Domain-Adaptive Pretraining for Low-N Antibody Fitness Prediction

Marco De Negri,[*] Giovanni Guidarini,[*] Scott Sussex, Andreas Krause

Abstract

Engineering an antibody for a new target usually starts with only a handful of labelled binding measurements, far too few to fit a fitness model from scratch. We study whether continuing the masked-language-modelling objective of a pretrained protein language model on unlabelled antibody sequence data (*domain-adaptive pretraining*), before aligning it to binding fitness with Direct Preference Optimization (DPO), improves low-data fitness prediction compared to applying DPO directly to a non-adapted model. Working on a deep mutational scanning (DMS) dataset for the C05 antibody, we compare several ways of choosing which sequences and which residue positions this adaptation stage sees. Preliminary results on C05 show three regimes: domain-adaptive pretraining removes a zero-shot rank inversion, improves Spearman correlation below roughly 500 labelled variants, is overtaken by the non-adapted backbone between 1,000 and 5,000 labels, and becomes indistinguishable from it by 50,000.

1 Introduction

Antibody engineering campaigns produce a deep mutational scan (DMS) around a single starting sequence: typically only a few hundred labelled variants, measured against one antigen, for one wild-type (WT). Fitting a fitness model directly on data this scarce is a low-N supervised learning problem, and the model has no chance to learn general antibody sequence structure from the labelled set alone.

Protein language models (PLMs) pretrained on large unlabelled sequence databases offer a way around this. Instead of learning sequence structure and fitness at the same time from a handful of labels, a model can learn sequence structure first, from millions of unlabelled sequences, and only then be pointed at the fitness signal. Two design choices structure how that pointing happens.

The first is *domain-adaptive pretraining*: continuing the model's own pretraining objective (masked language modelling, for our ESM2-family backbones) on sequences drawn from a chosen slice of antibody sequence space, before any labelled fine-tuning happens at all. We use two variants of this: pretraining on the general antibody repertoire (Observed Antibody Space, OAS), and WT-similar adaptation, which restricts that same continued-pretraining objective to a target-biased corpus of OAS sequences selected for similarity to one wild-type CDR-H3.

The second is how the labelled DMS data itself is used once domain-adaptive pretraining is done. We use Direct Preference Optimization [DPO, Rafailov et al., 2024]. Instead of regressing directly onto a fitness score, DPO reframes the labelled set as pairwise preferences (this sequence binds better than that one) and fine-tunes the model's own sequence likelihoods to respect those preferences, with the pre-DPO model acting as a KL anchor against overfitting. Domain-adaptive pretraining changes what the model knows before it sees any fitness label, and DPO changes what it prefers once it does.

Both pieces exist separately in the literature (see Section 2); prior work has already chained a similar adaptation-then-DPO recipe [Ferragu et al., 2025] for general protein fitness landscapes,

[1] Equal contribution.

Preprint.

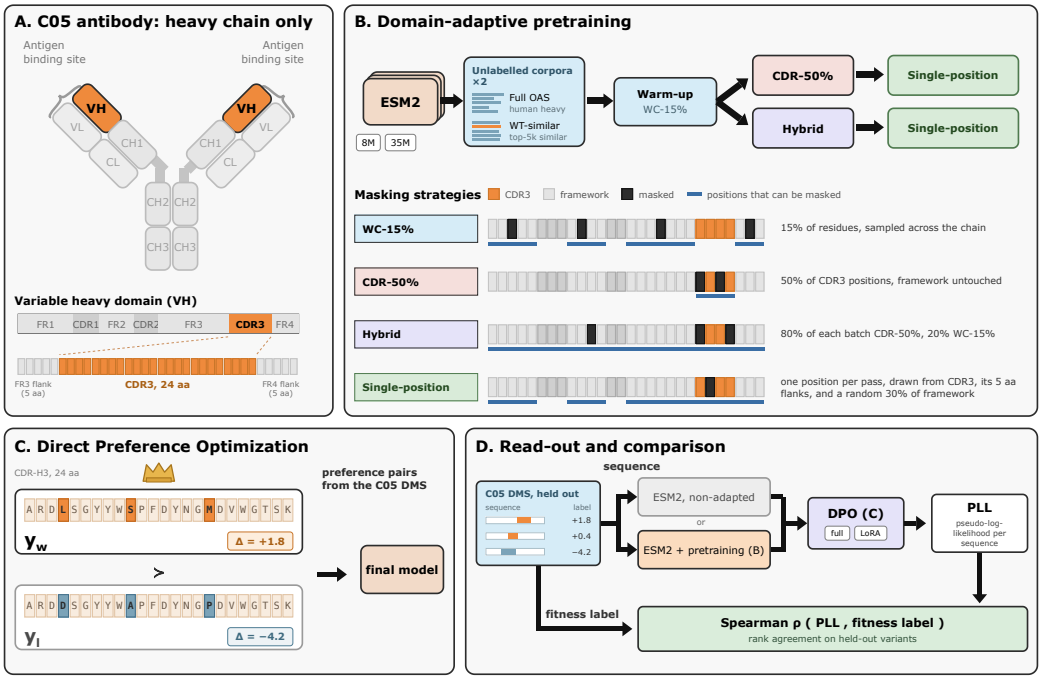


Figure 1: **Overview of the domain-adaptive pretraining and preference optimization framework.** **(a)** The C05 construct. Only the heavy chain is modelled: the variable domain (V_H, orange) and, below it, the framework and CDR layout of that domain, with the 24-residue CDR-H3 varied by the DMS libraries and the 5-residue flanks on either side. **(b)** Domain-adaptive pretraining. Both ESM2 backbones are warmed up under whole-chain masking on one of two unlabelled corpora, then branched across CDR-50% and Hybrid masking, and each branch is continued under single-position masking. The schematics below give one example mask per policy: orange marks CDR-H3, gray the framework, black the positions masked on that pass, and the blue rule underneath spans the positions the policy is allowed to mask. **(c)** Preference optimization. The labelled DMS data enters as pairs: the variant with the higher enrichment (y^w) is preferred over the lower one (y^l), and DPO moves the model's own sequence likelihoods to respect that ordering. Highlighted residues mark the positions at which the two sequences differ, which are the positions the pairwise score is restricted to. **(d)** Read-out. Every model, adapted or not and fine-tuned with full or LoRA DPO, scores held-out variants by pseudo-log-likelihood. The Spearman rank correlation (ρ) between those scores and the measured enrichment labels is the reported metric.

but not, to our knowledge, for antibody binding fitness specifically. Our first contribution is to test that combination directly on a C05 DMS dataset, comparing a domain-adaptively pretrained and a non-adapted ESM2-35M under the same DPO recipe as the number of labelled sequences grows. The C05 libraries are unusually large for a DMS, which is what lets us subsample them down to the budgets a real campaign starts with and read off how the comparison changes with the number of labels. Adaptation removes a zero-shot rank inversion and improves correlation below roughly 500 labelled variants. Past that point the ordering reverses, at K = 5,000 the non-adapted backbone is ahead by 0.05 to 0.06 Spearman on both evaluation sets, and the two converge only at the largest budgets. Figure 1 summarizes the full pipeline.

Contributions.

- **Framework Integration:** We evaluate the combination of domain-adaptive pretraining (DAP) and Direct Preference Optimization (DPO) on an antibody binding benchmark (C05 DMS) across low-data scaling regimes (Sections 3 and 4.1).
- **Pretraining and Masking Ablations:** We compare pretraining corpora (broad OAS vs. WT-similar) and masking strategies (whole-chain, CDR-50%, hybrid, single-position), finding

that restricting sequences to target-similar CDR-H3 loops improves zero-shot prediction of measured C05 binding fitness compared to broad repertoire MLM (Sections 4.2 and 4.3, Appendix D.4).

- **Label-Free Checkpoint Selection:** We test a self-supervised rule for choosing when to stop domain-adaptive pretraining, and find that in Phase 1 the checkpoint it returns lands on the rising part of the zero-shot fitness trajectory, close to its maximum (Sections 3.2, 4.3).
- **Data-Scaling Observations:** We evaluate fitness prediction across two backbones (ESM2-8M, ESM2-35M) and fine-tuning variants (full DPO, LoRA DPO), showing that DAP improves rank correlation in low-data regimes, whereas at intermediate sample sizes the non-adapted backbone overtakes it, and the two converge only at the largest budgets (Section 4.1, Appendix D.5).

2 Related Work

Domain-Adaptive Pretraining for Protein Language Models. General protein language models learn broad structural and evolutionary patterns across large sequence databases [Lin et al., 2023, Olsen et al., 2022]. However, global pretraining often fails to capture fine-grained functional distributions around specific target proteins. To address this, Biswas et al. [2021] introduced "evotuning," continuing unsupervised pretraining on evolutionary neighbors of a target sequence before fitting supervised models on small experimental datasets. In antibody modeling, domain adaptation has focused on tailoring pretraining corpora and masking strategies. For example, AbLang-2 [Olsen et al., 2024] and CDR-aware language models [Talaei et al., 2025] demonstrated that filtering antibody databases like OAS and applying region-targeted masking (such as focusing on CDRs) improves token recovery and downstream representation quality. Existing antibody adaptation studies evaluate representations primarily through frozen embedding regression on full datasets, while we focus on examining low-N data-scaling curves.

Likelihood Adaptation and Preference Alignment. When downstream labelled data is scarce, supervised prediction heads trained on frozen embeddings can easily overfit or fail to generalize to distant mutations. An alternative approach directly adapts the language model's likelihoods using sequence-level ranking or preference objectives. Hawkins-Hooker et al. [2024] showed that fine-tuning PLM likelihoods with a Bradley-Terry ranking loss over sequence pairs improves few-shot fitness prediction compared to frozen embeddings. Adapting Direct Preference Optimization [DPO, Rafailov et al., 2024] to protein models, ProteinDPO [Widatalla et al., 2026] aligned model likelihoods with experimental stability measurements by converting continuous fitness scores into pairwise preferences. Recently, g-DPO [Ferragu et al., 2025] combined unsupervised fine-tuning with preference optimization to reduce pairwise training redundancy across protein landscapes.

Preference Optimization in Low-Data Antibody Regimes. While domain adaptation and preference optimization have been studied separately or on general proteins, their combination remains unexamined for antibody binding under severe data constraints. Antibody binding presents distinct challenges: sequence variation is heavily concentrated in hypervariable loops like CDR-H3, and early experimental rounds yield only small sets of labelled variants. Our work connects these two paradigms by systematically evaluating how domain-adaptive pretraining, varying corpus selection (broad vs. wild-type similar) and masking policies (whole-chain, CDR-focused, hybrid, and single-position), interacts with subsequent DPO fine-tuning across low-N data-scaling regimes on antibody binding benchmarks.

3 Methods

Our method consists of two stages: (i) *domain-adaptive pretraining* on unlabelled antibody sequences using masked language modeling (MLM), followed by (ii) *preference alignment* on deep mutational scanning (DMS) fitness measurements using Direct Preference Optimization (DPO). The two stages are panels (b) and (c) of Figure 1, which also shows the target construct (Figure 1a) and the scoring protocol every checkpoint is evaluated under (Figure 1d). Model selection during pretraining relies strictly on label-free validation metrics, ensuring a realistic low-data downstream setup.

Model Backbones. We evaluate our pipeline using two backbones from the ESM2 family [Lin et al., 2023]: `esm2_t6_8M_UR50D` (8M parameters, 6 layers; referred to as **ESM2-8M**) and `esm2_t12_35M_UR50D` (35M parameters, 12 layers; referred to as **ESM2-35M**).

3.1 Datasets

Unlabelled OAS Corpus. To train models on general antibody sequence space, we extract human heavy-chain sequences from the Observed Antibody Space (OAS) database [Olsen et al., 2022]. Following the AbLang-2 processing pipeline [Olsen et al., 2024], we remove non-canonical Ig-fold sequences and heavy fragments, map non-standard amino acids to `X`, and deduplicate sequences at 95% identity using Linclust [Steinegger and Söding, 2018]. The deduplicated pool of 201M sequences is split 90/5/5 into training, validation, and test sets.

WT-Similar Corpus. To test whether pretraining on sequences closely surrounding a target antibody outperforms broad antibody exposure, we construct a target-restricted corpus from OAS. We score every candidate CDR-H3 against the target C05 wild-type (WT) CDR-H3 using a normalized BLOSUM62 global alignment and select the top $K = 5{,}000$ ranked sequences.

C05 DMS Benchmark. Our downstream fitness benchmark comprises deep mutational scanning (DMS) data for the C05 antibody scored by binding enrichment against its target antigen, two strains of the influenza virus. The dataset spans two independently sorted libraries varying the same 24-residue CDR-H3 loop: **ED3** (275,354 variants, median edit distance 3 from WT) and **ED5** (428,821 variants, median edit distance 5). Within-library enrichment is defined as:

$$\text{enrichment} = \log_2\left(\frac{f_{\text{pos}}}{f_{\text{neg}}}\right), \tag{1}$$

where f_{pos} and f_{neg} are a variant's read frequencies in binding-positive and binding-negative sorted pools. Models are trained and selected exclusively on ED3 across subset sizes ranging from $N = 50$ to $N = 50{,}000$, while ED5 is held out entirely as an independent evaluation library. Appendix D gives the full grid.

Full dataset characteristics are detailed in Appendix A.

3.2 Domain-Adaptive Pretraining

Domain-adaptive pretraining (DAP) continues ESM2's original MLM objective on unlabelled sequences. Optimization uses AdamW with a fixed peak learning rate of 1×10^{-4} (selected via hyperparameter sweep, Appendix C). Pretraining proceeds in three sequential phases:

Pretraining Curriculum.

- **Phase 0 (P_0, Whole-Chain Warm-up):** A single whole-chain pass over the OAS corpus starting from base ESM2 weights to establish baseline antibody statistics.
- **Phase 1 (P_1, Domain Adaptation):** Branches P_0 four ways across two masking strategies (*CDR-50%* vs. *Hybrid*) and two corpora (*OAS* vs. *WT-similar*).
- **Phase 2 (P_2, Single-Position Continuation):** Continues each P_1 branch under *single-position masking* to align pretraining conditionals directly with point-mutation scoring.

We refer throughout to three points on a trajectory: the branch point, where P0 stops and the four P1 arms diverge; the switch point, where a P1 arm stops and its P2 continuation begins; and the final checkpoint, the P2 checkpoint carried forward to preference optimization.

Figure 1b draws this branching, with the positions each policy may mask marked beneath the sequence schematic.

Masking Strategies. We evaluate four masking policies targeting different regions of the antibody sequence:

- **Whole-Chain (WC-15%):** Standard BERT-style masking [Devlin et al., 2019] applying a 15% uniform mask across the entire sequence ([Talaei et al., 2025]).

- **CDR-Focused (CDR-50%):** Restricts masking strictly to the CDR-H3 loop at a 50% masking rate ([Talaei et al., 2025]).
- **Hybrid:** A batch-level mixture where 80% of sequences receive CDR-50% masking and 20% receive WC-15% masking, designed to mitigate framework degradation ([Talaei et al., 2025]).
- **Single-Position:** Replaces exactly one eligible position per pass with a `[MASK]` token, matching the exact conditional distribution evaluated during scoring.

Implementation details and token corruption rules for each policy are expanded in Appendix B.

Label-Free Checkpoint Selection. To avoid leaking downstream fitness information during pretraining, checkpoints are selected using self-supervised token-recovery metrics on held-out OAS validation sequences. Among checkpoints satisfying a framework recovery constraint, we select the one maximizing CDR-H3 recovery, breaking ties toward the earliest [Talaei et al., 2025]:

$$\mathrm{FR}(\mathrm{ref}) - \mathrm{FR}(t) \leq 0.1 \text{ percentage points}, \tag{2}$$

where $\mathrm{FR}(\mathrm{ref})$ is the framework accuracy of the reference checkpoint (P_0 warm-up for base branches, or the P_1 switch point for continuations). This one-sided rule prevents framework degradation while allowing framework accuracy gains.

3.3 Preference Optimization

Following pretraining checkpoint selection, models are fine-tuned on preference pairs constructed from the labelled DMS fitness benchmark using Direct Preference Optimization [DPO, Rafailov et al., 2024].

Masked Pseudo-Log-Likelihood Score. Because masked protein language models do not compute direct sequence-level joint probabilities, we define a restricted pseudo-log-likelihood score $S_\theta(y; M)$ restricted to the mutation set $M = \{i : y_i^w \neq y_i^l\}$ where the preferred sequence y^w and dispreferred sequence y^l differ:

$$S_\theta(y; M) = \sum_{i \in M} \log p_\theta\big(y_i \mid y_{\setminus i}\big)\,. \tag{3}$$

Evaluation score. Equation 3 is restricted to a mutation set and is defined only for a pair. At evaluation time each held-out variant is scored on its own by pseudo-log-likelihood over full sequence:

$$PLL_\theta(y) = \sum_{i \in P} \log p_\theta\big(y_i \mid y_{\setminus i}\big) \tag{4}$$

where P are all positions in y. This is the score correlated against enrichment in all reported Spearman values, including the N = 0 anchor.

Rank-Based Pair Construction. Preference pairs are constructed using rank-based gap sampling [Widatalla et al., 2026]. Sequence variants are ordered by enrichment relative to wild-type $\Delta = s - s_{\mathrm{WT}}$ and paired at a fixed rank gap level $g = 0.5$. Pairs are drawn across four categories defined by the sign of : (i) both variants above WT, (ii) both below WT, (iii) wild-type anchored, and (iv) one above and one below. Sampling parameters and score thresholds are detailed in Appendix D.

Weighted DPO Objective. Models are optimized using weighted DPO [Widatalla et al., 2026], scaling loss gradients by continuous enrichment scores:

$$\ell(y^w, y^l) = -w^w \log \sigma(\beta\delta) - w^l \log \sigma(-\beta\delta), \tag{5}$$

where $\delta = (S_\theta(y^w; M) - S_{\mathrm{ref}}(y^w; M)) - (S_\theta(y^l; M) - S_{\mathrm{ref}}(y^l; M))$, $\beta = 0.1$, and sample weights $(w^w, w^l) = \mathrm{softmax}(r^w/T, r^l/T)$ with temperature $T = 1.0$ derived from raw enrichment scores r^w, r^l.

Parameter-Efficient Fine-Tuning (LoRA). We compare full-parameter DPO against Low-Rank Adaptation [LoRA, Hu et al., 2022] on query and value projections (rank $r = 8$, $\alpha = 16$). Restricting parameter movement acts as an explicit regularizer against overfitting when fine-tuning on small N preference pairs.

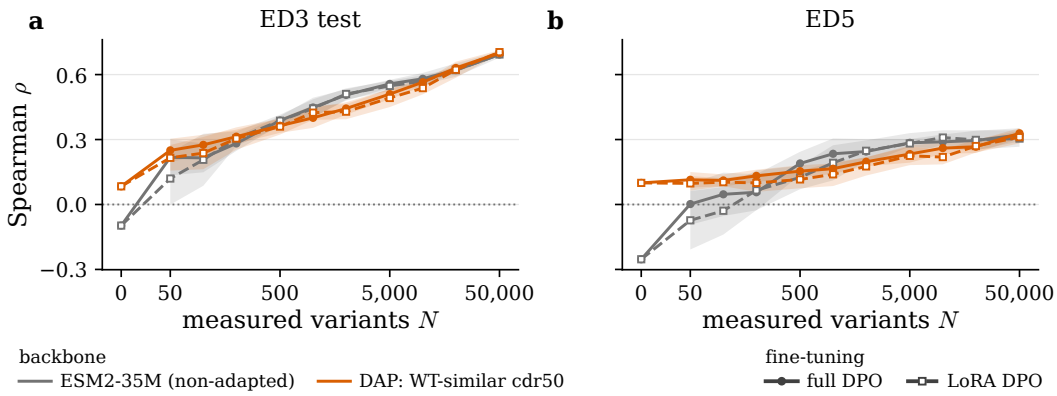


Figure 2: **Domain-adaptive pretraining improves low-N fitness prediction across fine-tuning paradigms.** In both panels, color identifies the backbone (gray, non-adapted; orange, domain-adaptively pretrained) and fill identifies the fine-tuning paradigm (filled and solid, full DPO; open and dashed, LoRA DPO). **(a)** Spearman rank correlation ρ on the ED3 test split as a function of labeled training variants N for ESM2-35M backbones. Bands are ± 1 standard deviation over the 10 random draws of N training variants at each point. **(b)** Corresponding downstream evaluation on the held-out ED5 dataset across N. The paired differences between the two backbones, with their confidence intervals, are given in Table 1.

4 Results

4.1 Domain-adaptive pretraining improves low-N fitness prediction

We now inspect how domain-adaptive pretraining impacts downstream binding fitness prediction across varying amounts of labeled training variants N. We compare the non-adapted ESM2-35M backbone against a domain-adaptively pretrained (DAP) backbone, whose label-free checkpoint selection procedure will be detailed in Section 4.3, holding the DPO preference alignment protocol fixed while varying the number N of training sequences taken from ED3 train split. Evaluations are conducted under both full DPO and LoRA DPO across the ED3 test split (Figure 2a) and the held-out ED5 dataset (Figure 2b), with paired differences summarized in Table 1.

In the zero-shot regime ($N = 0$), the non-adapted backbone yields negative Spearman correlation on ED5 (Figure 2b) and near-zero correlation on ED3 (Figure 2a). Domain-adaptive pretraining eliminates this rank inversion, establishing a positive correlation baseline that produces the largest paired difference observed across the scaling curve (Table 1). In low-data regimes ($N < 500$), the DAP backbone achieves higher Spearman correlations than the non-adapted baseline under both full DPO and LoRA DPO (Figure 2a, b). As N increases, this performance advantage systematically shrinks as experimental labels substitute for the sequence prior supplied by pretraining. Around $N \approx 500$, the paired difference between backbones crosses zero (Table 1). Between $N = 1{,}000$ and $N = 5{,}000$ the difference is negative on both evaluation sets with 95% intervals excluding zero, so the non-adapted backbone is not merely matched but ahead by 0.05 to 0.07 Spearman (Table 1). The two converge only at the largest budgets, sitting within 0.02 Spearman of each other on ED5 by $N = 50{,}000$ (Figure 2b). Domain-adaptive pretraining therefore helps when labels are scarce, costs accuracy at intermediate budgets, and makes no difference once labels are abundant.

4.2 Region-level token recovery through the pretraining phases

We next examine how the different masking strategies affect the model during pretraining. We track token recovery separately for the framework regions (FR1-FR4) and the hypervariable regions (CDR1-CDR3), with particular attention to CDR-H3 (Figure 3).

Phase 0: Whole-Chain Warm-up. We first continue standard whole-chain masked language modelling on the broad OAS corpus (P_0). During this phase, framework recovery increases very quickly and reaches a plateau early in training: more than 90% of framework tokens are recovered

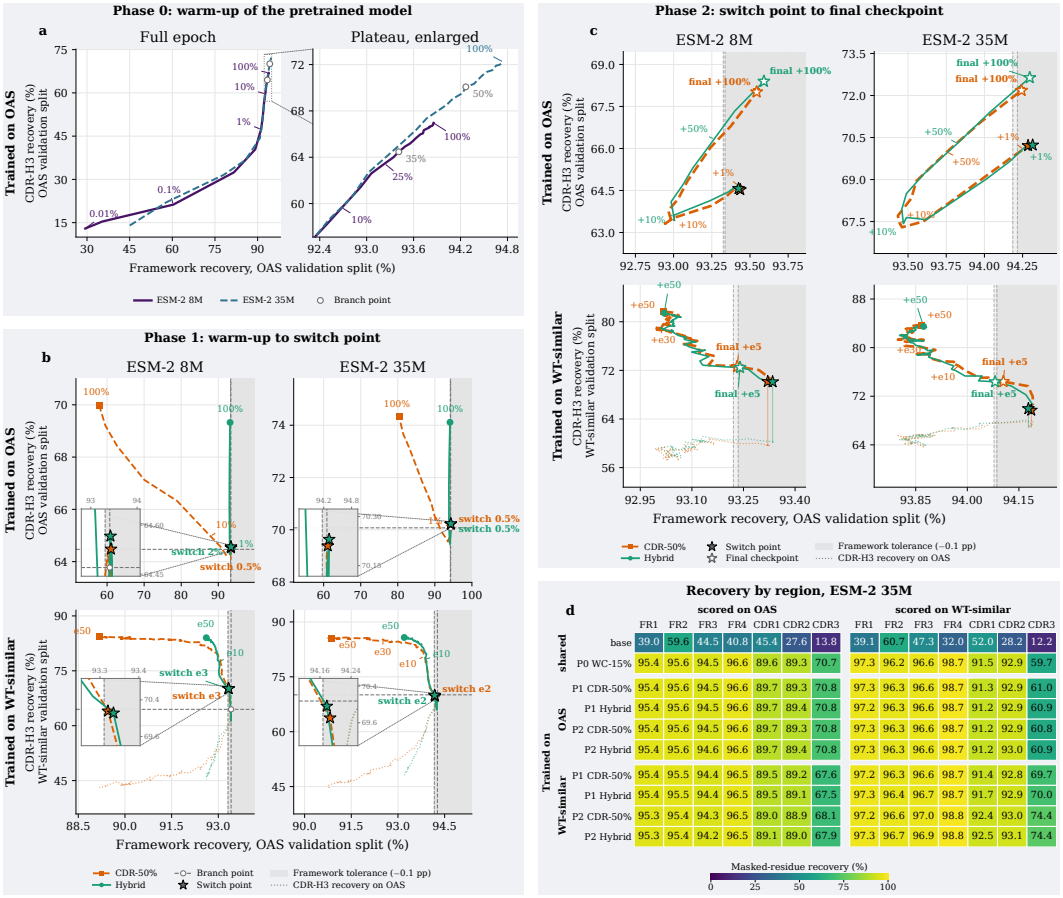


Figure 3: **Self-supervised token recovery dynamics across pretraining phases. (a-c)** Trajectories plotting framework recovery (%) against CDR-H3 recovery (%) on held-out validation splits across: **(a)** Phase 0 (P_0) whole-chain warm-up (0.01% to 100% of epoch; inset expands plateau region), **(b)** Phase 1 (P_1) domain adaptation under CDR-50% (orange) and Hybrid (green) masking, and **(c)** Phase 2 (P_2) single-position continuation for ESM2-8M (left) and ESM2-35M (right). Open circles indicate branch points in Phase 0; filled stars mark Phase 1 switch points; open stars mark final label-free selected checkpoints. The dotted gray boundary represents the framework degradation constraint (≤ 0.1 percentage points from reference). For WT-similar branches, framework recovery is measured on the OAS validation split, keeping it comparable across corpora, and CDR-H3 recovery is measured on the WT-similar held-out split. OAS branches use the OAS split for both. **(d)** Matrix of masked token recovery accuracy (%) by sequence region (FR1-FR4, CDR1-CDR3) for ESM2-35M. Rows group the checkpoints shared by both branches (base model and P_0 warm-up) above one block per pretraining corpus; columns report both validation splits, general OAS (left) and WT-similar (right).

Table 1: **Downstream binding fitness correlation and performance gap ($\Delta\rho$) across training regimes.** Paired differences ($\Delta\rho = \rho_{\text{DAP}} - \rho_{\text{Base}}$) between the domain-adaptively pretrained and non-adapted ESM2-35M backbones, on the ED3 test split and ED5. Each cell gives the mean difference over the 10 random draws of N training variants and, in brackets, its 95% confidence interval. The deterministic $N = 0$ anchor carries a single value and so no interval.

	ED3 test		ED5	
N	full DPO	LoRA DPO	full DPO	LoRA DPO
0	$+0.18$	$+0.18$	$+0.35$	$+0.35$
50	$+0.03$ [-0.03, $+0.10$]	$+0.10$ [$+0.02$, $+0.17$]	$+0.11$ [$+0.05$, $+0.18$]	$+0.17$ [$+0.08$, $+0.26$]
200	$+0.03$ [$+0.00$, $+0.06$]	$+0.00$ [-0.05, $+0.05$]	$+0.08$ [$+0.01$, $+0.14$]	$+0.03$ [-0.03, $+0.10$]
1,000	-0.05 [-0.08, -0.01]	-0.01 [-0.03, $+0.02$]	-0.07 [-0.12, -0.02]	-0.05 [-0.11, $+0.01$]
5,000	-0.05 [-0.07, -0.02]	-0.05 [-0.09, -0.02]	-0.05 [-0.08, -0.02]	-0.06 [-0.11, -0.01]
20,000	$+0.00$ [-0.03, $+0.04$]	-0.00 [-0.03, $+0.03$]	-0.03 [-0.06, $+0.01$]	-0.03 [-0.05, -0.01]
50,000	$+0.01$ [-0.01, $+0.02$]	$+0.01$ [$+0.01$, $+0.02$]	$+0.01$ [-0.01, $+0.03$]	$+0.01$ [-0.03, $+0.04$]

within the first part of an epoch (Figure 3a). This is expected because framework residues make up most of the sequence and are relatively conserved across antibodies. The model therefore learns common sequence patterns in the framework quickly, while CDR-H3 recovery continues to improve more gradually.

Once framework recovery has plateaued, continuing whole-chain pretraining provides little additional benefit relative to its computational cost. The checkpoint-selection rule of Section 3.2 cannot apply to P_0 as framework and CDR-H3 recovery rise together throughout the first phase, so there is no trade-off between regions to resolve. We therefore branch on compute instead, at the earliest checkpoint whose framework recovery is within 0.5 percentage points of the accuracy reached at the end of the epoch (Appendix B). This selects 50% of an epoch for ESM2-35M and 35% for ESM2-8M, leaving the remaining training budget for targeted adaptation.

Phase 1: Domain Adaptation. We next branch the P_0 checkpoint across two masking strategies and two sequence corpora (Figure 3b,d). The masking strategies are CDR-50%, which restricts prediction targets to CDR-H3, and Hybrid, which applies CDR-50% masking to 80% of sequences and whole-chain masking to the remaining 20%.

With CDR-50% masking the model gradually loses some of its framework representations because those positions no longer receive direct training. Hybrid retains enough whole-chain supervision to preserve the framework, while still supplying 86% of the CDR-H3 supervision that CDR-50% provides (Appendix B). The two policies therefore reach almost the same CDR-H3 recovery and differ mainly in what happens to the framework (Figure 3b).

The choice of pretraining corpus has a similarly strong effect. Continuing on the full OAS corpus preserves general antibody sequence patterns, but gives only modest improvements in the C05 CDR-H3 region. In contrast, restricting the training data to sequences similar to the C05 wild type focuses the model much more strongly on the local sequence space around the target. This produces a clear increase in CDR-H3 recovery on the WT-similar validation set, with only a small reduction in performance on the broader OAS sequences (Figure 3d). In other words, the model becomes more specialized to the C05 sequence neighborhood.

Phase 2: Single-Position Continuation. The final pretraining stage replaces multi-position masking with single-position masking (P_2). Each training example therefore asks the model to predict one masked residue, matching the conditional probability used later to score individual mutations.

This change further improves CDR-H3 recovery across both pretraining corpora, shown by the Phase 2 trajectories for both backbones (Figure 3c) and by the per-region breakdown for ESM2-35M (Figure 3d), with the ESM2-8M counterpart in Figure 6. Unlike the earlier masking strategies, the purpose of this stage is not primarily to increase the amount of supervision on CDR-H3, but to make the pretraining objective resemble the conditional probabilities used for mutation scoring.

These results show that the three stages play different roles: whole-chain pretraining rapidly establishes general antibody sequence statistics, CDR-focused adaptation shifts learning toward the target

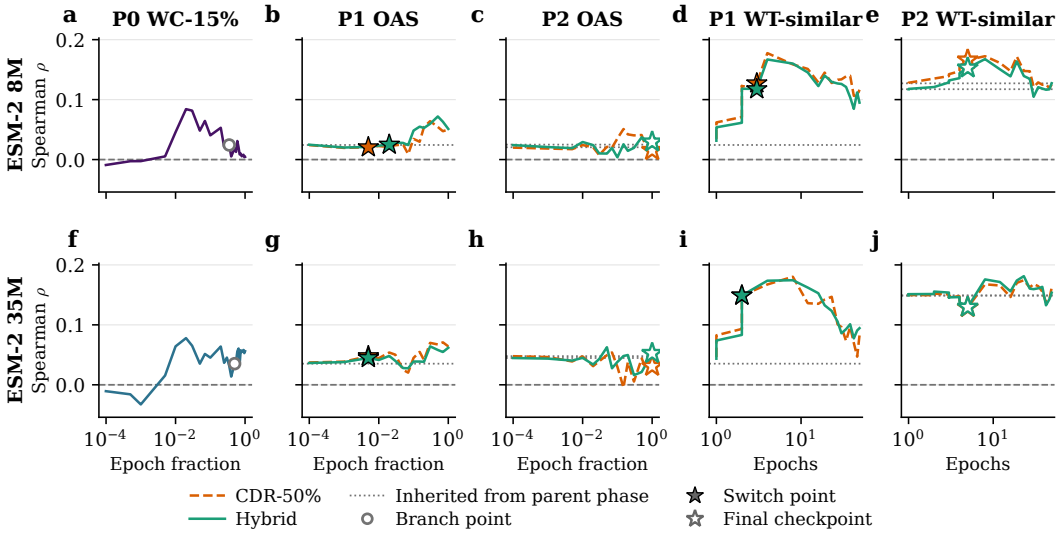


Figure 4: **Zero-shot DMS fitness prediction correlation across pretraining phases and corpora.** Zero-shot Spearman rank correlation (ρ) between sequence pseudo-log-likelihood and C05 DMS enrichment labels plotted against pretraining epoch fractions for **(a-e)** ESM2-8M and **(f-j)** ESM2-35M across curriculum phases: Phase 0 warm-up **(a, f)**, Phase 1 OAS adaptation **(b, g)**, Phase 2 OAS continuation **(c, h)**, Phase 1 WT-similar adaptation **(d, i)**, and Phase 2 WT-similar continuation **(e, j)**. Orange curves represent CDR-50% masking; green curves represent Hybrid masking. Dotted gray segments show parent phase trajectories inherited at branch points. Open circles, filled stars, and open stars denote branch points, Phase 1 switch points, and final label-free selected checkpoints, respectively.

loop, and single-position masking aligns the training objective with the subsequent mutation-scoring task.

4.3 Label-free recovery indicates when the model is improving

In a real antibody-engineering setting, downstream DMS labels are unavailable during pretraining. We therefore ask whether the self-supervised token-recovery metric of Section 3.2, computed on unlabelled sequences, can indicate when the model is becoming better at predicting the C05 fitness landscape. To test this, we track the zero-shot Spearman correlation between the model's pseudo-log-likelihood scores and the C05 DMS enrichment values throughout the pretraining trajectories (Figure 4).

Zero-Shot Correlation on the Broad Corpus. During P_0, whole-chain pretraining on the broad OAS corpus produces a small early increase in zero-shot correlation, followed by a decline toward near-zero values (Figure 4a,f). This limited signal is consistent with the distance between the broad pretraining corpus and the C05 target. C05 contains a 24-residue CDR-H3 with no close relative in general antibody repertoires. Across 400,000 unique OAS CDR-H3 loops the closest match covers only half of it (50% identity), and selecting the most similar 5,000 raises the median without moving that ceiling (Appendix A).

The model therefore becomes better at learning general antibody sequence patterns, but this does not translate into a better understanding of the specific mutations that affect C05 binding. Continuing pretraining on the broad OAS corpus produces little further improvement in zero-shot fitness correlation (Figure 4b,c,g,h).

Zero-Shot Correlation on the WT-Similar Corpus. The picture changes when pretraining is restricted to the WT-similar corpus (Figure 4d,i). Training on sequences surrounding the C05 wild type produces a clear increase in zero-shot Spearman correlation from the near-zero baseline. This

improvement occurs during P_1, when the model is both exposed to a more target-relevant sequence distribution and trained with CDR-focused or Hybrid masking.

In contrast, continuing this adaptation with single-position masking in P_2 does not provide evidence of a further improvement in zero-shot ranking (Figure 4e,j). The main gain therefore comes from target-specific adaptation during P_1, not from the additional P_2 continuation.

Label-Free Checkpoint Selection. The self-supervised checkpoint-selection rule provides a useful way to identify this improvement without access to C05 fitness labels. For P_1, the selected checkpoint falls along the rising portion of the zero-shot correlation trajectory and close to its subsequent maximum (Figure 4d,i). The rule therefore identifies a point at which CDR-H3 recovery has improved while framework recovery remains preserved, without using the downstream DMS measurements.

This agreement matters because the checkpoint-selection criterion is based entirely on token recovery, whereas Figure 4 evaluates the model against experimental fitness labels that are never used during pretraining. The results suggest that self-supervised token recovery can provide a practical way to choose a pretraining checkpoint without the downstream DMS labels.

5 Discussion

Evaluating downstream preference alignment across data regimes, we have shown that domain-adaptive pretraining eliminates zero-shot rank inversion ($N = 0$) and improves mean Spearman correlation in low-data settings ($N < 500$), though substantial variance across subsampling draws at $N \leq 200$ indicates that early performance gains remain sensitive to subset composition. Beyond $N \approx 500$ the ordering reverses: between $N = 1{,}000$ and $N = 5{,}000$ the non-adapted backbone is ahead with 95% intervals excluding zero, and the two converge only at $N \geq 20{,}000$. Within the pretraining stage itself, we extended the evaluation of sequence masking strategies from Talaei et al. [2025] to a single-chain CDR-H3 setting, showing that while CDR-50% masking maximizes loop recovery, a hybrid policy (80% CDR-50%, 20% whole-chain) is required to prevent framework degradation while preserving CDR-H3 performance. Consequently, for campaigns with small experimental budgets ($N < 500$), we recommend combining target-restricted sequence filtering with CDR-focused masking as an initialization step prior to DPO; CDR-50% and hybrid masking are statistically indistinguishable on downstream fitness, so hybrid is the safer default because it also holds the framework in place. For $N \geq 1{,}000$, domain-adaptive pretraining is not merely unnecessary but counterproductive, and direct preference alignment on a non-adapted backbone is the better choice. A primary limitation of our study is that we evaluated this pipeline on a single DMS benchmark centered on antibody C05; future work should evaluate this approach across multiple wild-types, multi-chain constructs, and target antigens to confirm the generalizability of these scaling thresholds.

References

Surojit Biswas, Grigory Khimulya, Ethan C. Alley, Kevin M. Esvelt, and George M. Church. Low-n protein engineering with data-efficient deep learning. *Nature Methods*, 18:389–396, 2021. URL `https://www.nature.com/articles/s41592-021-01100-y`.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, 2019. URL `https://arxiv.org/abs/1810.04805`.

Constance Ferragu, Jonathan D. Ziegler, Nicolas Deutschmann, Arthur Lindoulsi, Eli Bixby, and Cradle ML Team. g-dpo: Scalable preference optimization for protein language models, 2025. URL `https://arxiv.org/abs/2510.19474`.

Alex Hawkins-Hooker, Shikha Surana, Jake Simons, Jakub Kmec, Oliver Bent, and Paul Duckworth. Likelihood-based fine-tuning of protein language models for few-shot fitness prediction and design. *bioRxiv*, 2024. URL `https://www.biorxiv.org/content/10.1101/2024.05.28.596156v4`.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International*

Conference on Learning Representations (ICLR), 2022. URL https://arxiv.org/abs/2106.09685.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. URL https://www.science.org/doi/10.1126/science.ade2574.

Tobias H. Olsen, Fergus Boyles, and Charlotte M. Deane. Observed antibody space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Science*, 31(1):141–146, 2022. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4205.

Tobias H. Olsen, Iain H. Moal, and Charlotte M. Deane. Addressing the antibody germline bias and its effect on language models for improved antibody design. *Bioinformatics*, 40(11):btae618, 2024. URL https://academic.oup.com/bioinformatics/article/40/11/btae618/7845256.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model, 2024. URL https://arxiv.org/abs/2305.18290.

Martin Steinegger and Johannes Söding. Clustering huge protein sequence sets in linear time. *Nature Communications*, 9:2542, 2018. URL https://www.nature.com/articles/s41467-018-04964-5.

Mohammad Talaei, Kevin C. Walker, et al. Cdr-aware masked language models for paired antibodies enable state-of-the-art binding prediction. *bioRxiv*, 2025. URL https://www.biorxiv.org/content/10.1101/2025.10.31.685149v1.full.pdf.

Talal Widatalla, Ashir A. Borah, Samuel H. King, Claudia L. Driscoll, Rafael Rafailov, and Brian L. Hie. Aligning protein-generative models to experimental fitness with ProteinDPO. *Nature Methods*, 23(9):1805–1813, 2026. doi: 10.1038/s41592-026-03137-3. URL https://doi.org/10.1038/s41592-026-03137-3.

A Dataset Statistics and Corpus Construction

Figure 5 summarizes the datasets used in this work. This section provides detailed yield statistics and processing protocols for each dataset.

A.1 C05 DMS Benchmark Libraries

The downstream benchmark consists of two independently sorted libraries for the C05 antibody, varying the same 24-residue CDR-H3 loop: **ED3** (275,354 variants) and **ED5** (428,821 variants). ED3 features a median edit distance of 3 from wild-type (WT) (spanning 0 to 5 mutations; 99% mass between 2 and 4), while ED5 features a median edit distance of 5 (spanning 0 to 8 mutations; 91% mass between 4 and 5) (Figure 5d-e). Their enrichment distributions differ accordingly, with ED3 concentrated below its wild type and ED5 spread across it (Figure 5b-c). Table 2 summarizes both libraries.

Table 2: Overview of the two C05 DMS libraries.

Library	Total Variants	WT Enrichment Score	% Variants Above WT
ED3	275,354	$+5.19$	10%
ED5	428,821	-0.03	52%

A set of 2,478 variants belong to both libraries. On this shared subset, enrichment scores exhibit a Spearman rank correlation of 0.215. This confirms that enrichment is a relative, within-library quantity tied to a specific sorting competitor pool, not an absolute physical scale. Models are trained

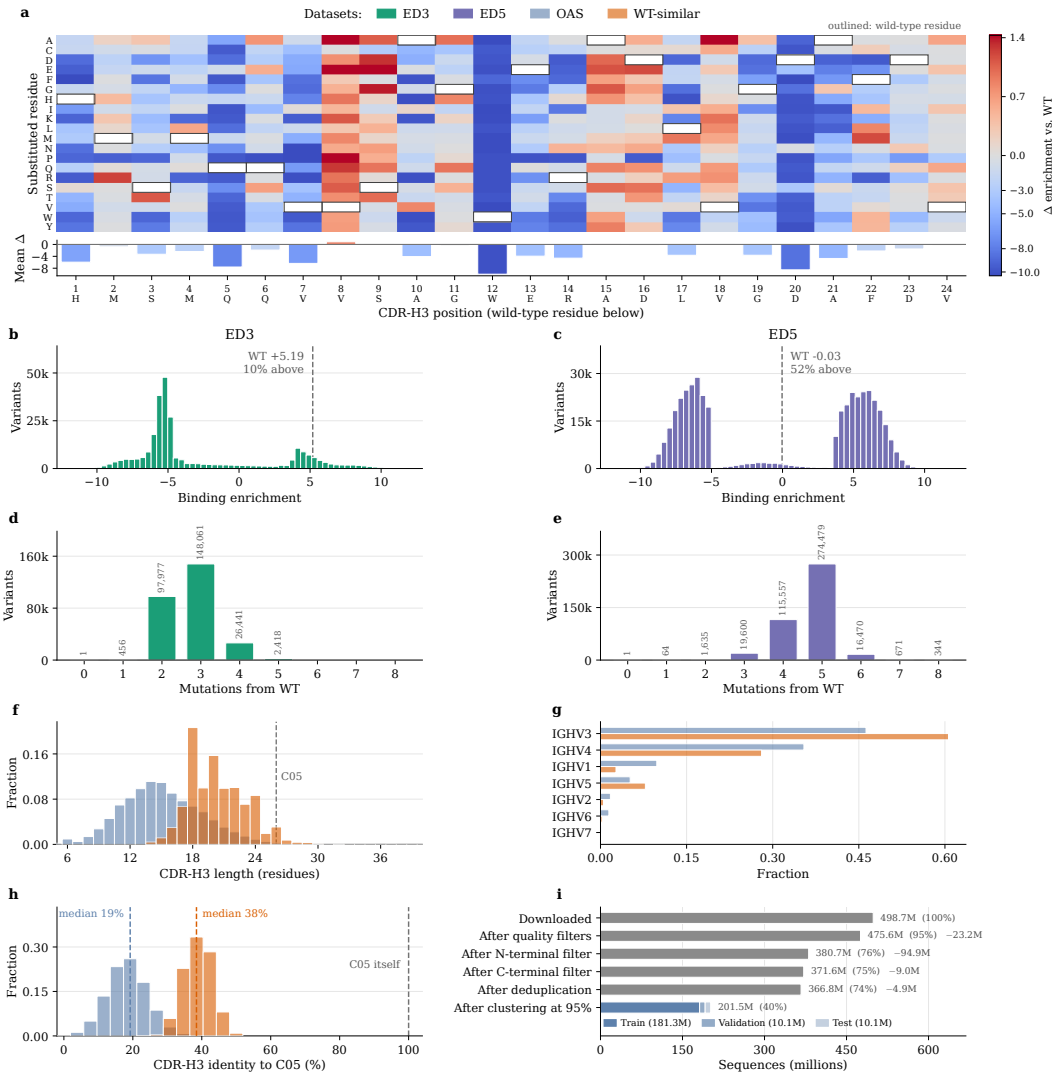


Figure 5: **Characteristics of the unlabelled pretraining corpora and downstream C05 DMS benchmark libraries.** **(a)** Single-amino-acid substitution landscape showing mean enrichment scores (Δ) relative to wild-type across all 24 positions of C05 CDR-H3. Outlined boxes indicate wild-type amino acids. **(b-c)** Distribution of binding enrichment scores for downstream DMS libraries ED3 ($N = 275{,}354$) and ED5 ($N = 428{,}821$). Dashed vertical lines indicate wild-type enrichment score. **(d-e)** Distribution of edit distance (Hamming distance from C05 WT) across sequence variants in ED3 and ED5. **(f-h)** Comparative distributions between the broad *Unlabelled OAS Corpus* (blue) and target-restricted *WT-Similar Corpus* (orange) across: **(f)** CDR-H3 length distribution, **(g)** heavy-chain IGHV gene family proportions, and **(h)** sequence identity to C05 WT CDR-H3. Vertical dashed lines denote C05 WT values. **(i)** Sequence filtering and deduplication funnel from raw downloaded OAS sequences to final train, validation, and test splits.

and selected exclusively on ED3, while ED5 is retained as an independent held-out library. Before evaluation, ED5 is filtered to retain only high-precision variants (SE ≤ 0.5) and to remove any overlap with ED3 training or validation splits. While this quality filter ensures clean, reliable labels, it also shifts the dataset composition, it removes most variants with more than five mutations and reduces the overall proportion of sequences scoring above wild type, from 52% to 27%. All reported ED5 evaluations use this final high-confidence benchmark. The single-substitution scan covers all amino acid mutations across the CDR-H3 loop (Figure 5a). Mutational tolerance varies a lot along the sequence: a few key residues (such as W12 and D20) are strictly intolerant to substitution, whereas flexible positions (such as V8) accept nearly any amino acid without disrupting binding.

A.2 Unlabelled OAS Corpus

Starting from 498,749,381 raw human heavy-chain sequences from the Observed Antibody Space (OAS) database [Olsen et al., 2022], filtering yields 201,491,665 unique sequences (40.4% yield). The pipeline applies the following sequential stages:

1. **Quality Filtering:** Removes non-productive sequences, missing conserved Cys residues anchoring the Ig-fold, sequences shorter than expected, or missing CDR-H3 annotations (23,197,539 removed).
2. **Fragmentation Filters:** Removes severe N-terminal (>16 missing residues; 94,884,285 removed) and C-terminal (>7 missing residues; 9,018,014 removed) truncations.
3. **Deduplication & Clustering:** Removes exact duplicate sequences (4,856,911 removed) and clusters the remaining pool at 95% sequence identity using Linclust [Steinegger and Söding, 2018], selecting one representative per cluster (165,300,967 removed).

The resulting pool is split 90/5/5 into 181,343,223 training, 10,074,328 validation, and 10,074,114 test sequences (Table 3, Figure 5i).

Table 3: Unlabelled OAS Corpus yield at each filtering stage.

Stage	Sequence Count	% of Initial
Downloaded Raw Sequences	498,749,381	100%
Quality Filtered	475,551,842	95.3%
N-terminal Filtered	380,667,557	76.3%
C-terminal Filtered	371,649,543	74.5%
Exact Deduplication	366,792,632	73.5%
Linclust 95% Clustering	201,491,665	40.4%
Train Split	181,343,223	36.4%
Validation Split	10,074,328	2.0%
Test Split	10,074,114	2.0%

A.3 WT-Similar Corpus

To build the WT-Similar Corpus, candidate CDR-H3 sequences from OAS are scored against C05 WT CDR-H3 using a normalized BLOSUM62 alignment score. Taking the top $K = 5{,}000$ ranked sequences establishes a minimum alignment score threshold of 0.261. These 5,000 sequences carry 2,082 distinct CDR-H3 loops and come from 266 studies in OAS. Similarity is scored under the IMGT CDR-H3 annotation used by OAS, which starts immediately after the conserved cysteine and includes the two V-gene residues preceding the mutated loop. C05's CDR-H3 is therefore 26 residues under this annotation, against the 24 residues varied by the DMS libraries. Selecting for similarity shifts the median sequence identity to C05 from 19% (unlabelled OAS corpus baseline) to 38% (Figure 5h). The closest OAS CDR-H3 reaches 50% identity (13 of 26 positions), and similarity selection raises the median but not that ceiling. Partitioning by parent OAS split yields 4,549 training, 228 validation, and 223 test sequences.

Selecting on CDR-H3 similarity also constrains properties the selection does not score directly (Figure 5f-g). Because C05 carries an unusually long loop, sequences resembling it are drawn from the long tail of the corpus, and the median CDR-H3 length rises from 15 to 20 residues

(Figure 5f). V-gene usage shifts over the same selection toward the IGHV3 family, from 46% to 61% (Figure 5g). Scoring the hypervariable loop alone is therefore sufficient to select a compatible germline background.

B Pretraining Protocol and Masking Definitions

B.1 Masking Policies and Corruption Rules

All sequences are truncated to 254 residues plus standard special tokens (`[CLS]`, `[SEP]`). Non-standard residues (`X`) are never selected for masking. Standard BERT token corruption (the 80/10/10 rule) replaces a selected position with: (i) the `[MASK]` token with 80% probability, (ii) a random standard amino acid with 10% probability, or (iii) the unchanged original residue with 10% probability.

The four masking policies operate as follows:

- **Whole-Chain (WC-15%):** Every position in the sequence is eligible. Each eligible position is selected independently with probability $p = 0.15$ and corrupted via the 80/10/10 rule.
- **CDR-Focused (CDR-50%):** Eligibility is restricted strictly to CDR-H3 residues. Each CDR-H3 position is selected independently with probability $p = 0.50$ and corrupted via the 80/10/10 rule.
- **Hybrid:** Within each training batch of size B, exactly $\mathrm{round}(0.8B)$ sequences follow CDR-50% masking and $0.2B$ sequences follow WC-15% masking. Corruptions apply the 80/10/10 rule.
- **Single-Position:** Eligibility spans the CDR-H3 loop, 5-residue flanking regions on each side, and a 30% random sample of framework positions (resampled per occurrence). Exactly *one* position is selected uniformly at random per forward pass and replaced directly with `[MASK]` (omitting 80/10/10 random corruption to match scoring conditionals).

Supervision Rates and Framework Preservation. Across ESM2-35M pretraining, WC-15% masks an average of 16.6 tokens/sequence, CDR-50% masks 7.6 tokens/sequence (all in CDR-H3), and Hybrid masks 9.4 tokens/sequence. In Hybrid masking, 80% of sequences contribute $0.8 \times 7.6 = 6.0$ CDR-H3 positions and 20% contribute $0.2 \times 16.6 = 3.3$ total positions (including 0.5 CDR-H3 positions). Thus, Hybrid supplies 86% of CDR-50%'s CDR-H3 supervision while maintaining a steady stream of 2.9 masked positions per sequence outside CDR-H3, sufficient to prevent framework degradation.

B.2 Training Schedule and Optimization

Pretraining optimization follows AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 0.01) with bf16 precision. Effective batch size is 512 (64 sequences/step $\times$ 8 gradient accumulation steps). Learning rate uses linear warm-up over 5% of total steps followed by linear decay to 0.

For the Unlabelled OAS Corpus, one epoch corresponds to 354,185 optimizer steps. For the WT-Similar Corpus (4,549 training sequences), one epoch is 8 optimizer steps, and runs train to a 100-epoch ceiling (800 steps). Figures 3 and 4 show the first 50 epochs, which contain every selected checkpoint.

B.3 Checkpoint Selection Rule

Checkpoints are evaluated on a fixed sample of 2,000 held-out OAS validation sequences. Token recovery accuracy is measured independently for CDR-H3 ($\mathrm{CDR}(t)$) and framework regions ($\mathrm{FR}(t)$) by enumerating all positions under single-token masking. Enumeration costs one forward pass per position, roughly 220,000 per evaluation point, and the sample is capped at 2,000 sequences to keep the full evaluation grid affordable.

Evaluation points sit at fixed fractions of an epoch, spaced densely over the early part of training where framework recovery does most of its moving. Phases P_1 and P_2 use $\{0.01, 0.1, 0.5, 1, 2, 3, 5, 7, 10, 15, 20, 30, 50, 70, 100\}\%$ of an epoch; the P_0 warm-up uses a denser 29-point grid that adds a point every 5% past the 20% mark. On the WT-Similar Corpus one epoch is

8 optimizer steps, so the grid resolves to five distinct points per epoch and repeats across every epoch of the run.

Let $\mathrm{FR}(\mathrm{ref})$ denote the framework accuracy of the reference checkpoint (P_0 warm-up for adaptation branches, or the P_1 switch point for single-position continuations). Checkpoint t is eligible if:

$$\mathrm{FR}(\mathrm{ref}) - \mathrm{FR}(t) \leq 0.1 \text{ percentage points.} \tag{6}$$

We select the earliest eligible checkpoint t maximizing $\mathrm{CDR}(t)$. Per-region recovery metrics for ESM2-8M are presented in Figure 6.

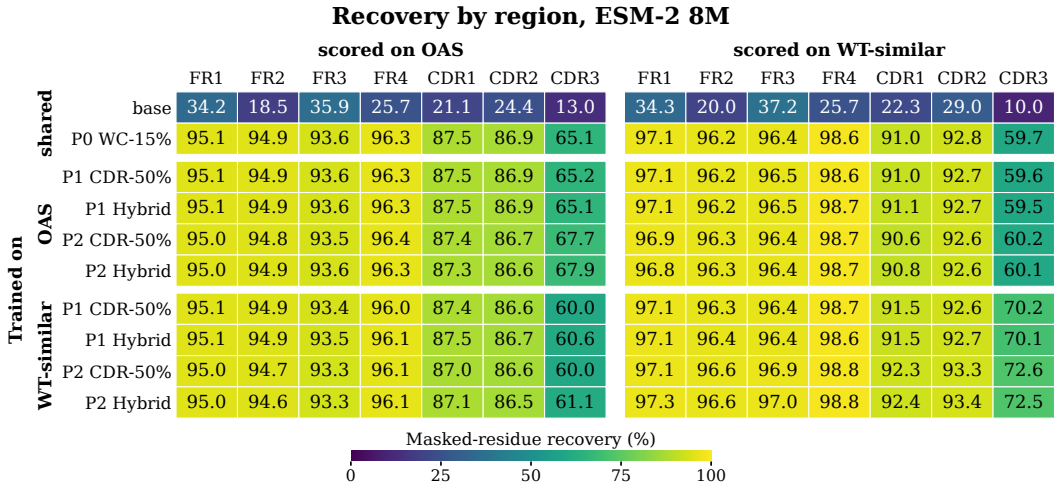


Figure 6: Per-region token recovery across pretraining phases for ESM2-8M. Layout and evaluation protocols match Figure 3d, which reports the same measurement for ESM2-35M.

Warm-up Branch Point. The P_0 warm-up has no preceding phase to be measured against, so the rule above does not apply to it. Framework and CDR-H3 recovery rise together throughout P_0, and the branch point is chosen on compute. We take the earliest checkpoint whose framework recovery stays within 0.5 percentage points of the accuracy reached at the end of the epoch, scanning backwards from the final point so that a single early crossing does not qualify. This selects 50% of the epoch for ESM2-35M (94.27% framework recovery, against 94.75% at the end) and 35% for ESM2-8M (93.41%, against 93.86%).

C Learning-Rate Hyperparameter Sweep

We sweep learning rate over $\{5 \times 10^{-6}, 1 \times 10^{-5}, 2 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}\}$ across Phase 0 (P_0) pretraining using a 10M-sequence OAS training subsample (3 epochs). Token recovery improves monotonically with learning rate for both backbones with no turnover (Table 4).

Table 4: Validation token recovery (%) after 3 pretraining epochs on the 10M OAS subsample across learning rates.

	ESM2-8M		ESM2-35M	
Learning Rate	CDR-H3	Framework	CDR-H3	Framework
5×10^{-6}	51.9	91.8	53.8	92.1
1×10^{-5}	53.9	92.0	56.0	92.3
2×10^{-5}	55.9	92.2	58.1	92.5
5×10^{-5}	58.3	92.5	61.3	93.0
1×10^{-4}	60.7	92.8	64.1	93.3

All five rates trace nearly the same path through the recovery plane and differ in how far along it a run travels (Figure 7). At matched framework recovery the curves agree on CDR-H3 recovery to within

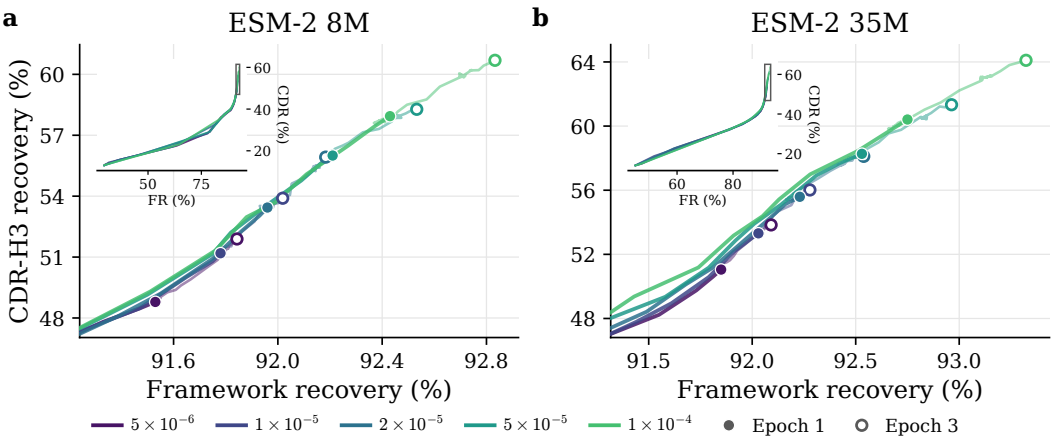


Figure 7: **Learning-rate sweep trajectories through the recovery plane.** Framework recovery against CDR-H3 recovery on held-out OAS validation sequences for **(a)** ESM2-8M and **(b)** ESM2-35M, over 3 pretraining epochs on the 10M OAS subsample, with one curve per learning rate. Filled circles mark one epoch, the budget the pipeline grants a full-OAS phase; open circles mark the end of the sweep at three epochs. Insets show the full climb from the base pretrained model, of which the main axes show the final portion.

0.6 percentage points on ESM2-8M, and on ESM2-35M their spread narrows from 1.8 percentage points early in the climb to 0.3 by the end of the sweep. The learning rate therefore sets the speed along a shared trajectory, not its shape.

Within the single epoch the pipeline grants a full-OAS phase, that speed is what separates the rates: ESM2-35M reaches 51.1% CDR-H3 recovery at 5×10^{-6} and 60.4% at 1×10^{-4}. A lower rate does not close the gap with more compute, since 2×10^{-5} after three epochs reaches 58.1% and still trails 1×10^{-4} after one. Consequently, 1×10^{-4} is fixed across all pretraining phases. Because recovery is still improving at the top of the grid, 1×10^{-4} is the best value tested and not necessarily the best available. Whether larger learning rates continue to help, and where the turnover lies, is left untested.

D Preference Optimization (DPO) Protocol

D.1 Pair Construction and Rank-Gap Sampling

Given a dataset of sequence variants with measured enrichment scores s, scores are normalized relative to wild-type: $\Delta = s - s_{\mathrm{WT}}$. For a pool of n sequences and gap parameter $g = 0.5$, variants are sorted by Δ and partitioned into consecutive blocks of size $2k$, where:

$$k = 1 + \lfloor g\,(\lfloor n/2 \rfloor - 1) \rfloor\,. \tag{7}$$

Within each block, sequence i is paired with sequence $i + k$ for $i < k$.

To cover the full binding landscape, pairs are sampled across four distinct functional categories:

- **Within-Positive (10,000 pairs):** Rank-gap pairs drawn among binder variants ($\Delta > 0$).
- **Within-Negative (3,000 pairs):** Rank-gap pairs drawn among non-binder variants ($\Delta < 0$).
- **WT Anchors (2,000 pairs):** WT paired against strong binders ($\Delta > 1.0$, WT dispreferred) and strong non-binders ($\Delta < -5.0$, WT preferred).
- **Cross-Mode (7,000 pairs):** Independent pairs drawn between one binder ($\Delta > 0$) and one non-binder ($\Delta < 0$).

A minimum score margin filter discards pairs with $|\Delta(y^w) - \Delta(y^l)| < 0.1$. This threshold, and the Δ cutoffs defining the four categories above, are absolute values on an enrichment scale specific to one library and its sorting competitor pool. All of them would need to be recalibrated to the score distribution of a new dataset.

D.2 Splits and Low-N Protocol

Sequences are clustered using Hamming distance ≤ 1 connected components prior to pairing, preventing single-substitution variants from leaking across splits. Pairs are generated independently within train, validation, and test splits.

For low-N data-scaling experiments, sequence training sets are uniformly subsampled prior to pair generation. The reported curve covers $N \in \{50, 100, 200, 500, 1000, 2000, 5000, 10{,}000, 20{,}000, 50{,}000\}$; the pretraining-recipe ablation uses the coarser grid $N \in \{0, 50, 200, 1000, 5000, 50{,}000\}$. Validation and test sets remain fixed across all N to ensure identical evaluation benchmarks. Each point on the reported curve is the mean of ten draws. The ablation uses five per arm, except the two arms it shares with the headline comparison, which carry ten; the $N = 0$ anchor is deterministic and carries one. Draws sharing a seed are nested, so the $N = 50$ subset is contained in the $N = 100$ subset.

D.3 Optimization Hyperparameters

Full-Parameter DPO. Optimized using Adam ($lr = 3 \times 10^{-6}$, no weight decay) with batch size 8 and gradient norm clipping at 1.0 in bf16 precision. Learning rate uses linear warm-up over the first 5% of steps followed by cosine decay over 50 epochs, with early stopping after 5 epochs of non-improving validation Spearman correlation.

LoRA DPO. Freezes base weights and trains Low-Rank Adapters [Hu et al., 2022] inserted into query and value projections across all layers ($r = 8, \alpha = 16$, dropout 0.05). Trained with Adam ($lr = 5 \times 10^{-4}$), batch size 10, and 4 gradient accumulation steps.

D.4 Choice of Pretraining Recipe

While Section 4.1 compares a single adapted backbone against the non-adapted baseline, this checkpoint was selected from a systematic sweep across nine candidate pretraining arms. Here, we outline the pretraining configurations and the label-free selection protocol used to identify the optimal backbone.

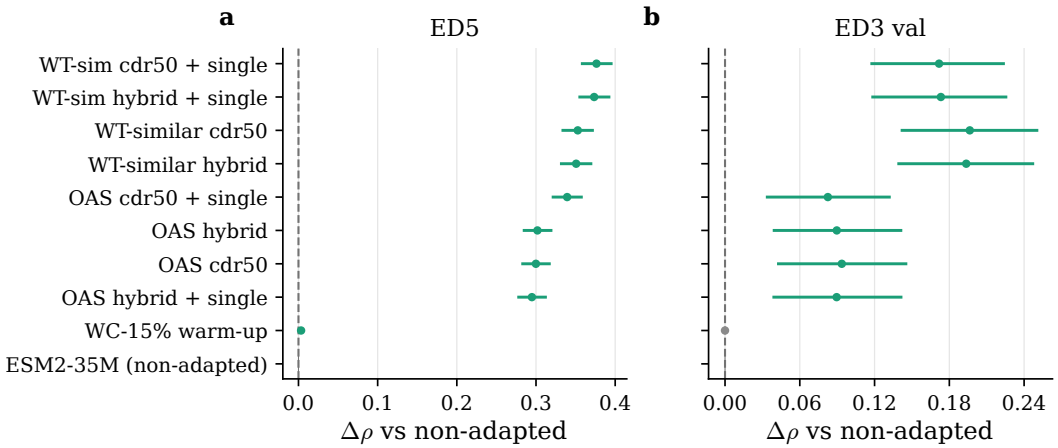


Figure 8: **Zero-shot performance across candidate pretraining arms** ($N = 0$)**.** Paired difference in Spearman rank correlation ($\Delta\rho$) relative to the non-adapted ESM2-35M backbone prior to preference alignment, evaluated with 95% bootstrap confidence intervals. All candidate arms are evaluated on identical test subsets to control for evaluation row variance. **(a)** Held-out ED5 test set. **(b)** ED3 validation split used for model selection.

Figure 8 categorizes the nine candidate arms into three distinct procedural tiers:

1. **Whole-Chain Warm-up:** Performs identically to the non-adapted baseline on the ED3 validation split, yielding negligible correlation gains on ED5.

2. **Broad-Repertoire Masked Language Modeling (MLM):** A single epoch of standard MLM without complementary determining region (CDR)-focused masking or sequence corpus filtering maintains baseline ranking behavior.
3. **CDR-Targeted and WT-Similar Pretraining:** Applying CDR-focused masking significantly decouples performance from the baseline, with the strongest gains realized when restricting the pretraining corpus to wild-type (WT)-similar sequences.

While single-position continuation training modestly improves correlation on ED5, it slightly reduces performance on the ED3 validation split. Because these differences fall within overlapping confidence intervals, single-position continuation effects remain inconclusive.

To prevent evaluation bias on downstream test sets, final model selection was executed strictly on the ED3 validation split. Under this protocol, the **WT-similar CDR-50%** arm achieved the highest validation correlation (with **WT-similar hybrid** performing within statistically indistinguishable bounds) and was selected as the representative DAP backbone for Section 4.1.

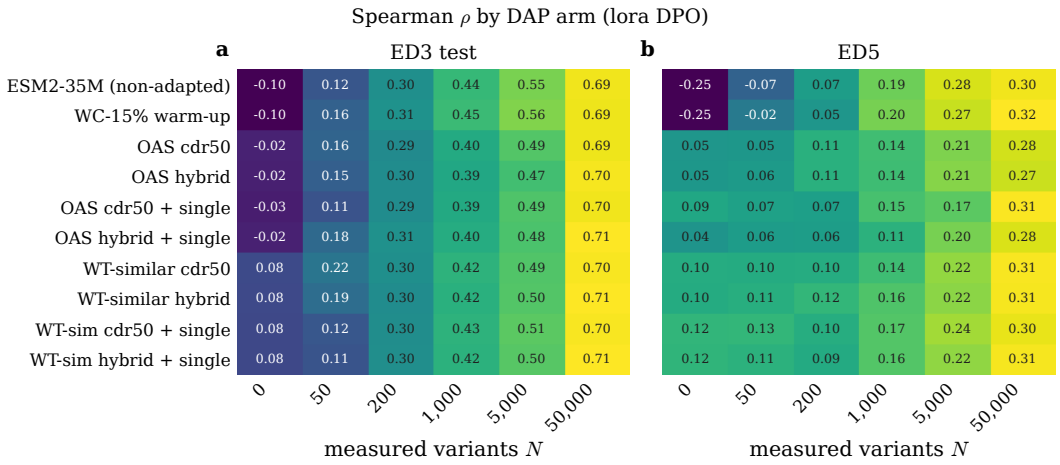


Figure 9: **Pretraining arm performance scaling across labeled training sample sizes N.** Spearman rank correlation per pretraining arm across varying N under LoRA DPO, averaged over at least five draws per cell (Appendix D). Rows reflect progression through the pretraining pipeline: non-adapted backbone, shared warm-up, individual corpus/masking configurations, and single-position continuations. Color intensity is normalized within each panel separately.

Evaluating performance across dataset sizes N (Figure 9) demonstrates that the relative ranking of pretraining arms is dynamic across scaling regimes. In data-scarce regimes ($N \to 0$), non-adapted backbones rank lowest while WT-similar arms perform best. As N expands, performance across all pretraining arms converges as experimental labels supersede pretraining priors, confirming that domain-adaptive pretraining primarily enhances initial sample efficiency.

D.5 Backbone Size

To evaluate whether low-data pretraining advantages extend across parameter scales, we replicated the benchmark evaluation using the smaller ESM2-8M backbone (Figure 10). The pretraining recipe applied to ESM2-8M was kept identical to the selected ESM2-35M configuration (WT-similar CDR-50%) to isolate model scale as the sole variable.

The ESM2-8M results reproduce all three core empirical trends observed at 35M parameters:

1. Zero-shot ($N = 0$) evaluations on the non-adapted backbone exhibit negative rank correlation.
2. Domain-adaptive pretraining resolves rank inversion and boosts low-N accuracy.
3. The advantage narrows as labelled variants accumulate, and the two backbones meet between $N = 200$ and $N = 1{,}000$ on three of the four curves; under full DPO on ED3 test the adapted backbone stays ahead across the measured grid.

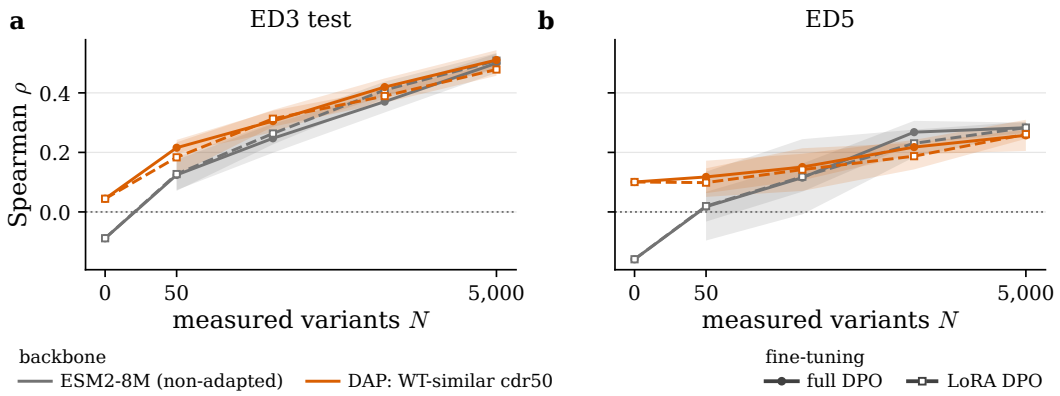


Figure 10: **Downstream binding fitness scaling on the ESM2-8M architecture.** Panels and encoding follow Figure 2, with five draws per point. The domain-adapted backbone uses the same pretraining recipe as ESM2-35M, isolating parameter scale.

The ESM2-8M curve rests on five draws per point against ten for ESM2-35M, so its bands are estimated from fewer samples and should be read as the rougher of the two.