DPO type of algorithm that can efficiently utilize the numerical scores of the candidate sequences. Given an $\mathbf{x}$ and candidate sequences $\{\mathbf{y}^{(i)}, r(\mathbf{x}, \mathbf{y}^{(i)})\}_{i=1}^{K}$, consider the parameterized reward

$$r_\theta(\mathbf{x}, \mathbf{y}^{(i)}) = \beta \log \frac{p_\theta(\mathbf{y}^{(i)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(i)}|\mathbf{x})} + \beta \log Z_\theta(\mathbf{x}). \tag{11}$$

That is a parameterized distribution $p_{\theta(\mathbf{y}|\mathbf{x})}$ induces a reward function $r_\theta(\mathbf{x}, \mathbf{y})$, such that the model p_θ is the optimal solution for the corresponding RL problem in equation (5). Now, instead of aligning the generative model p_θ with some sampled rankings or preferences, we will instead minimize a distance function of the type $D(r(\mathbf{x}, \mathbf{y}), r_\theta(\mathbf{x}, \mathbf{y}))$; that is, we will match the true score samples with the implicit ones. A remaining issue is that we still have the constraining partition function $Z_\theta(\mathbf{x})$ (because it depends on p_θ), which we would like to eliminate. When we consider

$$w_\theta^{(i)} = \frac{\exp\left(r_\theta\left(\mathbf{x}, \mathbf{y}^{(i)}\right)\right)}{\sum_i \exp\left(r_\theta\left(\mathbf{x}, \mathbf{y}^{(i)}\right)\right)} = \frac{\exp\left(\beta \log \frac{p_\theta(\mathbf{y}^{(i)}|\mathbf{x})}{p_{ref}(\mathbf{y}^{(i)}|\mathbf{x})}\right)}{\sum_i \exp\left(\beta \log \frac{p_\theta(\mathbf{y}^{(i)}|\mathbf{x})}{p_{ref}(\mathbf{y}^{(i)}|\mathbf{x})}\right)}, \tag{12}$$

notice that the partition function $Z_\theta(\mathbf{x})$ was eliminated. In the same manner we have that

$$w^{(i)} = \frac{\exp\left(r(\mathbf{x}, \mathbf{y}^{(i)})\right)}{\sum_i \exp\left(r(\mathbf{x}, \mathbf{y}^{(i)})\right)}. \tag{13}$$

Notice now that $\mathbf{w} = (w^{(1)}, \ldots, w^{(K)})$ and $\mathbf{w}_\theta = (w_\theta^{(1)}, \ldots, w_\theta^{(K)})$ are both valid probability distributions. In fact, the probability distributions resemble the Boltzmann distribution from statistical mechanics[65]

$$p(i) = \frac{\exp\left(-\frac{E_i}{k_B T}\right)}{\sum_i \exp\left(-\frac{E_i}{k_B T}\right)}, \tag{14}$$

where $p(i)$ is the probability of the system being in a state i with energy E_i, and $k_B T$ is the scaled absolute temperature of the system. Substituting $r(\mathbf{x}, \mathbf{y}^{(i)})$ for the negative energy of the state $-E_i$, and considering $k_B T$, a parameter which allows us to scale the numerical scores $r(\mathbf{x}, \mathbf{y}^{(i)})$, the weighted DPO objective is simply minimization of the Kullback–Leibler (KL) divergence $\mathbb{D}_{KL}[\mathbf{w}|\mathbf{w}_\theta]$ between the Boltzmann-like distributions parameterized by the generative model p_θ and numerical scores from the dataset $\mathcal{D}$. With some simple algebra, this yields:

$$\mathcal{L}_{p_\theta} = -\mathbb{E}_{\mathcal{D}}\left[\sum_i w^{(i)} \beta \log \frac{p_\theta(\mathbf{y}^{(i)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(i)}|\mathbf{x})} - \log \sum_i \exp\left(\beta \log \frac{p_\theta(\mathbf{y}^{(i)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(i)}|\mathbf{x})}\right)\right]. \tag{15}$$

In the case where we have only two candidate sequences $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$, this can be simplified to

$$\mathcal{L}_{p_\theta} = -\mathbb{E}_{\mathcal{D}}\left[w^{(i)} \log \sigma\left(\beta \log \frac{p_\theta(\mathbf{y}^{(1)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(1)}|\mathbf{x})} - \beta \log \frac{p_\theta(\mathbf{y}^{(2)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(2)}|\mathbf{x})}\right) + \right. \tag{16}$$

$$\left. w^{(2)} \log \sigma\left(\beta \log \frac{p_\theta(\mathbf{y}^{(2)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(2)}|\mathbf{x})} - \beta \log \frac{p_\theta(\mathbf{y}^{(1)}|\mathbf{x})}{p_{\text{ref}}(\mathbf{y}^{(1)}|\mathbf{x})}\right)\right]. \tag{17}$$

That is, this is essentially a weighted version of the DPO algorithm. One interpretation is that this objective is the expected DPO objective under $p(\mathbf{y}^{(1)} \succ \mathbf{y}^{(2)}|\mathbf{x}) = w^{(1)}$ and $p(\mathbf{y}^{(2)} \succ \mathbf{y}^{(1)}|\mathbf{x}) = w^{(2)}$. These objectives are theoretically optimal in the offline regime as they are both mode-seeking and use a 'negative gradient' formulation[64].

Implementation.

Algorithm 1 Compute sequence log-likelihood

Input: sequence $\mathbf{y}$, encoded structure $\mathbf{h}$
Output: log-likelihood $\log p_\theta(\mathbf{y}|\mathbf{h})$
// Get logits via teacher forcing
$\mathbf{O} \leftarrow \text{TransformerDecoder}(\mathbf{y}_{[0:|\mathbf{y}|-1]}, \mathbf{h})$ // Logits for next-token prediction
// Convert logits to probabilities
$\mathbf{P} \leftarrow \text{Softmax}(\mathbf{O})$ // Token probabilities over vocabulary
// Compute log-likelihood as negative cross-entropy
$\mathbf{l} \leftarrow -\text{CrossEntropy}(\mathbf{P}, \mathbf{y}_{[1:|\mathbf{y}|]})$ // Per-position negative loss
$\log p_\theta(\mathbf{y}|\mathbf{h}) \leftarrow \sum_t \ell_t$ // Sum over sequence positions
return $\log p_\theta(\mathbf{y}|\mathbf{h})$

Algorithm 2 DPO for structure-conditioned protein language model (pairwise)

Input: structure $\mathbf{X} \in \mathbb{R}^{L\times 3}$, preferred sequence $\mathbf{y}_w$, dispreferred sequence $\mathbf{y}_l$, reference log-likelihoods $\log p_{\text{ref}}(\mathbf{y}_w|\mathbf{X})$ and $\log p_{\text{ref}}(\mathbf{y}_l|\mathbf{X})$, KL penalty coefficient β
Output: DPO loss $\mathcal{L}_{\text{DPO}}$
// Encode structure once for both sequences
$\mathbf{h} \leftarrow \text{GVPEncoder}(\mathbf{X})$
// Compute policy log-likelihoods for both sequences
$\mathbf{O}_w, \mathbf{O}_l \leftarrow \text{TransformerDecoder}(\mathbf{y}_w, \mathbf{h}), \text{TransformerDecoder}(\mathbf{y}_l, \mathbf{h})$ // Logits
$\log p_\theta(\mathbf{y}_w|\mathbf{X}), \log p_\theta(\mathbf{y}_l|\mathbf{X}) \leftarrow \text{ComputeLogLikelihood}(\mathbf{y}_w, \mathbf{O}_w), \text{ComputeLogLikelihood}(\mathbf{y}_l, \mathbf{O}_l)$
// Compute pairwise DPO loss
$r_\theta \leftarrow \log p_\theta(\mathbf{y}_w|\mathbf{X}) - \log p_\theta(\mathbf{y}_l|\mathbf{X})$ // Policy log-ratio
$r_{\text{ref}} \leftarrow \log p_{\text{ref}}(\mathbf{y}_w|\mathbf{X}) - \log p_{\text{ref}}(\mathbf{y}_l|\mathbf{X})$ // Reference log-ratio
$\mathcal{L}_{\text{DPO}} \leftarrow -\log\sigma(\beta(r_\theta - r_{\text{ref}}))$
return $\mathcal{L}_{\text{DPO}}$

Algorithm 3 DPO for structure-conditioned protein language model (ranked)

Input: structure $\mathbf{X} \in \mathbb{R}^{L\times 3}$, ranked sequences $\{\mathbf{y}_1, \ldots, \mathbf{y}_K\}$ where $\mathbf{y}_1 \succ \mathbf{y}_2 \succ \ldots \succ \mathbf{y}_K$, reference log-likelihoods $\{\log p_{\text{ref}}(\mathbf{y}_k|\mathbf{X})\}_{k=1}^{K}$, KL penalty coefficient β
Output: DPO loss $\mathcal{L}_{\text{DPO}}$
// Encode structure once for all sequences
$\mathbf{h} \leftarrow \text{GVPEncoder}(\mathbf{X})$
// Compute policy log-likelihoods for all sequences
for $k = 1$ to K do
  $\mathbf{O}_k \leftarrow \text{TransformerDecoder}(\mathbf{y}_k, \mathbf{h})$ // Logits
  $\log p_\theta(\mathbf{y}_k|\mathbf{X}) \leftarrow \text{ComputeLogLikelihood}(\mathbf{y}_k, \mathbf{O}_k)$
end for
// Plackett–Luce ranking loss
$\mathcal{L}_{\text{DPO}} \leftarrow 0$
for $k = 1$ to $K-1$ **do**
  $r_k \leftarrow \beta(\log p_\theta(\mathbf{y}_k|\mathbf{X}) - \log p_{\text{ref}}(\mathbf{y}_k|\mathbf{X}))$ // Reward for rank k

$s_k \leftarrow \log \sum_{j=k+1}^{K} \exp\left(\beta(\log p_\theta(\mathbf{y}_j|\mathbf{X}) - \log p_{\text{ref}}(\mathbf{y}_j|\mathbf{X}))\right)$
$\mathcal{L}_{\text{DPO}} \leftarrow \mathcal{L}_{\text{DPO}} - \log\sigma(r_k - s_k)$
end for
return $\mathcal{L}_{\text{DPO}}$

Algorithm 4 DPO for structure-conditioned protein language model (weighted)

Input: structure $\mathbf{X} \in \mathbb{R}^{L\times 3}$, sequences $\{\mathbf{y}_k\}_{k=1}^{K}$, reference log-likelihoods $\{\log p_{\text{ref}}(\mathbf{y}_k|\mathbf{X})\}_{k=1}^{K}$, preference targets $\mathbf{w} \in \mathbb{R}^K$, KL penalty coefficient β, temperature T
Output: DPO loss $\mathcal{L}_{\text{DPO}}$
// Encode structure once for all sequences
$\mathbf{h} \leftarrow$ GVPEncoder($\mathbf{X}$)
// Compute policy log-likelihoods for all sequences
for $k = 1$ to K **do**
  $\mathbf{O}_k \leftarrow$ TransformerDecoder($\mathbf{y}_k$, $\mathbf{h}$) // Logits
  $\log p_\theta(\mathbf{y}_k|\mathbf{X}) \leftarrow$ ComputeLogLikelihood($\mathbf{y}_k$, $\mathbf{O}_k$)
end for
// Compute reward-weighted cross-entropy loss
$\mathbf{r} \leftarrow \beta[\log p_\theta(\mathbf{y}_1|\mathbf{X}) - \log p_{\text{ref}}(\mathbf{y}_1|\mathbf{X}), \ldots, \log p_\theta(\mathbf{y}_K|\mathbf{X}) - \log p_{\text{ref}}(\mathbf{y}_K|\mathbf{X})]^\top$
$\mathcal{L}_{\text{DPO}} \leftarrow -\sum_{k=1}^{K} w_k \log \frac{\exp(r_k/T)}{\sum_{j=1}^{K}\exp(r_j/T)}$ // Cross-entropy with temperature scaling
return $\mathcal{L}_{\text{DPO}}$

Datasets

Megascale. The Megascale dataset (v2 4/20/2023) served as the experimentally measured stability dataset from which we aligned the pretrained ESM-IF1. After filtering out data marked as unreliable for machine learning by Megascale authors, the dataset contained approximately 660,000 variants for 405 protein domains. The variants consisted of insertions, deletions, single and double substitutions. Similar to prior methods[13], insertions and deletions were not considered as DPO training was designed to compare stabilities across the same structure; however, due to the robust nature of the model we were able to train and model with the double-mutations in the dataset. We additionally removed duplicate sequences. FoldSeek[23] clustering with the 'easy-cluster algorithm was performed on all PDBs remaining in the Megascale dataset with an alignment threshold of 50%. From this, clusters were randomly assigned to respective train (90%), validation (5%) and test (5%) datasets to ensure generalizability of the model across different structures. Because the model is trained with sets of sequences for a common structure, either the ΔG (absolute) or $\Delta\Delta G$ (relative) of sequences were used for training. While these should be identical in theory after accounting for native ΔG, due to noise in the data the two metrics are different. We found negligible difference in loss and validation accuracy training with either metric, so all models were trained on ΔG as more data points had ΔG measurements. For SFT, all variants with $\Delta\Delta G < 0$ were selected and the same splits were used. For using model likelihoods to predict Megascale thermostability, the whole-sequence log-likelihood of the variant (algorithm 1), normalized by subtracting the log-likelihood of the wild-type sequence, was used.

SKEMPIv2. SKEMPIv2 is database of binding free energy changes upon single point mutations within a variety of protein complex interfaces[28]. After removal of PDBs with non-canonical amino acids, the final dataset contained 6,487 variants. To score on this dataset, we followed instructions and code from the ESM-IF1 Github[5], averaging the log-likelihood of simultaneous mutations given the entire complex structure and sequence as input. However we took the additional step of normalizing with the averaged wild-type likelihood of said mutated residues. Because certain complexes were only represented by very few variants, sometimes with a different PDB for each, we did not evaluate the average correlation within each complex but report an 'all versus all' comparison.

AB-Bind. AB-Bind is an antibody binding mutational database curated to train and evaluate computational affinity predictions from Sirin et al.[29]. The database includes 1,101 variants across 32 complexes curated from public data, each variant having experimentally determined changes in binding free energy ($\Delta\Delta G$). Duplicated variants, variants with missing mutated residues in the PDB, and those without an experimentally determined PDB were removed resulting in a final dataset size of 1,011 variants. We evaluated scoring on this dataset the same as with SKEMPIv2[28]; however, we reported the average correlations across each antibody–antigen complex as there were an ample number of variants per complex.

Antibody thermal unfolding. The antibody thermal stability dataset is derived from Widatalla et al.[32], which compiled scores and IgFold[66] predicted structures from the original Jain et al.[34] and Shehata et al.[33] datasets, consisting of thermal stability measurements for monoclonal antibodies (mAbs). These two studies collected several biophysical measurements for approximately 500 human B cell and LLPC derived mAbs[33] and mAbs that have been in clinical trials[34]. Among these includes 483 measurements of the temperature of thermal unfolding (T_m). While the entire fragment antigen binding (Fab) domain was measured in this experiment, only the structure and sequence of the variable domains (Fv) were taken as input to the models. We used no normalization for scoring on this dataset as we considered every sequence as a 'native' protein.

FireProt. The FireProt database is a curated dataset of changes in free energy ($\Delta\Delta G$) for 3,438 single mutations for 100 unique proteins[27]. To alleviate data-leakage from the Megascale training data, we used the Megascale 'Homolog-free' FireProt dataset from Diekhaus et al.[13], which has all homologous sequences to the Megascale dataset removed, resulting in 2,578 variants. For this dataset, probably due to the presence of only single mutations, we found the best performance taking the log-likelihood of only the mutated residue, and normalizing by subtracting the wild-type residue likelihood.

S669. The S669 dataset contains experimentally measured $\Delta\Delta G$ values of 669 single mutations of 94 proteins[26]. Due to the presence of non-canonical amino acids, four variants were removed from this dataset for evaluation of vanilla ESM-IF1 and ProteinDPO. Similar to FireProt[27], we found the best performance taking the log likelihood of the mutated residue normalized by subtracting the wild-type residue likelihood. Results for all other methods on this dataset derived from Cuturello et al. and Diekhaus et al.[13,25].

Evaluation

Model sampling. The sampling evaluation conducted for (Fig. 3d) is described here. In this evaluation, 500 sequences were sampled with the Paired, Vanilla and SFT models using the backbones from PDBs 1STN, 1PHP as input. Sequences were generated from each model across five sampling temperatures [0.01, 0.05, 0.1, 0.5, 0.1] with standard autoregressive sampling and no additional filtering. Sampling was conducted using sampling code from the ESM-IF1 Github repository[5], only differing the the use of model weights between experiments. The exact code used is made available at https://github.com/evo-design/protein-dpo/blob/main/sample.py

Sampling evaluation. Self-consistency template modeling (TM) was calculated between the ESMFold[1] predicted structure of the designed sequence and the experimental PDB structure. pLDDT was computed as the average per-residue predicted error from the ESMFold prediction. Given the generation and original sequence have the same sequence length, sequence similarity is calculated as the number of the residue matches divided by the total number of positions.

Rosetta generation evaluation. The same exact sequences from the above sampling evaluation was then evaluated with the Rosetta cart_ddg protocol[38]. To run this protocol, code from the RosettaDDGPrediction repository was used www.github.com/ELELAB/RosettaDDGPrediction. Below is the command used to run the protocol for 1PHP generations:

```
rosetta_ddg_run
-p /path/to/relaxed/pdb
-cr /path/to/config_run/cartesian2020_ref2015.yaml
-cs /path/to/config_settings/nompi.yaml
-r /home/talal/rosetta.source.release-371/
-d /path/to/out/dir/
-l /path/to/mutations.txt
```

Where the `mutations.txt` file simply includes a comma-separated list of all differing positions from the native sequence, for example, `A.N.2.G,A.K.3.Y,A.R.7.Y,A.V.9.I,A.R.12.S,A.F.17.L,A.C.18.L,A.R.19.F,A.F.22.L,`

`A.N.23.D,A.E.27.V,A.Q.28.D,A.A.30.R,A.T.32.L,A.D.34.P,A.T.35.S,A.R.38.L,`

`A.A.40.Q,A.K.53.R,A.L.56.I,A.A.57.I,A.H.59.Q,A.R.62.D,A.K.64.G,A.K.66`....

Rosetta benchmarking protocol. For benchmarking of REU in antibody thermal stability prediction in (Fig. 3c), all input structures were energy-minimized using the PyRosetta `FastRelax` protocol[67]. The protocol was run with both backbone and sidechain degrees of freedom enabled, while rigid-body movements were disabled. Minimization was performed using the LBFGS-Armijo nonmonotone minimizer with a maximum of 200 iterations. Energy evaluations were performed using the `ref2015` scoring function. Total REU was used and was not normalized by length, as length-normalization decreased performance.

Scoring metrics. All correlation coefficients were calculated using the `scipy` python package, particularly the `spearmanr` and `pearsonr` functions. AUROC scores were calculated using the `scikit-learn` python package, particularly the `roc_auc_score` function. When a correlation is reported as, for example, average Spearman ρ, this means that the correlation is computed for all mutations within each protein domain, then averaged across all domains. Any scoring comparisons between variants of ESM IF1 (vanilla, paired, ranked, weighted and SFT) were made with the exact same code and configuration, with the only difference being model weights. The exact input into the correlation calculations varies per dataset due to the presence of single versus double and multimutations and is noted in the Datasets section. ProteinMPNN[9] was used in the same manner as ESM-IF[5] and ProteinDPO for scoring, also varying per dataset. To run ProteinMPNN, the `score` function in the `ColabDesign` implementation was used, accessible at https://github.com/sokrypton/ColabDesign.

Model training

Input construction. Multiple methods for positive-negative pair selection from the scalar-labeled data were evaluated, including randomly selecting pairs then ordering them by scalar value, prioritizing pairs with small differences and prioritizing pairs with large differences. To have a tunable parameter that controls the distance between pairs for the two latter approaches, we implemented a 'gap level' heuristic, which is a parameter within [0,1]. A gap level of 1 corresponds to placing the largest gap possible in sequence ranking between pairs, whereas a gap level of 0 corresponds to the smallest possible gap. For any gap level choice, we ensure the gaps between sequences in terms of ranking is uniform across all pairs and that all sequences are used, except for a single excluded sequence in the case there is an odd number of sequences for a native protein. Although most choices of gap level had comparable performance, we found a gap level of 0.5 performed the best on validation data. Sets of sequences were randomly chosen for models trained with the ranked objective, and models trained with the weighted objective and set size (K) greater than two. For models trained with the weighted objective we experimented with various values temperature values (T) for scaling the scalar labels, finding 0.1 to be the most favorable.

DPO training details. Noise of 0.1 Å was applied to all PDB structures in line with the training of ESM-IF1[5]. For each set (ranked and weighted models) or pair (paired model) of sequence responses, the same native backbone was used as structural prompt, PDBs of which were derived from Tsuboyama et al.[18]. Hyperparameter tuning was performed with varying values for β, gap level (paired and weighted model), temperature T (weighted model), learning rate, K (weighted and ranked model) and batch size. For computational efficiency, reference model likelihoods were precomputed before training. Models were each trained on a single NVIDIA H100 GPU with 80 GB of memory. All three models (paired, ranked and weighted) share the same learning rate of 1×10^{-7}, β of 0.1, maximum of 30 epochs and identical Adam optimizer settings ($\beta_1=0.9, \beta_2=0.98, \epsilon=1\times10^{-8}$ and decay of 0.1). The paired and weighted models used a batch size of 32 and $K=2$, while the ranked model used a batch size of 8 and $K=3$. A gap level of 0.5 was applied to both the paired and weighted models, and the weighted model additionally used a temperature parameter $T=0.1$. All models were trained until validation loss convergence, up to the maximum number of epochs. Algorithms for training ESM-IF1 with all variants of DPO are reported above (algorithms 2–4).

SFT training details. SFT training largely mirrored DPO training with two distinctions. Fine-tuning used the same noise level, the same input PDBs, was initialized from the same weights and used the same optimizer. The major distinction is fine-tuning was only done on a subset of training and validation examples which had favorable stability, specifically variants with stability superior to the native protein ($\Delta\Delta G<0$) were selected. The second major distinction is loss. For SFT negative log-likelihood (or cross-entropy loss) was minimized. The algorithm for SFT is provided below (algorithm 5). Training and validation loss and sequence recovery curves across a learning rate sweep are provided in Supplementary Information figure 11 as well as performance on a subset of fitness evaluations (Fig. 4). The checkpoint with the lowest validation loss was selected.

Algorithm 5 SFT of ESM-IF model

Input: dataset $\mathcal{D}$, pretrained model θ_0
Output: fine-tuned model parameters θ
// Filter for favorable stability
$\mathcal{D}_{\text{stable}} \leftarrow \{(\mathbf{X}, \mathbf{y}, \Delta\Delta G) \in \mathcal{D} : \Delta\Delta G < 0\}$ // Select variants more stable than native
Initialize $\theta \leftarrow \theta_0$
for all $(\mathbf{X}, \mathbf{y}, \Delta\Delta G) \in \mathcal{D}_{\text{stable}}$ **do**
  // Encode structure
  $\mathbf{h} \leftarrow \text{GVPEncoder}(\mathbf{X})$
  // Get decoder outputs (logits) via teacher forcing
  $\mathbf{O} \leftarrow \text{TransformerDecoder}(\mathbf{y}_{[0:|\mathbf{y}|-1]}, \mathbf{h})$
  // Compute negative log-likelihood (see Algorithm 1)
  $\mathcal{L} \leftarrow -\text{ComputeLogLikelihood}(\mathbf{y}, \mathbf{O})$ // Cross-entropy loss
  $\text{UpdateParameters}(\theta, \nabla_\theta \mathcal{L})$ // One optimization step
end for
return θ

In vitro experiments

Variant selection. To select variants for experimental testing, the sequence likelihood of every possible single-site substitution was evaluated using ProteinDPO using an experimental structure of H5N1 HA (PDB: 6CFG). This was done by individually substituting all 19 alternate amino acids across every position in the native sequence

and scoring the whole-sequence likelihood of each variant using the algorithm defined above 1. Given HA is a homo-trimer, scores for identical chains were averaged. To alleviate structural bias due to experimental conditions during structural characterization of the input structure, this procedure was repeated across two other available structures of H5N1 hemagluttinin (PDBs: 2FK0, 6CF5), and the percentile ranking of mutants within each structure were averaged. The top-*n* variants in highest percentile rank averaged across structures were selected for characterization. We note that the first 20 single-site substitutions experimentally characterized (denoted as DPO-A1-20 in Extended Data Table 3) were selected using the monomeric form of HA, only utilizing PDB 6CFG, while the following 10 single-site substitutions (Extended Data Table 3, DPO-1,2,5,6,8,11,14,17,20,23), as well as all following multisite substitution variants were selected using the trimeric form of the PDB and all three experimental structures.

To select three-site, five-site and nine-site substitution variants in reasonable computational time, we utilized the top single-site substitutions predicted by the model given the trimeric structure as an initial pool to derive combinations from, removing six single-substitutions that were not significantly stabilizing (P160D, P160E, H26W, H26Y, G47W, H111F) in the first round of monomeric-structure-derived variants (Extended Data Table 3, DPO-A1-20). In addition, we excluded chain A of hemagluttinin (head region) as defined by PDB 6CFG from the pool of single-site substitution variants, given ProteinDPO heavily preferred variants in chain B (stem region). Finally, we trimmed down this pool to balance the benefits of coverage with the computational costs of exploring greater combinatorial space. This resulted in an initial pool of 16, 13 and 57 unique single-substitutions to be used as candidates for the three-site, five-site and nine-site substitution variants respectively. We then used the same procedure for scoring single-site substitutions, whole-sequence likelihood, for every possible combination of mutations from the pool for three-site, five-site and nine-site substitution variants. We once again repeated this across all three structures, calculated the ranking of each variant within each structure, and selected the best variants by average percentile rank across the structures for experimental testing. For combinations of three and five, we additionally included a variant consisting of a simple combination of the top-*n* stabilizing single-site substitution variants. This was not possible for combinations of nine, given some of these positions (for example, N95) had not been tested in the initial round of single variants (Extended Data Table 3, DPO-A1-20).

HA site-directed mutagenesis. The HA sequence was cloned into a pADD2 vector between the rBeta-globin intron and β-globin poly(A). HA constructs contain a Foldon trimerization domain, an AviTag and a C-terminal 6× His tag. We used H5N1 HA sequences from the A/Vietnam/1203/2004 for the initial experiments. We used the A/Texas/dairycattle/Texas/24009367011/2024 and A/British_Columbia/PHL-2032/2024 variants for the follow-up experiments, whose amino acid sequences were derived from GISAID[44]. Single amino acid mutations were introduced using mutagenic primers and Gibson assembly. The target plasmid was linearized by restriction digest outside the region of interest. PCR fragments were amplified from the wild-type plasmid using Q5 High-Fidelity DNA Polymerase (NEB) with mutagenic primers containing the desired mutations and Gibson-compatible overlaps. PCR products were gel-purified and assembled with the digested vector using Gibson Assembly Master Mix (NEB). Assembled products were transformed into *Escherichia coli* Stellar competent cells (Takara Bio, 636763) and plated on carbenicillin selection plates. Positive clones were verified by whole-plasmid sequencing before plasmid preparation.

Antibody cloning. Variable regions of antibody heavy and light chain sequences were ordered as eBlock gene fragments from Integrated DNA Technologies and cloned into a linearized CMV/R plasmid backbone by Gibson assembly, for expression under a constitutive CMV promoter as described[10]. Plasmid backbones possessed human IgG1 heavy or light chain constant domains; 13D4[68] was cloned with a kappa light chain, whereas CR6261[69] and AZD3152[70] were cloned with lambda light chains. Primers used for destination plasmid linearization were also synthesized by Integrated DNA Technologies.

DNA preparation. Plasmids were transformed into Stellar competent cells, and transformed cells were plated and grown at 37 °C overnight. Colonies were midi-prepped per the manufacturer's protocols (ZymoPure II Plasmid Midiprep Kit, Zymo Research). Plasmids were sterile filtered using a 0.22-µm syringe filter and stored at 4 °C.

Protein expression. All proteins were expressed in Expi293F cells (Thermo Fisher Scientific, A14527). For variants which where to be further tested for antibody binding, a biotinylation tag (AviTag) were also expressed in the presence of a BirA enzyme, resulting in spontaneous biotinylation during protein expression[10]. Expi293F cells were cultured in media containing 66% FreeStyle 33% Expi media (Thermo Fisher Scientific) and grown in polycarbonate shaking flasks at 37 °C in 8% carbon dioxide. Upon sufficient growth, cells were then diluted and transfected at a density of approximately 3–4 million cells per milliliter and volume of 45 ml. Transfection mixtures were made by adding the following components: 15 µg of midi-prepped DNA, 3 ml of culture media, 300 µl of glucose, 300 µl of valproic acid and FectoPRO (Polyplus) would be added to cells following manufacturer protocols. After mixing and a 10-min incubation, the transfection cocktail would be added to the 45 ml of cells. The cells were collected 3–5 days after transfection by spinning the cultures at 10,000*g* for 20 min. Supernatants were filtered using a 0.45 µm filter.

Protein purification. All variants of HA were His-tagged and purified using HisPur Cobalt resin (Thermo Fisher Scientific, 89964)[10]. A total of 1 ml of of resin was added to cell supernatants, and the samples were then incubated with rotation at 4 °C for at least 1 h. Resin/supernatant mixtures were added to chromatography columns for gravity flow purification. The resin in the column was washed with wash buffer (20 mM imidazole, phosphate buffered saline (PBS) pH 7.4, 150 mM NaCl). The proteins were then eluted with 250 mM imidazole, PBS 7.4, 150 mM NaCl. This was followed by size-exclusion chromatography on an ÄKTA pure system (Cytiva). ÄKTA pure FPLC with a Superdex 200 Increase (S200) gel filtration column was used for purification, for which 0.5 ml of sample was injected and run over the S200, which had been pre-equilibrated in degassed PBS and flash-frozen before storage at −80 °C.

Antibody expression. Antibody expression was similar to protein expression, except cells were transfected with a 1:2 mixture of heavy chain DNA (7.5 µg):light chain DNA (15 µg) in a total expression volume of 30 ml.

Antibody purification. We purified antibodies using a 5-ml HiTrap MabSelect PrismA column on the ÄKTA pure fast protein liquid chromatography (FPLC) instrument (Cytiva). The ÄKTA system was pre-equilibrated with PBS pH 7.4, before the filtered supernatant from Expi293 expression culture was applied to the column directly from a 50 ml Superloop at a flow rate of 2.5 ml min^{-1}. The column was washed with at least five column volumes (CV) of PBS pH 7.4, after which bound IgG was eluted with up to 20 CV 0.1 M sodium citrate pH 3.0. Samples were eluted with into collection tubes containing 1 M Tris–HCl, pH 9.0 alkaline buffer for immediate neutralization of the antibody samples. Between runs, the column was equilibrated with 7 CV PBS pH 7.4, washed with 2 CV 2.5 M NaCl and re-equilibrated into PBS pH 7.4. Using 30-kDa molecular weight cutoff centrifugal concentrators, the eluted samples were concentrated and either used directly in experiments or flash-frozen and stored at −80 °C.

Biolayer interferometry binding experiments. All reactions were run on an Octet RED96 at 30 °C, and samples were run in 1× PBS with 0.1% BSA and 0.05% Tween 20 (Octet buffer)[10]. IgGs were assessed for binding to biotinylated antigens using streptavidin biosensors (Sartorius/ForteBio). Antigen was loaded at a concentration of 200 nM. Tips were then washed and baselined in wells containing only Octet buffer. Samples were then associated in wells containing IgG at 100 nM concentration. Tips without loaded antigen were also associated in wells containing 100 nM IgG and used as a baseline subtraction for data analysis. Association and dissociation binding curves were fit in Octet System Data Analysis Software version 13.0.3.52 using a 1:2 bivalent model for IgGs to determine apparent K_D. The broadly neutralizing antibodies (bnAbs) targeting the head and stem region of HA were 13D4[68] and CR6261[69], respectively. A control antibody used to assess non-specific binding was the anti-SARS-CoV-2 RBD antibody AZD3152[70].

Thermal melting experiments. Thermal melting profiles of HA variants were measured by differential scanning fluorimetry on a Prometheus NT.48 instrument (NanoTemper). A total of 30 µl of each protein was concentrated or dilluted to at least 0.2 mg ml^{-1} in PBS, and triplicate samples were loaded onto Monolith Premium Capillaries (NanoTemper) and then subjected to a temperature gradient from 20 °C to 95 °C at a heating rate of 1 °C min^{-1}. Intrinsic fluorescence (350 nm and 330 nm) was recorded as a function of temperature. Thermal melting curves were plotted using the first derivative of the ratio (350 nm/330 nm). Melting temperatures were calculated automatically by the instrument (PR.ThermControl software, version 2.3.1) and represented positive or negative peaks in the thermal melting curves.

CD experiments. CD spectra were obtained for selected HA variants with a JASCO CD spectrophotometer in a 1-mm-pathlength cuvette (Hellma). Protein samples were diluted or concentrated to be within the range of 0.1 mg ml^{-1} to 0.3 mg ml^{-1} in PBS at 20 °C. Concentration was calculated using a Thermo Scientific NanoDrop OneC Microvolume Spectrophotometer measuring for absorbance at 280 nm. For obtaining a melting profile, absorption signal was monitored at 217 nm from 20 °C to 95 °C in 1 °C increments per minute, with 10 s of equilibration time and 1 s of digital integration time.

Reporting summary

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

Data availability

We used the Megascale[18] dataset for training, either with our own training splits (derived from v2 4/20/2023), those from Diekhaus et al.[13] or those from Cuturello et al.[25], depending on subsequent evaluation. We used the following datasets for evaluation: Homolog-free FireProt[13,27], S669[26], antibody thermal stability[33,34], SKEMPIv2[28] and AB-Bind[29]. Additional details on these datasets are provided in the Methods. Source data are provided with this paper.

Code availability

Code for ProteinDPO is available via GitHub at https://github.com/evo-design/protein-dpo. Weights for the model trained with the paired DPO objective are available via Zenodo at https://doi.org/10.5281/zenodo.11218181 (ref. 71).

References

49. OpenAI. Introducing ChatGPT. *OpenAI* https://openai.com/index/chatgpt/ (2022).
50. Meta AI. Introducing Meta Llama 3: the most capable openly available LLM to date. *Meta AI Blog* https://ai.meta.com/blog/meta-llama-3/ (Meta, 2024).
51. Stiennon, N. et al. Learning to summarize from human feedback. *Adv. Neural Inf. Process. Syst.* **33**, 3008–3021 (2020).
52. Bai, Y. et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. Preprint at https://arxiv.org/abs/2204.05862 (2022).
53. Plackett, R. L. The analysis of permutations. *J. R. Stat. Soc. C* **24**, 193–202 (1975).
54. Luce, R. D. *Individual Choice Behavior: A Theoretical Analysis* (Courier Corporation, 2012).
55. Bradley, R. A. & Terry, M. E. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* **39**, 324–345 (1952).
56. Lambert, N. & Calandra, R. The alignment ceiling: objective mismatch in reinforcement learning from human feedback. Preprint at https://arxiv.org/abs/2311.00168 (2023).
57. Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. Preprint at https://arxiv.org/abs/1707.06347 (2017).
58. Zheng, R. et al. Secrets of RLHF in large language models part I: PPO. Preprint at https://arxiv.org/abs/2307.04964 (2023).
59. Huang, S., Dossa, R.F.J., Raffin, A., Kanervisto, A. and Wang, W. The 37 implementation details of proximal policy optimization. *ICLR Blog Track* https://iclr-blog-track.github.io/2022/03/25/ppo-implementation-details/ (2022).
60. Gao, L., Schulman, J. & Hilton, J. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning* (ICML, 2023).
61. Peng, X. B., Kumar, A., Zhang, G. & Levine, S. Advantage-weighted regression: Simple and scalable off-policy reinforcement learning (OpenReview, 2021); https://openreview.net/forum?id=ToWi1RjuEr8
62. Nair, A., Gupta, A., Dalal, M. & Levine, S. AWAC: accelerating online reinforcement learning with offline datasets. Preprint at https://arxiv.org/abs/2006.09359 (2020).
63. Kostrikov, I., Nair, A. & Levine, S. Offline reinforcement learning with implicit Q-learning. In *Proc. 10th International Conference on Learning Representations* https://openreview.net/forum?id=68n2s9ZJWF8 (ICLR, 2022).
64. Tajwar, F. et al. Preference fine-tuning of LLMs should leverage suboptimal, on-policy data. In *Proc. 41st International Conference on Machine Learning* (eds Salakhutdinov, R. et al.) Vol. 235, 47441–47474 (PMLR, 2024).
65. Noé, F., Olsson, S., Köhler, J. & Wu, H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019).
66. Ruffolo, J. A., Chu, L.-S., Mahajan, S. P. & Gray, J. J. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nat. Commun.* **14**, 2389 (2023).
67. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
68. Zheng, Q. et al. Properties and therapeutic efficacy of broadly reactive chimeric and humanized h5-specific monoclonal antibodies against H5N1 influenza viruses. *Antimicrob. Agents Chemother.* **55**, 1349–1357 (2011).
69. Ekiert, D. C. et al. Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246–251 (2009).
70. Cai, Y. et al. Azd3152 neutralizes sars-cov-2 historical and contemporary variants and is protective in hamsters and well tolerated in adults. *Sci. Transl. Med.* **16**, eado2817 (2024).
71. Widatalla, T., Rafailov, R. & Hie, B. Aligning protein generative models with experimental fitness via direct preference optimization. *Zenodo* https://doi.org/10.5281/zenodo.11218181 (2026).

Acknowledgements
We thank C. Im, R. Shuai, B. Kang and G. Brixi for helpful discussions and J. Sullivan for cluster support. We thank A. Merchant, V. Shanker, D. Pham, N. Sambold, J. Carozza, R. Mardjuki and J. Liu for assistance with experimental characterization of HA variants. T.W. and S.H.K. acknowledge support by the NSF Graduate Research Fellowship. T.W acknowledges support by the Stanford Graduate Fellowship. B.L.H. acknowledges funding support from the Gates Foundation, the Chan Zuckerberg Initiative, Arc Institute, Schmidt Sciences AI2050, Stanford Center for Digital Health and Stanford Institute for Human-Centered Artificial Intelligence (HAI) Hoffman-Yee Research Grants. This research was funded in part by the Advanced Research Projects Agency for Health (ARPA-H) agreement nos. 1AY2AX000054 and D24AC00412-00. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Government.

Author contributions
T.W. and B.L.H conceived the research. T.W. performed dataset curation, model training, generations and evaluation. T.W., A.A.B., S.H.K. and C.L.D. performed cloning, expression, purification and characterization of HA variants. R.R. derived the DPO algorithms. B.L.H. supervised the research. T.W. wrote the initial version of the manuscript. All authors wrote the final version of the manuscript.

Competing interests
B.L.H. acknowledges outside interest in Arpelos Biosciences and Genyro as a scientific cofounder. All other authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41592-026-03137-3.

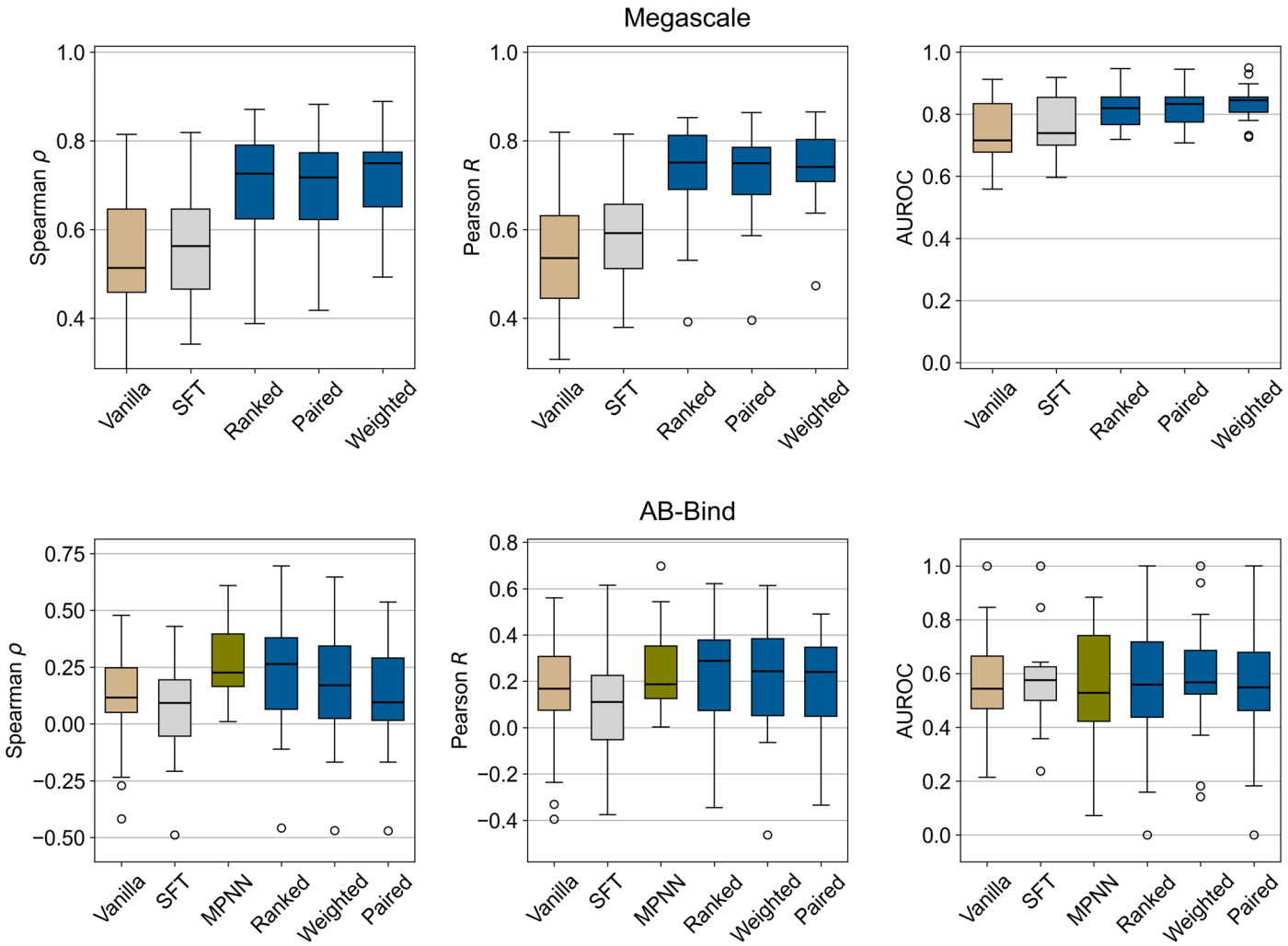
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-026-03137-3.

Correspondence and requests for materials should be addressed to Brian L. Hie.

Peer review information *Nature Methods* thanks Gabriel Rocklin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Arunima Singh, in collaboration with the *Nature Methods* team. Peer reviewer reports are available.

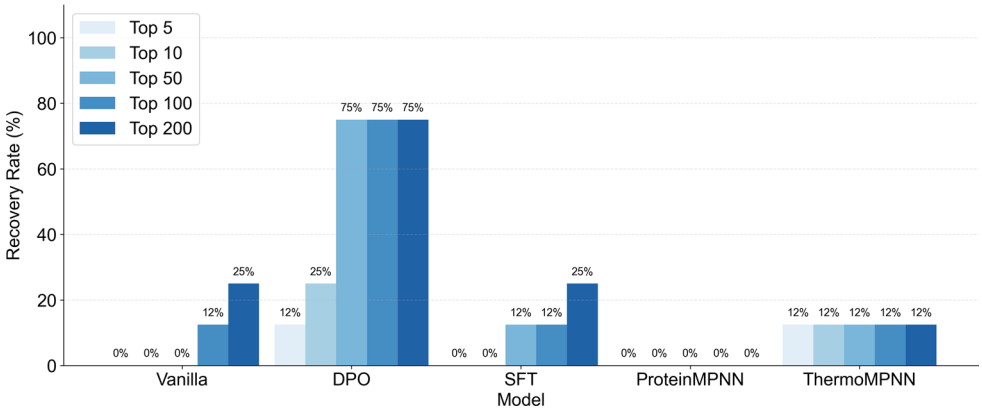
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41592-026-03137-3



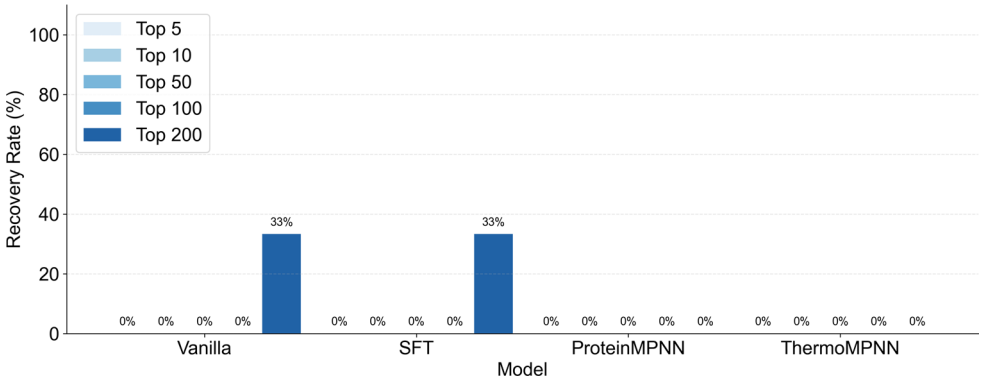
Extended Data Fig. 1 | Performance distribution across protein domains for stability and binding prediction. Distribution of performance across varying protein domains (n=14) for predicting stability changes upon mutation ddG in curated Megascale holdout dataset, and predicting binding-affinity changes upon mutation across varying antibody-antigen complexes (n=27) in AB-Bind dataset.

https://doi.org/10.1038/s41592-026-03137-3



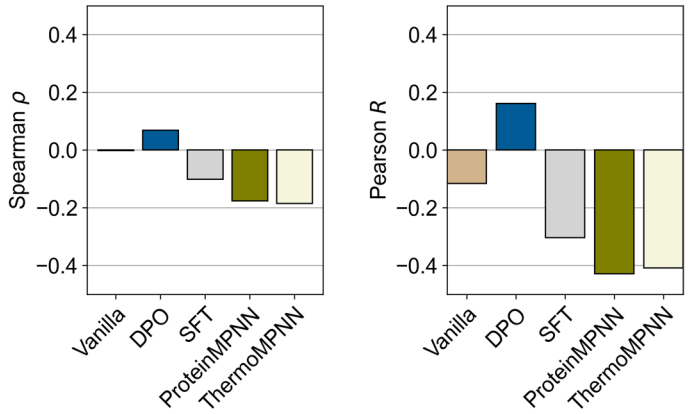
Extended Data Fig. 2 | Recovery of literature-validated stabilizing mutations by design models. Recovery of literature-validated stabilizing mutations across protein design models. Bar plot showing the percentage of n=8 known stabilizing mutations (Milder et al. and Yassine et al.) recovered within the top-ranked predictions of each model. Models were evaluated using ensemble scoring across three PDB structures (6cf5, 6cfg, 2fk0).

https://doi.org/10.1038/s41592-026-03137-3



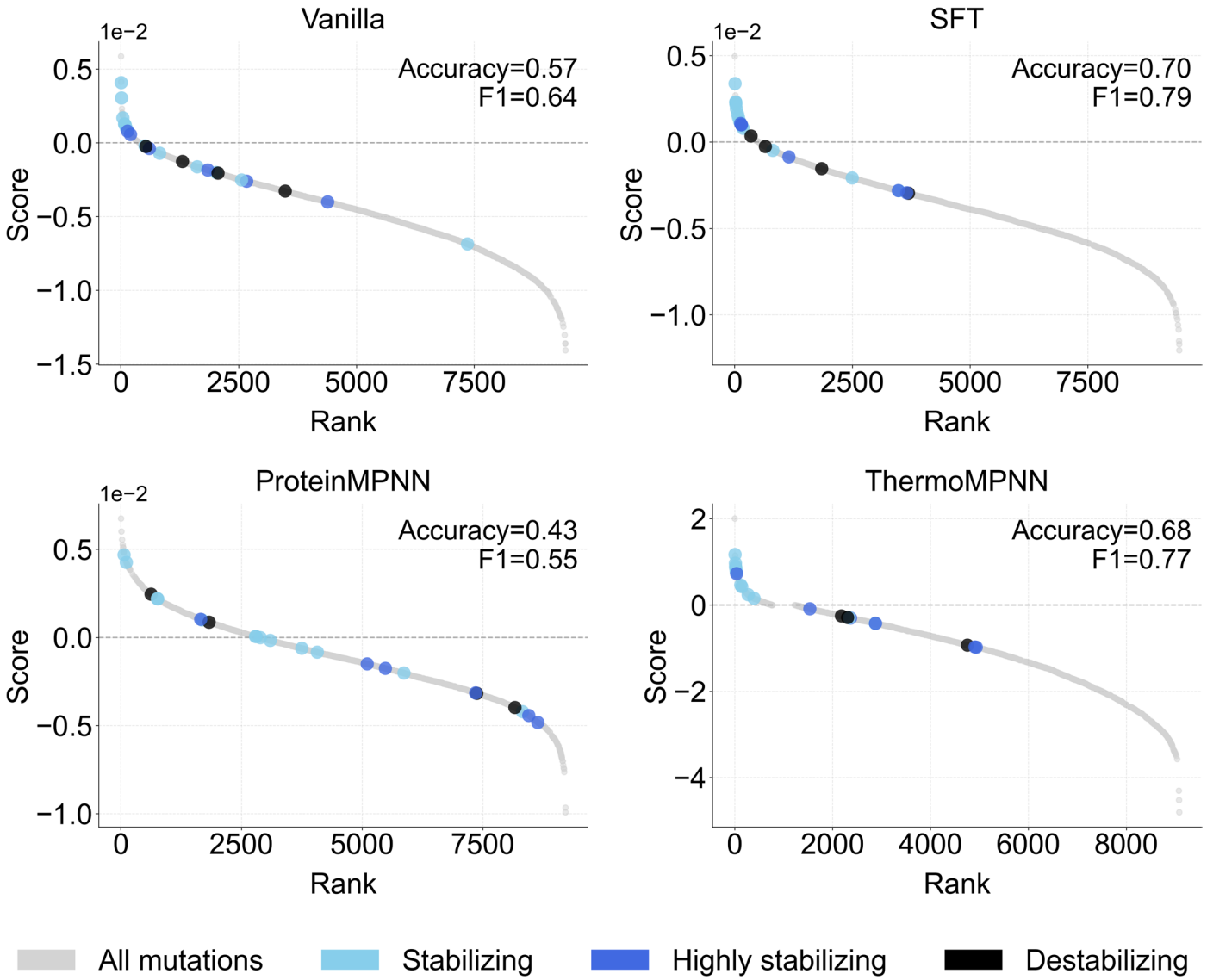
Extended Data Fig. 3 | Recovery of highly stabilizing ProteinDPO mutations by design models. Recovery of highly-stabilizing ProteinDPO-identified single-substitutions across protein design models. Bar plot showing the percentage of n=6 significantly stabilizing (>= 9 C) DPO-identified mutations recovered within the top-ranked predictions of each model. Models were evaluated using ensemble scoring across three PDB structures (6cf5, 6cfg, 2fk0).

https://doi.org/10.1038/s41592-026-03137-3



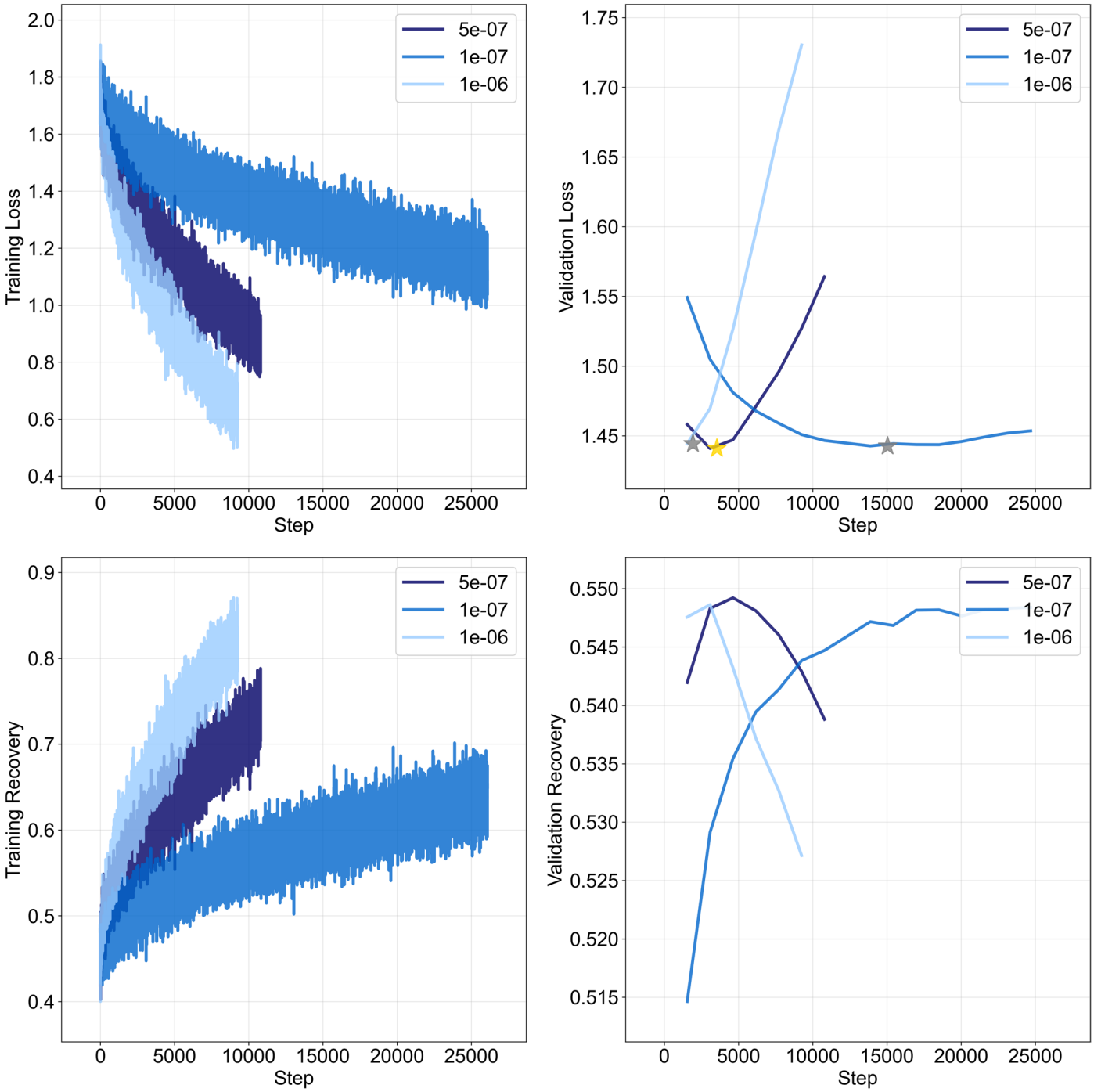
Extended Data Fig. 4 | Ranking correlation between model scores and experimental melting temperatures. Comparison of methods in ability to rank the n=30 experimentally tested single-substitution variants identified by ProteinDPO, measured by Pearson R and Spearman rho correlation coefficients between model scores and temperature of thermal unfolding Tm.

https://doi.org/10.1038/s41592-026-03137-3



Extended Data Fig. 5 | Score distribution placement of stabilizing versus destabilizing mutations. Method comparison in placement of stabilizing (light blue) and destabilizing (black) ProteinDPO-identified and experimentally-tested single-substitution variants within the score distributions of all possible mutations (gray). Mutations are subset to those which had temperature of thermal unfolding (Tm) that was at least 1 degree higher or lower than the native sequence n=23. Accuracy and F1 scores derived from the models' prediction of a given mutation being favorable or unfavorable is also shown. Additionally, the position of the six significantly stabilizing mutations (dark blue) analyzed Figure S3 are shown.

https://doi.org/10.1038/s41592-026-03137-3



Extended Data Fig. 6 | Training loss curves and sequence recovery during supervised fine-tuning. Training and validation loss and sequence recovery curves for training the supervised finetuned model. The selected checkpoint was the checkpoint which best minimized validation loss across all runs, and is denoted with a star. Checkpoints which minimized validation loss for other runs, are denoted with a gray star, and are compared against the selected checkpoint in Table S4.

https://doi.org/10.1038/s41592-026-03137-3

Extended Data Table 1 | Additive versus combined scoring comparison for ProteinDPO models on double mutations

Model	Strategy	Rho	R
Paired	Combined	**0.468**	**0.475**
Weighted	Combined	0.439	0.436
Ranked	Combined	<u>0.463</u>	<u>0.437</u>
Paired	Additive	0.374	0.395
Weighted	Additive	0.340	0.364
Ranked	Additive	0.328	0.358
ThermoMPNN	Additive	0.438	0.415

Additive and whole-sequence scoring on Megascale holdout double mutations (n=20,202) using ProteinDPO models and ThermoMPNN. Additive scoring refers to adding the respective log-likelihoods of each mutation evaluated individually, while combined scoring refers to adding the log-likelihood of both mutations while evaluating them simultaneously.

https://doi.org/10.1038/s41592-026-03137-3

Extended Data Table 2 | Self-consistency evaluation of protein structure generation at different temperatures

Protein	Temp	ProteinDPO			Vanilla			SFT		
		scTM	pLDDT	%Sim	scTM	pLDDT	%Sim	scTM	pLDDT	%Sim
4P79	0.01	0.96	90.2	37.5	0.95	88.8	45.5	0.94	80.5	41.8
	0.05	0.96	90.6	37.9	0.95	88.8	45.2	0.93	81.1	41.4
	0.1	0.96	90.6	37.7	0.95	88.0	45.2	0.93	82.1	41.5
	0.5	0.96	89.3	35.5	0.95	85.6	45.7	0.94	82.0	39.4
	1.0	0.95	86.2	33.4	0.94	82.4	41.5	0.94	82.5	34.7
1STN	0.01	0.97	83.9	40.7	0.99	87.2	64.6	0.99	88.4	66.6
	0.05	0.97	84.1	41.1	0.99	87.2	63.9	0.99	87.8	61.2
	0.1	0.97	83.9	40.5	0.99	87.2	63.5	0.99	88.4	66.2
	0.5	0.96	84.2	41.5	0.99	87.4	62.7	0.98	86.7	54.0
	1.0	0.97	84.1	40.7	0.98	86.4	55.9	0.99	88.4	66.5
1PHP	0.01	0.99	93.4	60.9	0.99+	90.1	79.8	0.99+	91.4	75.2
	0.05	0.99	93.8	58.6	0.99+	91.3	77.4	0.99+	92.2	73.7
	0.1	0.99	93.5	61.2	0.99+	89.9	80.0	0.99+	91.0	75.0
	0.5	0.99	95.0	55.1	0.99+	93.1	69.9	0.99+	92.6	67.5
	1.0	0.99	93.8	60.4	0.99+	90.1	79.4	0.99+	91.4	75.6

Self-consistency of generations for 4P79, 1STN, and 1PHP, evaluated with the TM score of the input backbone and ESMFold structure (scTM) and confidence of the predicted structure (pLDDT). For each temperature metrics were averaged across n=100 samples, for a total of 500 samples per model per protein.

https://doi.org/10.1038/s41592-026-03137-3

Extended Data Table 3 | Experimental melting temperatures for ProteinDPO-identified mutations

Variant	Mutations	T_m (°C)	ΔT_m (°C)
Wild-type	–	45.22	–
Vietnam 2004 Single Mutations			
DPO-A1	AG205T	47.23	+2.02
DPO-A2	AH183L	n.d	–
DPO-A3	A44F	47.57	+2.35
DPO-A4	A44L	47.63	+2.42
DPO-A5	A44Y	47.00	+1.78
DPO-A6	G47F	46.73	+1.52
DPO-A7	G47L	46.57	+1.35
DPO-A8	G47W	45.90	+0.68
DPO-A9	G47Y	47.37	+2.15
DPO-A10	H26W	45.40	+0.18
DPO-A11	H26Y	45.30	+0.08
DPO-A12	H111F	42.43	-2.78
DPO-A13	H111Y	46.63	+1.42
DPO-A14	N81W	56.10	+10.88
DPO-A15	N114I	47.37	+2.15
DPO-A16	N114L	47.80	+2.58
DPO-A17	N114M	47.03	+1.82
DPO-A18	N114V	47.70	+2.48
DPO-A19	P160D	45.80	+0.58
DPO-A20	P160E	45.50	+0.28
DPO-1	K51L	44.29	-0.92
DPO-2	K51C	40.81	-4.41
DPO-5	K58L	54.37	+9.15
DPO-6	E103M	57.93	+12.71
DPO-8	E103L	48.79	+3.58
DPO-11	K51I	43.74	-1.48
DPO-14	N95L	44.98	-0.24
DPO-17	K51M	54.21	+9.00
DPO-20	E103I	58.01	+12.80
DPO-23	N95F	62.21	+16.99
Vietnam 2004 Combination Mutations			
DPO-3	A44F, G47F, H111Y, N81W, N114L	42.64	-2.58
DPO-4	A44F, E103L, G47M, H111Y, K51M, K58L, N114L, N81Y, N95L	55.04 (76.93)	+9.82 (+31.71)
DPO-7	A44F, G47L, H111Y, N81W, N114L	n.d	–
DPO-9	N81W, N114L, A44L	47.05	+1.83
DPO-10	A44F, G47Y, H111Y, N81W, N114L	43.27	-1.99
DPO-12	N81W, N114L, A44F	44.38	-0.84
DPO-13	A44Y, G47F, H111Y, N81W, N114L	42.90	-2.32
DPO-15	G47F, N114L, A44F	48.12	+2.90
DPO-16	A44F, E103L, G47M, H111Y, K51M, K58L, N114L, N81Y, N95I	62.60	+17.38
DPO-18	N81W, N114L, A44Y	44.28	-0.94
DPO-19	A44F, E103L, G47M, H111Y, K51M, K58L, N114L, N81W, N95F	57.43	+12.21
DPO-21	G47L, N114L, A44F	49.97	+4.75
DPO-22	A44F, E103L, G47M, H111Y, K51L, K58L, N114L, N81W, N95I	49.82	+4.60
DPO-24	A44L, G47Y, H111Y, N81W, N114L	41.49	-3.73
DPO-25	A44F, E103L, G47M, H111Y, K51L, K58C, N114L, N81Y, N95F	62.04	+16.82
Texas 2024			
Wild-type	–	61.57	–
DPO-16-TX	A44F, E103L, G47M, H111Y, K51M, K58L, N114L, N81Y, N95I	74.12	+12.55
British Columbia 2024			
Wild-type	–	42.94	–
DPO-16-BC	A44F, E103L, G47M, H111Y, K51M, K58L, N114L, N81Y, N95I	75.05	+32.11

Melting temperatures of protein variants as determined by NanoTemper Prometheus software and improvement over their respective wild-type sequence. At least three replicates were performed per measurement, and reported value is average over replicates. Unless otherwise denoted, all mutations are applied to chain B of the structure as demarcated in PDB 6CFG.

https://doi.org/10.1038/s41592-026-03137-3

Extended Data Table 4 | Learning rate comparison for supervised fine-tuning checkpoints

Learning rate	Megascale (mean)	FireProt	AB-Bind (mean)	Antibody thermal stability
1e-7	0.584/0.561	0.430/0.441	0.10/0.05	0.250/0.236
5e-7	0.588/0.570	0.415/0.429	0.10/0.06	0.258/0.246
1e-6	0.577/0.556	0.422/0.433	0.10/0.04	0.249/0.235

Comparison of different SFT checkpoints trained with varying learning rate on stability and binding prediction in terms of (mean) Pearson R / Spearman rho correlation coefficients.

nature portfolio

Corresponding author(s): Brian Lance Hie

Last updated by author(s): Mar 12, 2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Thermal stability - PR.ThermControl software, version 2.3.1; Biolayer Interferometry (BLI) - Octet System Data Analysis Software version 13.0.3.52
Data analysis	N/A

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

Data

Policy information about availability of data

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

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

Section 8 (Data availability) and section 7.5 (Datasets)

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	The same size of hemagglutinin (HA) variants tested was intended to be as large as possible while falling within a reasonable experimental workload.
Data exclusions	Thermal melting profiles which were clearly highly variable due to improper loading of the sample were excluded, and experiments were re-ran until a reliable measurement could be obtained.
Replication	Thermal melting experiments were performed in triplicate, and BLI experiments used at least two replicate experiments to determine affinity.
Randomization	The order and batch assignments for the production and testing of hemagglutinin (HA) variants was completely randomized.
Blinding	Blinding was not necessary due to the use of objective and unbiased external software for data collection.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	Anti-H5 IgG antibodies 13D4 (Q. Zheng et al., 2011) and CR6261 (D. C. Ekiert et al., 2009), and anti-SARS-CoV-2 RBD antibody AZD3152 (Y. Cai et al., 2024) were produced in-house.

Validation	In-house produced antibodies anti-H5 13D4 and CR6261 have been verified to bind to recombinant hemagglutinin (HA) extracellular domain from A/Vietnam/1203/04 (H5N1) by biolayer interferometry, while anti-SARS-CoV-2 RBD AZD3152 was verified to be non-HA binding. Crystal structures are available for the Fab domains of each antibody bound to their target: 13D4 (6A0Z); CR6261 (3GBN); AZD3152 (8SUO).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Expi293F cells were obtained from Thermo Fisher Scientific (A14527)
Authentication	No further authentication of the cell lines was carried after out after acquisition from their suppliers.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	There were no commonly misidentified lines used in this study.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A