AbEvolve: Aligning Protein Language Models with Preference Signals from Antibody Natural Selection

Anonymous authors
Paper under double-blind review

Abstract

While protein language models (PLMs) extract general biological syntax from sequence databases, they often exhibit an alignment gap in antibody engineering. This gap stems from antibodies' unique evolutionary search for optimal binding specificity—a logic that self-supervised PLMs trained on static sequence collections do not explicitly model, because the experimentally labeled data that would reveal this search directly are scarce and costly to produce. In this paper, we instead explore the large-scale, coarse redundancy signal that is readily available in unlabeled antibody sequencing data. To bridge this gap, we propose **AbEvolve**, a systematic framework that transforms the clonal abundance within antibody sequencing data—a coarse yet massive proxy for natural selection—into explicit abundance-derived evolutionary trajectories for PLM alignment training. By leveraging abundance as a proxy for natural selection, AbEvolve converts latent sequence distributions into actionable signals that help PLMs better capture maturation-associated sequence preferences. To absorb these signals, we adopt two alignment objectives—reward-weighted maximum-likelihood (RW-MLE), which reweights terminal variants by their abundance reward, and GFlowNet alignment, which propagates the scalar abundance rewards along constructed mutation paths and preserves intermediate path structure. We integrate AbEvolve with pre-trained models including ESM2 and ProteinReasoner and evaluate their zero-shot performance across representative antibody property prediction tasks. AbEvolve improves zero-shot ranking on all eight benchmarks—spanning binding, thermostability, and expression—for ESM2-GFlowNet; for ProteinReasoner, GFlowNet alignment improves binding and thermostability, whereas expression gains vary by benchmark. Together, AbEvolve offers a new supervision source and modeling pathway for addressing this alignment gap in therapeutic antibody design.

1 Introduction

Antibody therapeutics have transformed the treatment of complex diseases, including oncology and autoimmune disorders, due to their high specificity and favorable safety profiles (Weiner, 2015; Lu et al., 2020). Yet antibody design remains challenging because engineering must jointly optimize binding affinity, and other functional properties, which often impose competing biophysical constraints (Jain et al., 2017; Tharp et al., 2026). Conventional directed evolution and screening can improve antibodies, but they are labor-intensive and explore only a minute fraction of the enormous antibody sequence space, estimated to exceed 10^{60} variants (Arnold, 2018; Dryden et al., 2008). This sparsity of functional variants motivates computational frameworks that leverage latent evolutionary signals to guide efficient antibody design (Tran et al., 2026; Shakked et al., 2022).

The recent rise of protein language models (PLMs) has introduced promising avenues for *in silico* protein design (Lin et al., 2023; Hie et al., 2024; Kenlay et al., 2024; Ruffolo et al., 2021). By pre-training on massive datasets of sequences and structures, models such as ESM2 (Lin et al., 2023) and ProteinReasoner (Liu et al., 2025) have demonstrated a remarkable capacity to capture the underlying "grammar" and biophysical constraints of protein sequences. Yet, despite their success in grasping general biological syntax, these models often suffer from a fundamental alignment gap when applied to antibody engineering (Meier et al., 2021; Widatalla et al., 2026). While

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

PLMs capture the broad evolutionary plausibility shaped over geological time (Lin et al., 2023; Liu et al., 2025; Hie et al., 2024; Hayes et al., 2025), therapeutic optimization requires navigating a distinct fitness landscape driven by the rapid, antigen-specific selection pressures of the immune system (Mesin et al., 2016; Victora & Nussenzweig, 2022). A fundamental limitation is the gap between global evolutionary priors and the local fitness signals necessary for trait-specific refinement. The experimentally labeled antibody data that would directly expose these local signals are scarce and costly to produce, whereas unlabeled repertoire sequences are abundant. Current PLMs, including antibody specific PLMs are typically trained on large, static collections of natural protein sequences using self-supervised learning paradigms, such as maximum-likelihood language-modeling objectives, thereby capturing broad evolutionary plausibility rather than explicitly modeling task-specific selection trajectories (Rives et al., 2021; Lin et al., 2023; Elnaggar et al., 2021; Meier et al., 2021). This creates an alignment gap between the rules learned from self-supervised sequence distributions and the functional objectives of interest (Widatalla et al., 2026). Furthermore, antibody evolution is inherently lineage-dependent: selection acts on mutational trajectories rather than isolated sequences (Hie et al., 2024; Burnett et al., 2018). As illustrated in Figure 1, within an antibody lineage, highly abundant clones often reflect clonal expansion under antigen-driven selection, suggesting that abundance can be interpreted as an observable proxy for relative survival and proliferative advantage, and hence for lineage-level fitness (Merkenschlager et al., 2025; Victora & Mesin, 2014; Mesin et al., 2016). We treat this signal as a coarse, repertoire-proximal surrogate for selection rather than a validated experimental fitness assay. This motivates a lineage-aware view of antibody optimization, where intra-lineage abundance provides local evolutionary signals that complement the global priors learned by PLMs.

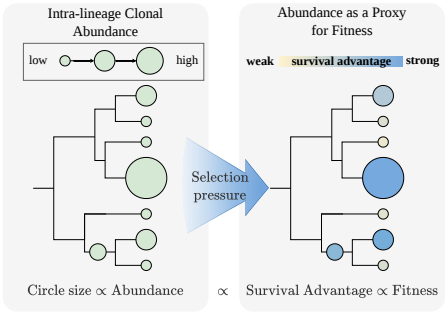

Figure 1: Lineage-aware fitness signal. Intra-lineage clonal abundance reflects selection-driven expansion, providing a coarse proxy for survival advantage, proliferative advantage, and local fitness—not a validated experimental assay.

We introduce AbEvolve (Figure 2), a model-agnostic framework that extracts evolutionary supervision from antibody repertoires without requiring experimental labels. By interpreting clonal abundance in the Observed Antibody Space (OAS) (Olsen et al., 2022a)—a coarse but massive signal in unlabeled antibody sequencing data—as a biological "breadcrumb trail" of natural selection, AbEvolve constructs abundance-labeled evolutionary trajectories from sequence clusters, allowing models to learn the mutational logic of antibody maturation. This approach converts unstructured repertoire data into supervised-like signals for trait refinement.

We evaluate AbEvolve's versatility by aligning diverse pre-trained architectures—sequence-based ESM2 (Lin et al., 2023) and multi-modal ProteinReasoner (Liu et al., 2025)—under trajectory-aware GFlowNet alignment (Bengio et al., 2021; 2023) on both backbones, and additionally endpoint reward-weighted maximum-likelihood (RW-MLE) on ProteinReasoner, whose native autoregressive likelihood supports seed-conditioned terminal scoring. Across multiple antibody design benchmarks, AbEvolve's clonal abundance signals show strong transferability across architectures, e.g., raising Spearman's ρ on SKEMPI-V2 (0.075 vs. 0.214) on ESM2 and (0.019 vs. 0.257) on ProteinReasoner.

The primary contributions of this work are summarized as follows:

- **Repurposing Clonal Abundance as Evolutionary Supervision**: We propose a model-agnostic framework that converts coarse clonal abundance in OAS—a massive, repertoire-proximal proxy for natural selection—into explicit abundance-derived *evolutionary trajectories*, repurposing unlabeled repertoires into structured supervision for maturation-associated PLM alignment without target-specific labels.
- **Trajectory-aware Alignment across Diverse Paradigms**: We introduce RW-MLE and GFlowNet alignment as complementary paradigms for integrating clonal abundance into pre-trained PLMs. While RW-MLE reweights terminal variants in isolation, GFlowNet alignment

propagates scalar abundance rewards along constructed mutation paths and preserves tree-like mutational structure, enabling more principled use of clonal-abundance signals.
- **Enhanced Zero-shot Antibody Evaluation**: We demonstrate that AbEvolve improves zero-shot ranking on all eight benchmarks—spanning binding, thermostability, and expression—for ESM2, and on binding and thermostability for ProteinReasoner over the corresponding vanilla backbones.

2 Related Work

The synergy between protein language modeling and functional alignment has redefined the boundaries of computational antibody engineering. Our work situates itself at the intersection of evolutionary representation, zero-shot fitness landscapes, and generative alignment paradigms.

Protein Language Models for Antibody Fitness Prediction. General protein language models such as ESM2 (Lin et al., 2023) and ProteinReasoner (Liu et al., 2025) are pre-trained on large, diverse protein corpora and capture broad evolutionary syntax; they can also be applied to antibody engineering without antigen-specific supervision (Hie et al., 2024). Yet model likelihood is not equivalent to *design fitness* (Widatalla et al., 2026; Hie et al., 2024), a gap especially pronounced in antibodies, whose fitness is shaped by somatic hypermutation, clonal selection, and hypervariable mutational signals (Mesin et al., 2016; Burnett et al., 2018; Victora & Nussenzweig, 2022; Merkenschlager et al., 2025). Antibody-specific models instead pre-train directly on repertoire data such as OAS (Olsen et al., 2022a), including AntiBERTy (Ruffolo et al., 2021), IgBert (Kenlay et al., 2024), AbLang (Olsen et al., 2022b), and AbLang2 (Olsen et al., 2024). These self-supervised antibody PLMs nevertheless treat repertoires largely as static independent sequences, underusing clonal abundance. AbEvolve converts this repertoire-level information into evolutionary alignment signals, refining PLM representations toward abundance-derived maturation preferences rather than sequence likelihood alone.

Preference Alignment and Fitness Alignment for Protein Models. Aligning protein models with functional objectives bridges sequence likelihood and biological fitness. Supervised fine-tuning on experimental labels (Rao et al., 2019; Meier et al., 2021; Notin et al., 2023) is effective but requires task-specific assays, generalizes poorly across targets, and is limited by scarce binding-affinity data, particularly for antibodies. Preference-based alignment relaxes the need for calibrated scalar labels by learning from relative fitness comparisons (Rafailov et al., 2023); for example, ProteinDPO (Widatalla et al., 2026) aligns structure-conditioned PLMs with stability preferences.

A complementary line of work aligns PLMs using maturation structure inferred from repertoires rather than experimental labels. CoSINE (Lu et al., 2026) post-trains a continuous-time Markov chain on reconstructed clonal trees with parent–child branch lengths, modeling affinity maturation along inferred tree edges using curated parent–child pairs from phylogenetic pipelines in the DASM tradition (Matsen IV et al., 2026). AbEvolve also requires a non-trivial construction pipeline—sequence restoration, clonal clustering, and abundance-derived pseudo-trajectories—but it is more permissive about data sources. Whereas CoSINE depends on curated BCR cohorts and inferred parent–child pairs from clonal phylogenies, AbEvolve trains from bulk OAS repertoires with clonal abundance counts as a model-agnostic alignment pipeline, enabling broader reuse of heterogeneous public sequencing data.

GFlowNets for Biological Sequence Optimization. Generative Flow Networks (GFlowNets) learn distributions over discrete objects proportional to a user-defined reward (Bengio et al., 2021; 2023). Unlike reward-maximizing RL, they model diverse high-reward candidates, suiting large, multimodal, sparsely enriched biological sequence spaces (Malkin et al., 2022; Pandey et al., 2025) and motivating applications in protein and nucleic-acid design, active learning, and multi-objective optimization (Jain et al., 2023; Ghari et al., 2024). AbEvolve instead uses GFlowNets not as a sequence generator but as alignment method for PLMs: rewards are derived from clonal abundance, and models are trained on abundance-informed mutational trajectories.

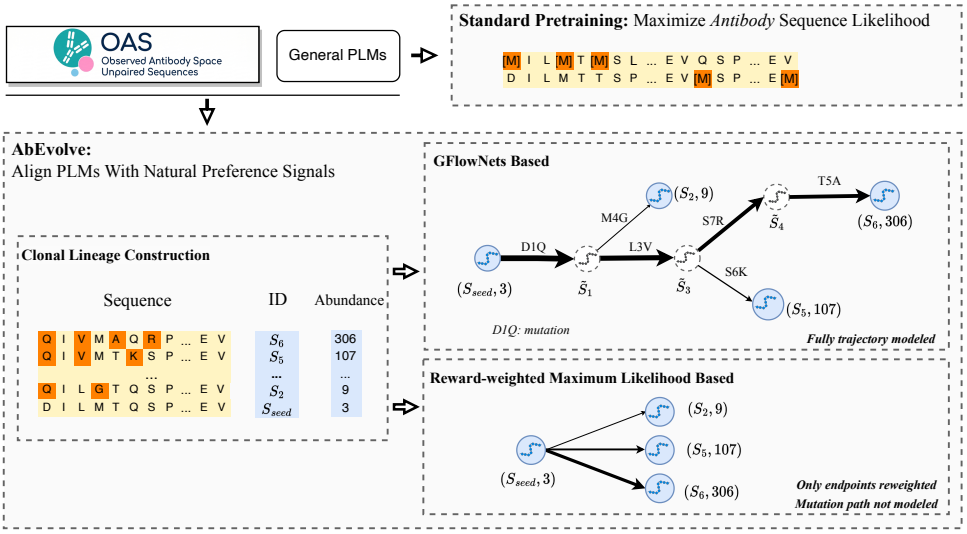

Figure 2: **AbEvolve framework overview.** Standard pre-training (Top) learns from static collections of repertoire sequences, without explicitly capturing lineage-level selection or mutational accessibility. AbEvolve (Bottom) derives abundance-informed evolutionary trajectories from raw repertoires and aligns pre-trained models via RW-MLE and GFlowNet alignment, converting clonal expansion signals into supervision for property-aware representation learning.

3 Methodology

AbEvolve aligns pre-trained protein models with abundance-derived selection signals by converting static antibody repertoires into directed evolutionary trajectories. Unlike general protein language models such as ESM2 (Lin et al., 2023), which primarily learn from self-supervised objectives without explicit evolutionary supervision, AbEvolve models abundance-informed transitions across antibody maturation process. We use clonal abundance from the unpaired subset of OAS (Olsen et al., 2022a) as a proxy for selection and evolutionary success. As shown in Figure 2, AbEvolve consists of two modules:

1. **Evolutionary Trajectory Construction**: We construct lineage-specific mutational trajectories from abundance-labeled antibody repertoires. These trajectories encode abundance-ordered mutation patterns that provide coarse, lineage-level supervision—relative clonal expansion within each cluster, rather than calibrated experimental property labels (see Appendix A.1 for background on antibody structure and lineage concepts).
2. **Trajectory Alignment Paradigms**: We integrate these signals into pre-trained models through RW-MLE, which reweights terminal variants by abundance in each lineage, and GFlowNet alignment, which propagates scalar abundance rewards along constructed mutation trajectories.

3.1 Problem Formulation

Let $\mathcal{V}$ denote the 20-amino-acid alphabet. An antibody is represented as a single-chain sequence $\mathbf{s} \in \mathcal{V}^L$. Each antibody has latent biophysical traits $f(\mathbf{s}) \in \mathbb{R}^k$, such as affinity, stability, and expression. These traits are typically unobserved during training; AbEvolve instead uses clonal abundance $A(\mathbf{s}) \in \mathbb{Z}^+$ as an abundance-derived selection signal: higher abundance reflects stronger survival and expansion during antibody maturation (Bashford-Rogers et al., 2013; Victora & Nussenzweig, 2022; Merkenschlager et al., 2025).

We model maturation as an abundance-informed mutational trajectory $\tau_i = (\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(K_i)})$, where each consecutive transition is a single-point mutational step. Intermediate states $\mathbf{s}^{(t)}$ for $t = 1, \ldots, K_i - 1$ may be observed lineage sequences or FGMO-constructed virtual nodes $\hat{\mathbf{s}}^{(t)}$ when the corresponding intermediate is absent from the repertoire (Figure 2). Given a pre-trained model M_θ, AbEvolve converts each constructed trajectory τ_i with abundance labels A into objective-specific supervision. RW-MLE reweights terminal variants by their abundance reward, while GFlowNet alignment exploits the full mutation path, propagating the endpoint abundance reward along reachable intermediate states. We optimize

$$\min_\theta \mathcal{L}_{\text{align}}(\theta; \tau_i, A),$$

where $\mathcal{L}_{\text{align}}$ is instantiated by the corresponding RW-MLE or GFlowNet alignment objective.

3.2 Evolutionary Trajectory Construction

We hypothesize that antibody sequence abundance records the cumulative outcome of successful selection (Molari et al., 2021; Merkenschlager et al., 2025; Ralph et al., 2026). To extract such evolutionary signals, we use the unpaired subset of OAS (Olsen et al., 2022a) as the repertoire source and construct abundance-labeled clonal lineages as follows.

Antibody sequence restoration. Raw OAS sequences often contain N- or C-terminal truncations due to primer constraints (Yaari & Kleinstein, 2015; Friedensohn et al., 2017; Robins, 2013). To reduce the impact of sequencing artifacts, we remove sequences with abundance lower than 3. We then restore full-length variable regions using a germline-anchored completion strategy: V/J-gene metadata are used to retrieve matched templates from IMGT (Lefranc et al., 2015), which anchor missing terminal segments (Appendix A.2).

Clonal lineage identification and clustering. Within each donor repertoire, sequences are first grouped by shared V/J genes and CDR3 length (Gupta et al., 2015; Yaari & Kleinstein, 2015). We then apply agglomerative hierarchical clustering (Müllner, 2011) using Levenshtein distance (Levenshtein et al., 1966) with threshold $d = 6$, which captures the mutational breadth of antibody maturation within a lineage (Appendix Figure S2) (Nouri & Kleinstein, 2018). A strict length-uniformity constraint is imposed so that transitions can be represented as amino-acid substitutions.

This pipeline yields about 260K discrete clonal lineages $\{\mathcal{C}_i\}$, comprising roughly 19 million sequences. Within each lineage, sequences are length-standardized and annotated with clonal abundances $\{A(\mathbf{s}) \mid \mathbf{s} \in \mathcal{C}_i\}$, averaging over 70 sequences per lineage (see Appendix A.5.1 for details).

Trajectory construction. We convert each static clonal lineage $\mathcal{C}$ into abundance-informed pseudo-trajectories that orient mutations from low-abundance variants toward clonally expanded sequences. Rather than constructing germline-centered phylogenies, our goal is to construct selection-consistent training signals. Although abundance does not uniquely determine temporal ancestry, it provides a useful ordering signal for relative clonal success. For each lineage, the lowest-abundance sequence is selected as the computational seed:

$$\mathbf{s}_{\text{seed}} = \arg\min_{\mathbf{s} \in \mathcal{C}} A(\mathbf{s}),$$

and the remaining sequences form the terminal sequence set $\mathcal{S}_{\text{terminal}} = \mathcal{C} \setminus \{\mathbf{s}_{\text{seed}}\}$. Since sequences within $\mathcal{C}$ are length-standardized, the difference between $\mathbf{s}_{\text{seed}}$ and each terminal sequence $\mathbf{s}_i$ is decomposed into single-point mutations $M_i = \{m_{i,1}, \ldots, m_{i,K_i}\}$.

Each terminal sequence carries an abundance-shift reward relative to the seed,

$$R(\tau_i) = \log A(\mathbf{s}_i) - \log A(\mathbf{s}_{\text{seed}}), \tag{1}$$

which measures the clonal-expansion signal exploited by AbEvolve's alignment objectives (Sections 3.3.1 and 3.3.2).

3.3 Trajectory Alignment across Learning Paradigms

We test whether abundance-informed trajectories provide useful supervision across multiple learning paradigms. Our hypothesis is that learning from clonal-expansion directionality, rather than only

static sequence–label pairs, helps models capture selection-consistent mutational structure more robustly. We instantiate two alignment paradigms—RW-MLE, which reweights terminal variants by their abundance reward, and GFlowNet alignment, which preserves the full mutational trajectory. Both paradigms consume identical trajectories and rewards, differing only in whether intermediate path structure is exploited; we compare them on ProteinReasoner in Section 4.

3.3.1 Abundance-Guided Endpoint Alignment via RW-MLE

As a lightweight baseline, we first integrate abundance supervision at the trajectory endpoints. RW-MLE reweights the model's terminal log-likelihood by the abundance-shift reward of Eq. equation 1, so that variants favored by clonal expansion contribute more to training. Given a lineage $\mathcal{C}$ with terminal set $\mathcal{S}_{\text{terminal}} = \{\mathbf{s}_i\}$, AbEvolve optimizes

$$\mathcal{L}_{\text{RW-MLE}}(\theta) = -\mathbb{E}_{\mathbf{s}\sim\mathcal{S}_{\text{terminal}}}\left[\,\bar{R}(\mathbf{s})\,\log \pi_\theta(\mathbf{s} \mid \mathbf{s}_{\text{seed}})\right], \tag{2}$$

where $\pi_\theta(\cdot \mid \mathbf{s}_{\text{seed}})$ is the pre-trained model's sequence-level structure–MSA-conditioned log-likelihood of the terminal variant under the seed-anchored context (Appendix A.4), and $\bar{R}(\mathbf{s})$ is a strictly positive, within-lineage normalization of $R(\cdot)$, e.g., the softmax $\bar{R}(\mathbf{s}_i) = \exp(R(\tau_i))/\sum_{\mathbf{s}_j}\exp(R(\tau_j))$. Because $\mathbf{s}_{\text{seed}}$ has the minimum abundance within the lineage, $R(\tau_i) \geq 0$ for every terminal. RW-MLE keeps the standard maximum-likelihood target but reweights terminal variants by $\bar{R}(\mathbf{s})$, skewing learning toward clonally expanded sequences while treating each endpoint independently. Further implementation details are in Appendix A.3.1.

3.3.2 Abundance-Guided Trajectory Alignment via GFlowNet

Frequency-Guided Mutation Ordering (FGMO). To construct a selection-informed pseudo-order over mutations, we introduce Frequency-Guided Mutation Ordering. For each mutation m, we compute its lineage-level occurrence frequency

$$f(m) = \sum_{\mathbf{s}_i\in\mathcal{S}_{\text{terminal}}} \mathbb{I}(m \in M_i),$$

and order mutations in each M_i by decreasing $f(m)$. This produces a path $\tau_i = (\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(K_i)})$ for each terminal $\mathbf{s}_i$, where $\mathbf{s}^{(0)} = \mathbf{s}_{\text{seed}}$ and $\mathbf{s}^{(K_i)} = \mathbf{s}_i$. For $t = 1, \ldots, K_i - 1$, each intermediate $\mathbf{s}^{(t)}$ is either an observed sequence in $\mathcal{C}$ or a virtual node $\hat{\mathbf{s}}^{(t)}$ constructed by FGMO when that intermediate is not present in the repertoire; frequent mutations appear earlier along each path. As illustrated in Figure 2, shared observed and virtual intermediates merge these paths into a directed mutation tree over reachable states, revealing a lineage-level evolutionary backbone while reducing noise from rare mutations.

We use GFlowNet alignment to model this tree-structured set of trajectories, using the abundance-shift reward $R(\tau_i)$ from Eq. equation 1—globally min–max normalized to $(0, 1]$ across training trajectories so that $\log R(\tau_i)$ in the Trajectory Balance objective below remains well defined—rather than the within-lineage softmax-normalized endpoint weights $\bar{R}(\mathbf{s})$ used by RW-MLE. Given $\tau_i = (\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(K_i)})$, GFlowNet alignment learns transition policies whose terminal probabilities are proportional to $R(\tau_i)$; virtual nodes $\hat{\mathbf{s}}^{(t)}$ carry no abundance label and receive reward only as it is propagated backward by the learned policies. Unlike static sequence-level training, this objective preserves intermediate mutational structure and encourages the model to capture how abundance-derived selection signals accumulate along accessible evolutionary paths. We optimize the Trajectory Balance (TB) objective (Malkin et al., 2022):

$$\mathcal{L}_{\text{TB}}(\tau_i) = \left(\log \frac{Z_\theta \prod_{t=1}^{K_i} P_F\left(\mathbf{s}^{(t)} \mid \mathbf{s}^{(t-1)}; \theta\right)}{R(\tau_i)\prod_{t=1}^{K_i} P_B\left(\mathbf{s}^{(t-1)} \mid \mathbf{s}^{(t)}; \theta\right)}\right)^2, \tag{3}$$

where $\mathbf{s}^{(0)} = \mathbf{s}_{\text{seed}}$, $\mathbf{s}^{(K_i)} = \mathbf{s}_i$, and intermediate states $\mathbf{s}^{(t)}$ for $t = 1, \ldots, K_i - 1$ are observed lineage sequences or virtual nodes $\hat{\mathbf{s}}^{(t)}$ as in Figure 2; P_F and P_B denote forward and backward transition policies, and Z_θ is a learnable normalizing constant. Both policies are implemented as lightweight action heads over a shared pre-trained backbone. This objective preserves intermediate mutational structure and aligns transition scores with abundance-derived evolutionary signals. Details are provided in Appendix A.3.2.

4 Experiments

We evaluate AbEvolve's capacity to convert clonal abundance into abundance-derived selection signals, hypothesizing that constructed evolutionary trajectories provide richer supervision than static sequence distributions. Crucially, we focus on zero-shot generalization: without task-specific fine-tuning, AbEvolve improves ranking over vanilla backbones in a property- and architecture-dependent manner: **ESM2-GFlowNet** improves on all eight benchmarks, whereas ProteinReasoner-GFlowNet improves binding and thermostability and expression gains vary by benchmark (RW-MLE thermostability is mixed).

4.1 Experimental Setup

Base models and benchmarks. We evaluate AbEvolve on two representative pre-trained architectures: ESM2, a BERT-style PLM (Lin et al., 2023), and ProteinReasoner (PR), a GPT-style multimodal PLM (Liu et al., 2025). For each backbone, we compare the **Vanilla** pre-trained model with **AbEvolve** under trajectory alignment (GFlowNet; Section 3.3.2); for ProteinReasoner only, we additionally evaluate endpoint alignment (RW-MLE; Section 3.3.1), whose native autoregressive likelihood supports seed-conditioned terminal scoring. Table 1 summarizes the evaluated variants; main results use models aligned on the top 260K largest lineages. We assess zero-shot performance on eight antibody property evaluations spanning binding affinity, thermostability, and expression (Table 2); model scoring protocols, dataset statistics, and chain handling are in Appendix A.4, Appendix A.5, and Appendix A.6. Section 4.3 isolates continuous pre-training, abundance-label shuffling, and FGMO ordering; lineage-scale results are in Appendix A.7.

Scoring protocol. Vanilla ESM2 is scored with pseudo-log-likelihood; Vanilla ProteinReasoner and ProteinReasoner-RW-MLE use structure–MSA-conditioned likelihood. At evaluation, GFlowNet-aligned models score the log-probability of the FGMO path from the *benchmark parent* (wild-type) to the candidate under the learned forward policy—not from an OAS training seed. Details are in Appendix A.4.

Metrics. We report Spearman's ρ, which measures monotonic correlation between model scores and experimental labels.

Implementation details. Hyperparameters for each alignment strategy are provided in Appendix A.9.

Table 1: Pre-trained models used in AbEvolve. ✓ indicates the aligned variant evaluated on each backbone.

	Params	GFlowNet	RW-MLE
ESM2	650M	✓	
ProteinReasoner	650M	✓	✓

Table 2: Antibody property benchmarks used in the zero-shot evaluation.

Objective	Dataset	#Seeds	#Samples
Binding	SKEMPI-V2	18	343
	AB-Bind	11	257
	IgDesign	7	310
Thermostability	Hutchinson2023	2	216
	BC-T07[†] (T)	1	140
Expression	Koenig2017	1	4,274
	BC-T07[†] (E)	1	48
	Adam2017 (Adams et al., 2016)	1	10,857

[†] Internal benchmark; not publicly released.

4.2 Zero-shot Evaluation

We evaluate AbEvolve in a zero-shot setting, comparing Vanilla pre-trained backbones, AbEvolve-aligned variants, self-supervised antibody PLMs, and CoSINE (Lu et al., 2026)—a continuous-time Markov chain post-trained on antibody lineage data—all ranked together on these benchmarks in Table 3.

Comparison with antibody-specific baselines. We compare AbEvolve with self-supervised antibody PLMs that lack explicit evolutionary supervision—IgBert (Kenlay et al., 2024), AntiBERTy (Ruffolo et al., 2021), and AbLang2 (Olsen et al., 2024)—and with CoSINE (Lu et al., 2026).

Table 3: Zero-shot Spearman ρ on binding, thermostability, and expression. AbEvolve is compared with Vanilla backbones, self-supervised antibody PLMs, and CoSINE (Lu et al., 2026).

	Binding			Thermostability		Expression		
Method	SKEMPI-V2	AB-Bind	IgDesign	Hutchinson2023	BC-T07†	Koenig2017	BC-T07†	Adam2017
IgBert	0.057	0.033	-0.156	0.214	0.171	0.238	0.171	0.067
AntiBERTy	0.089	-0.067	-0.036	-0.142	-0.116	0.099	0.144	-0.117
AbLang2	0.046	0.042	-0.187	-0.200	0.306	0.240	**0.519**	0.000
ESM2	0.075	0.122	0.004	0.260	0.164	0.360	0.418	0.147
ProteinReasoner	0.019	0.188	0.211	0.208	<u>0.607</u>	0.292	0.399	**0.425**
CoSINE	**0.290**	0.196	0.203	0.379	0.474	**0.494**	0.438	<u>0.404</u>
ESM2-GFlowNet	0.214	0.196	0.250	**0.418**	0.447	0.399	0.453	0.304
PR-GFlowNet	0.257	**0.273**	**0.258**	<u>0.405</u>	**0.635**	<u>0.413</u>	0.420	0.275
PR-RW-MLE	<u>0.265</u>	<u>0.245</u>	<u>0.252</u>	0.361	0.560	0.127	<u>0.461</u>	0.343

Note. Bold indicates the best performance in each column, and underlining indicates the second-best. † denotes an internal benchmark not publicly released. PR-GFlowNet and PR-RW-MLE abbreviate ProteinReasoner-GFlowNet and ProteinReasoner-RW-MLE.

Table 4: Unified ablation of AbEvolve on ProteinReasoner. Data-exposure compares vanilla, CP, and PR-GFlowNet at 260K; label-shuffling applies the same control to PR-GFlowNet and PR-RW-MLE at 2.6K; mutation-ordering varies only the FGMO rule. Spearman ρ is reported.

			Binding			Average	
Ablation	Configuration	Scale	SKEMPI-V2	AB-Bind	IgDesign	Thermostability	Expression
Data exposure	PR-Vanilla	–	0.019	0.188	0.211	0.408	**0.372**
	PR-CP	260K	0.106	0.150	0.236	0.015	0.239
	PR-GFlowNet	260K	**0.257**	**0.273**	**0.258**	**0.520**	0.369
Label shuffling	PR-GFlowNet	2.6K	**0.208**	**0.256**	0.183	**0.520**	0.331
	PR-GFlowNet (shuffled)	2.6K	0.078	0.161	**0.220**	0.446	**0.343**
	PR-RW-MLE	2.6K	**0.130**	**0.260**	**0.266**	**0.468**	0.294
	PR-RW-MLE (shuffled)	2.6K	0.065	0.185	0.206	0.396	**0.316**
Mutation ordering	FGMO	2.6K	**0.208**	**0.256**	**0.183**	**0.520**	**0.331**
	Random	2.6K	0.105	0.090	0.127	0.455	0.257
	Rev.-FGMO	2.6K	0.167	0.136	0.140	0.443	0.302

Note. PR abbreviates ProteinReasoner; shuf. = abundance-label shuffling within each lineage. Thermostability and Expression columns report category averages (Hutchinson2023 and BC-T07; Koenig2017, BC-T07, and Adam2017, respectively). Bold marks the better result within each block.

Against the self-supervised PLMs, AbEvolve improves all three binding benchmarks and both thermostability benchmarks, except AbLang2 on one expression benchmark (BC-T07, $\rho = 0.519$; best AbEvolve 0.461). Versus CoSINE, the best AbEvolve variant in each column wins five of eight benchmarks, most clearly on thermostability (Hutchinson2023 0.418 vs. 0.379; BC-T07 0.635 vs. 0.474); CoSINE remains best on SKEMPI-V2 (0.290 vs. 0.265) and on Koenig2017 (0.494 vs. 0.413) and Adam2017 (0.404 vs. 0.343).

AbEvolve improves zero-shot property ranking, with the clearest gains on binding and thermostability. Relative to each vanilla backbone, GFlowNet alignment improves binding on all three datasets for both ESM2 and ProteinReasoner and improves thermostability on both architectures (Table 3). ProteinReasoner-RW-MLE also improves binding but is mixed on thermostability, gaining on Hutchinson2023 (0.361 vs. 0.208) while falling below vanilla on BC-T07 (0.560 vs. 0.607). **ESM2-GFlowNet** improves on all eight benchmarks, whereas ProteinReasoner expression gains vary by dataset. Clonal-abundance trajectories thus provide the most reliable supervision for binding- and thermostability-oriented ranking.

4.3 Ablation Study

Unless noted, ablations use ProteinReasoner based models on the 2.6K largest lineages for its computational cost and check 260K in Appendix A.7. Table 4 reports data-exposure, label-shuffling, and mutation-ordering; Table 5 reports seed re-selection under the same shuffle.

Impact of continuous pre-training. To test whether gains come from extra OAS exposure rather than trajectory supervision, we compare ProteinReasoner-GFlowNet with continuous pre-training (CP) on the same OAS data without abundance or trajectory labels. Table 4 shows that CP only

Table 5: Start-sequence sensitivity under abundance shuffling. Canonical and shuffled-R differ only in whether redundancy labels are permuted within each lineage before seed selection.

Dataset	# Lineages (seeds)	# Unique seed sequences	Matching seeds vs. canonical
Canonical	2,600	593	–
Shuffled-R	2,600	2,579	25 / 2,600 (1.0%)

mildly changes the binding average (0.139 vs. 0.164), while thermostability collapses (0.408 vs. 0.015) and expression falls (0.372 vs. 0.239). At the same 260K scale, ProteinReasoner-GFlowNet recovers binding and thermostability and restores expression near vanilla, exceeding CP on all three binding columns and both category averages in Table 4. On ESM2, GFlowNet alignment exceeds CP on all eight benchmarks (Appendix A.7).

Impact of abundance label shuffling. We next shuffle abundance labels within each lineage and rebuild trajectories from the permuted counts, which also re-selects the min-abundance seed. Shuffling reduces binding and thermostability averages for both ProteinReasoner-GFlowNet (0.216 vs. 0.153; 0.520 vs. 0.446) and ProteinReasoner-RW-MLE (0.219 vs. 0.152; 0.468 vs. 0.396), while expression stays roughly unchanged.

The same permutation is load-bearing in data construction (Table 5): canonical min-R collapses 2,600 seeds to 593 unique starts, whereas shuffled-R yields 2,579 unique starts (Jaccard 0.0076; 1.0% record-level match). Thus the sequence–redundancy association, not lineage structure alone, drives seed selection and the main binding and thermostability gains.

Impact of FGMO ordering. To isolate mutation ordering, we compare FGMO with random order and reverse-FGMO on the 2.6K lineages, holding data, labels, and the GFlowNet alignment objective fixed; variants are scored under the same rule used in training. Table 4 shows that FGMO improves all three averages over both alternatives (binding 0.216 vs. 0.107/0.148; thermostability 0.520 vs. 0.455/0.443; expression 0.331 vs. 0.257/0.302), so frequency-guided order supplies a stronger abundance-derived selection signal than random or reversed alternatives.

5 Conclusion

We introduce AbEvolve, which aligns protein language models with antibody maturation signals from clonal abundance trajectories. RW-MLE and GFlowNet alignment absorb scalar abundance labels along constructed mutational paths, going beyond additional repertoire exposure. ESM2-GFlowNet improves over vanilla ESM2 on all eight benchmarks. For ProteinReasoner, GFlowNet alignment improves binding on all three benchmarks and thermostability on both benchmarks, and expression gains vary by benchmark. Looking ahead, we will scale AbEvolve to additional large-scale weak-label repertoire sources beyond OAS unpaired heavy or light chains—including nanobody repertoires in INDI (Deszyński et al., 2022), paired VH–VL BCR repertoires mined from public sequencing archives such as the Sequence Read Archive (SRA) (Leinonen et al., 2011), and curated antigen-linked metadata so abundance signals can be conditioned on the selecting target. We will also refine pseudo-trajectory construction by incorporating phylogenetically inferred parent–child structure where reliable clonal trees are available, complementing coarse abundance ordering with higher-resolution maturation paths, and explore GFlowNets for diverse mutation generation and multi-objective optimization of affinity and developability.

6 Limitations

AbEvolve has two limitations that bound the scope of our claims. First, training relies on a coarse OAS-based dataset: clonal abundance provides repertoire-proximal weak labels rather than validated experimental fitness assays, and pseudo-trajectories are constructed from observed lineage structure without high-resolution phylogenetic reconstruction. Second, improvements are property- and architecture-dependent: abundance-centric supervision is most reliable for binding and thermostability, whereas ProteinReasoner expression ranking can degrade on some benchmarks even though ESM2-GFlowNet improves on all eight evaluations.

AI use statement

In this work, we used generative AI tools for editing and polishing the manuscript text, specifically for language refinement and figure formatting assistance. We have not used generative AI tools for formulating research ideas, designing experiments, or interpreting results. All experimental design, data analysis, model development, and scientific conclusions were conducted independently by the authors. We have reviewed all AI-assisted editorial work and take full responsibility for the final content of this paper, including all claims, methodology, and presented results.

Ethics statement

This work uses publicly available antibody sequence data from the Observed Antibody Space (OAS) database, together with one internal developability benchmark (BC-T07) that contains only de-identified protein sequences and is not publicly released. All data used in this study consist of protein sequences without associated personal or clinical information. Our framework is designed for therapeutic antibody optimization, with the goal of improving drug development efficiency. We have not identified significant ethical concerns with this research, as it does not involve human subjects, dual-use risks, or potentially harmful applications beyond standard biomedical research.

Reproducibility statement

We have made efforts to ensure the reproducibility of our results. The AbEvolve framework, including the trajectory construction pipeline and alignment objectives (RW-MLE and GFlowNet alignment), is described in detail in Sections 3.2, 3.3.1, and 3.3.2. The hyperparameter settings and experimental configurations are provided in the appendix. The figure source files and experimental scripts will be made available in an anonymous repository upon acceptance. The BC-T07 internal benchmark cannot be released; all public benchmarks and the OAS-derived training pipeline will be documented.

References

Rhys M Adams, Thierry Mora, Aleksandra M Walczak, and Justin B Kinney. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife*, 5:e23156, 2016. doi: 10.7554/eLife.23156.

Frances H Arnold. Directed evolution: Bringing new chemistry to life. *Angewandte Chemie*, 2018.

Rachael JM Bashford-Rogers, Anne L Palser, Brian J Huntly, Richard Rance, George S Vassiliou, George A Follows, and Paul Kellam. Network properties derived from deep sequencing of human b-cell receptor repertoires delineate b-cell populations. *Genome research*, 23(11):1874–1884, 2013.

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in neural information processing systems*, 34:27381–27394, 2021.

Yoshua Bengio, Salem Lahlou, Tristan Deleu, Edward J Hu, Mo Tiwari, and Emmanuel Bengio. Gflownet foundations. *Journal of Machine Learning Research*, 24(210):1–55, 2023.

Deborah L Burnett, David B Langley, Peter Schofield, Jana R Hermes, Tyani D Chan, Jennifer Jackson, Katherine Bourne, Joanne H Reed, Kate Patterson, Benjamin T Porebski, et al. Germinal center antibody mutation trajectories are determined by rapid self/foreign discrimination. *Science*, 360(6385):223–226, 2018.

Piotr Deszyński, Jakub Młokosiewicz, Adam Volanakis, Iga Jaszczyszyn, Natalie Castellana, Stefano Bonissone, Rajkumar Ganesan, and Konrad Krawczyk. INDI—integrated nanobody database for immunoinformatics. *Nucleic Acids Research*, 50(D1):D1273–D1281, 2022.

David TF Dryden, Andrew R Thomson, and John H White. How much of protein sequence space has been explored by life on earth? *Journal of the Royal Society Interface*, 5(25):953–956, 2008.

Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. Prottrans: toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127, 2021.

Simon Friedensohn, Tarik A Khan, and Sai T Reddy. Advanced methodologies in high-throughput sequencing of immune repertoires. *Trends in biotechnology*, 35(3):203–214, 2017.

Elena Gabrielli, Eva Pericolini, Elio Cenci, Federica Ortelli, Walter Magliani, Tecla Ciociola, Francesco Bistoni, Stefania Conti, Anna Vecchiarelli, and Luciano Polonelli. Antibody complementarity-determining regions (cdrs): a bridge between adaptive and innate immunity. *PloS one*, 4(12):e8187, 2009.

Pouya M Ghari, Alex M Tseng, Gökcen Eraslan, Romain Lopez, Tommaso Biancalani, Gabriele Scalia, and Ehsan Hajiramezanali. Gflownet assisted biological sequence editing. *Advances in Neural Information Processing Systems*, 37:106841–106869, 2024.

Namita T Gupta et al. Change-o: a highly-scalable computational framework for bcr repertoire analysis. *Bioinformatics*, 31(20):3350–3352, 2015.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q Tran, Jonathan Deaton, Marius Wiggert, et al. Simulating 500 million years of evolution with a language model. *Science*, 387(6736):850–858, 2025.

Brian L Hie et al. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 2024.

Mark Hutchinson, Jeffrey A Ruffolo, Nantaporn Haskins, Michael Iannotti, Giuliana Vozza, Tony Pham, Nurjahan Mehzabeen, Harini Shandilya, Keith Rickert, Rebecca Croasdale-Wood, et al. Toward enhancement of antibody thermostability and affinity by computational design in the absence of antigen. *mAbs*, 16(1):2362775, 2024. doi: 10.1080/19420862.2024.2362775.

Moksh Jain, Sharath Chandra Raparthy, Alex Hernández-Garcıa, Jarrid Rector-Brooks, Yoshua Bengio, Santiago Miret, and Emmanuel Bengio. Multi-objective gflownets. In *International conference on machine learning*, pp. 14631–14653. PMLR, 2023.

Tushar Jain et al. The developability of therapeutic antibodies. *Proceedings of the National Academy of Sciences*, 2017.

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H Moal. Skempi 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019.

Henry Kenlay, Frédéric A Dreyer, Aleksandr Kovaltsuk, Dom Miketa, Douglas Pires, and Charlotte M Deane. Large scale paired antibody language models. *PLOS Computational Biology*, 20 (12):e1012646, 2024.

Patrick Koenig, Chingwei V Lee, Benjamin T Walters, Vasantharajan Janakiraman, Jeremy Stinson, Thomas W Patapoff, and Germaine Fuh. Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proceedings of the National Academy of Sciences*, 114(4):E486–E495, 2017.

Marie-Paule Lefranc, Véronique Giudicelli, Patrice Duroux, Joumana Jabado-Michaloud, Géraldine Folch, Safa Aouinti, Emilie Carillon, Hugo Duvergey, Amélie Houles, Typhaine Paysan-Lafosse, et al. Imgt®, the international immunogenetics information system® 25 years on. *Nucleic acids research*, 43(D1):D413–D422, 2015.

Rasko Leinonen, Hideaki Sugawara, Martin Shumway, and International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Research*, 39(suppl_1):D19–D21, 2011. doi: 10.1093/nar/gkq1022.

Vladimir I Levenshtein et al. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pp. 707–710. Soviet Union, 1966.

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

Chaozhong Liu, Linlin Chao, Shaomin Ji, Hao Wang, Taorui Jiang, Zhangyang Gao, Yucheng Guo, Ming Yang, and Xiaoming Zhang. Proteinreasoner: A multi-modal protein language model with chain-of-thought reasoning for efficient protein design. *bioRxiv*, pp. 2025–07, 2025.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

RM Lu et al. Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, 2020.

Stephen Zhewen Lu, Aakarsh Vermani, Kohei Sanno, Jiarui Lu, Frederick A Matsen, Milind Jagota, and Yun S Song. Conditionally site-independent neural evolution of antibody sequences. *arXiv preprint arXiv:2602.18982*, 2026.

Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: Improved credit assignment in gflownets. *Advances in Neural Information Processing Systems*, 35:5955–5967, 2022.

Frederick A Matsen IV, Will Dumm, Kevin Sung, Mackenzie M Johnson, David H Rich, Tyler N Starr, Yun S Song, Julia Fukuyama, and Hugh K Haddox. Separating selection from mutation in antibody language models. *eLife*, 15:RP109644, 2026. doi: 10.7554/eLife.109644.

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

Julia Merkenschlager, Andrew GT Pyo, Gabriela S Silva Santos, Dennis Schaefer-Babajew, Melissa Cipolla, Harald Hartweger, Alexander D Gitlin, Ned S Wingreen, and Michel C Nussenzweig. Regulated somatic hypermutation enhances antibody affinity maturation. *Nature*, 641(8062):495–502, 2025.

Luka Mesin, Jonatan Ersching, and Gabriel D Victora. Germinal center b cell dynamics. *Immunity*, 45(3):471–482, 2016.

Marco Molari, Rémi Monasson, and Simona Cocco. Survival probability and size of lineages in antibody affinity maturation. *Physical Review E*, 103(5):052413, 2021.

Daniel Müllner. Modern hierarchical agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*, 2011.

Pascal Notin, Aaron Kollasch, Daniel Ritter, Lood Van Niekerk, Steffanie Paul, Han Spinner, Nathan Rollins, Ada Shaw, Rose Orenbuch, Ruben Weitzman, et al. Proteingym: Large-scale benchmarks for protein fitness prediction and design. *Advances in neural information processing systems*, 36: 64331–64379, 2023.

Namita Nouri and Steven H Kleinstein. Optimizing thresholded distance-based b-cell clonal grouping methods. *Frontiers in Immunology*, 9:2090, 2018.

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, 2022a.

Tobias H. Olsen, Iain H. Moal, and Charlotte M. Deane. Ablang: an antibody language model for completing antibody sequences. *Bioinformatics Advances*, 2(1):vbac046, 2022b. doi: 10.1093/bioadv/vbac046.

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. doi: 10.1093/bioinformatics/btae618.

Mohit Pandey, Gopeshh Subbaraj, Artem Cherkasov, Martin Ester, and Emmanuel Bengio. Pretraining generative flow networks with inexpensive rewards for molecular graph generation. *arXiv preprint arXiv:2503.06337*, 2025.

William E Paul. *Fundamental immunology*. Lippincott Williams & Wilkins, 2012.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

Duncan K Ralph, Athanasios G Bakis, Jared Galloway, Ashni A Vora, Tatsuya Araki, Gabriel D Victora, Yun S Song, William S DeWitt, and Frederick A Matsen IV. Inference of germinal center evolutionary dynamics via simulation-based deep learning. *eLife*, 14:RP108880, 2026. doi: 10.7554/eLife.108880.

Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yan Duan, Peter Chen, John Canny, Pieter Abbeel, and Yun Song. Evaluating protein transfer learning with tape. *Advances in neural information processing systems*, 32, 2019.

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

Harlan Robins. Immunosequencing: applications of immune repertoire deep sequencing. *Current opinion in immunology*, 25(5):646–652, 2013.

Jeffrey A Ruffolo, Jeffrey J Gray, and Jeremias Sulam. Deciphering antibody affinity maturation with language models and weakly supervised learning. *arXiv preprint arXiv:2112.07782*, 2021.

David G Schatz and Patrick C Swanson. V (d) j recombination: mechanisms of initiation. *Annual review of genetics*, 45(1):167–202, 2011.

Gila Shakked et al. Prospective prediction of protein-protein orientation using evolutionary patterns. *Nature Communications*, 2022.

Amir Shanehsazzadeh, Julian Alverio, George Kasun, Simon Levine, Ido Calman, Jibran A Khan, Chelsea Chung, Nicolas Diaz, Breanna K Luton, Ysis Tarter, et al. Igdesign: In vitro validated antibody design against multiple therapeutic antigens using inverse folding. *bioRxiv*, 2023. doi: 10.1101/2023.12.08.570889.

Sarah Sirin, James R Apgar, Eric M Bennett, and Amy E Keating. Ab-bind: antibody binding mutational database for computational affinity predictions. *Protein Science*, 25(2):393–409, 2016.

Cole R Tharp et al. Biophysical trade-offs in antibody evolution are resolved by conformation-mediated epistasis. *bioRxiv*, 2026.

Susumu Tonegawa. Somatic generation of antibody diversity. *Nature*, 302(5909):575–581, 1983.

Vincent Q Tran, Matthew Nemeth, Liam J Bartie, Sita S Chandrasekaran, et al. Rapid directed evolution guided by protein language models and epistatic interactions. *Science*, 392(6798):eaea1820, 2026. doi: 10.1126/science.aea1820.

Gabriel D Victora and Luka Mesin. Clonal and cellular dynamics in germinal centers. *Current opinion in immunology*, 28:90–96, 2014.

Gabriel D Victora and Michel C Nussenzweig. Germinal centers. *Annual review of immunology*, 40:413–442, 2022.

George J Weiner. Building better antibodies. *Nature Reviews Cancer*, 2015.

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*, 23:1805–1813, 2026. doi: 10.1038/s41592-026-03137-3.

Gur Yaari and Steven H Kleinstein. Practical guidelines for b-cell receptor repertoire sequencing analysis. *Genome medicine*, 7(1):121, 2015.

A Appendix

A.1 Antibody-Specific Evolution

Antibodies, or immunoglobulins, as shown in Figure S1, are Y-shaped proteins produced by B cells as part of the adaptive immune system. Each antibody specifically binds to a target antigen, allowing the immune system to recognize and neutralize pathogens. Structurally, antibodies consist of variable regions (VH, VL) and constant regions (CH, CL) (Paul, 2012). The variable regions of these chains contain hypervariable loops called complementarity-determining regions (CDRs), which primarily mediate antigen recognition. Among these, heavy-chain CDR3 (HCDR3) typically exhibits the highest sequence diversity and plays a major role in determining binding specificity (Gabrielli et al., 2009).

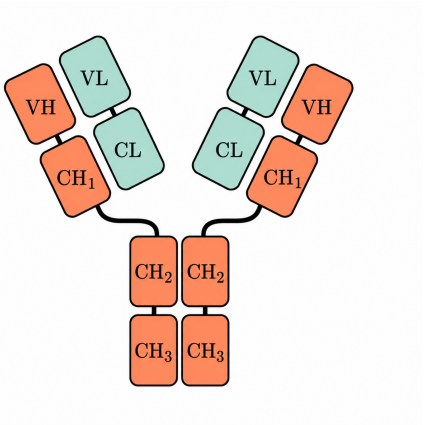

Figure S1: Schematic of antibody structure.

Antibody evolution is a multi-stage process. Each B cell originates from a **germline** precursor, whose immunoglobulin genes consist of variable (V), diversity (D), and joining (J) segments (Mesin et al., 2016). During early B cell development in the bone marrow, these segments undergo **VDJ recombination**, a stochastic DNA rearrangement process that assembles one V, one D, and one J segment for the heavy chain (and one V and one J for the light chain) (Tonegawa, 1983). This process occurs *before antigen exposure* and generates a highly diverse primary antibody repertoire, providing the raw material for immune recognition (Schatz & Swanson, 2011).

During affinity maturation in germinal centers (GCs), B cells accumulate somatic hypermutations (SHM), generating diverse related sequences (Mesin et al., 2016; Burnett et al., 2018; Victora & Nussenzweig, 2022). In this stage, point mutations accumulate in the variable regions, and B cells with higher antigen-binding affinity are preferentially expanded, forming a hierarchical network of variants that reflects the mutational history of the lineage (Merkenschlager et al., 2025). A **clonal lineage** is defined as a set of sequences that are evolutionarily related and share a common origin (Gupta et al., 2015). Within a lineage, sequences differ by the number and positions of mutations. This concept provides a framework for analyzing antibody diversity and evolutionary relationships, without requiring any computational trajectory to start from the germline sequence.

As shown in Figure 2, in AbEvolve, evolutionary trajectories are reconstructed from low- to high-abundance sequences within each clonal lineage. These trajectories do not necessarily begin at the germline sequence; instead, they leverage observed clonal abundance as a proxy for selection pressure. The ordering of mutations along each trajectory is determined using Frequency-Guided Mutation Ordering (FGMO), which prioritizes frequently observed mutations to construct biologically plausible intermediate states.

A.2 Detailed Methodology for Antibody Trajectory Construction

In this section, we provide the technical specifics and biological rationale for the AbEvolve data construction pipeline, expanding on the summary provided in Section 3.2.

A.2.1 Sequence Restoration via Germline Anchoring

A primary challenge in utilizing the Observed Antibody Space (OAS) (Olsen et al., 2022a) for lineage analysis is the high frequency of N- and C-terminal truncations. These artifacts are intrinsic to high-throughput library preparation, where multiplex primers typically target framework regions (FR1 and FR4) to ensure broad amplification across diverse V-gene families (Friedensohn et al., 2017). This often results in the loss of terminal residues, which, if left uncorrected, introduces artificial mutational gaps that confound phylogenetic construction.

Restoration Pipeline: For each truncated sequence $\mathbf{s}$, we implement the following restoration steps:

1. **Gene-Call Retrieval**: We extract the V- and J-gene allele assignments (e.g., *IGHV3-23*01*) provided by the OAS metadata.
2. **Template Matching**: We query the IMGT/V-QUEST reference directory (Lefranc et al., 2015) to retrieve the corresponding full-length germline leader, framework, and constant region templates.
3. **Anchored Completion**: The truncated sequence is aligned to the germline template. We then substitute the missing N-terminal (FR1) and C-terminal (FR4) segments with the exact germline segments.
4. **Root Identification**: The resulting fully restored sequence provides a standardized germline reference frame for calculating mutational distance and directionality. Note that the computational trajectory seed is later selected by minimum abundance within each lineage (Section 3.2), not by the germline sequence.

A.2.2 Clonal Identification and Hierarchical Clustering

To ensure that trajectories represent true biological maturation rather than random sequence variation, we employ a multi-stage clustering approach.

Initial Grouping: We first partition the individual repertoires into V-J-CDR3L groups. Following the conventions of (Yaari & Kleinstein, 2015; Gupta et al., 2015), sequences must share the same V-gene, J-gene, and CDR3 length. This stage isolates sequences that likely share a common V(D)J recombination event, significantly reducing the search space for lineage construction.

Clonal Refinement: Within each V-J-CDR3L group, we perform agglomerative hierarchical clustering based on the Levenshtein distance. We enforce a distance threshold of $d = 6$ to delineate distinct clonal lineages $\mathcal{C}$. This specific threshold is selected based on a balanced consideration of the following factors:

- **Somatic Hypermutation (SHM) Modeling:** It provides a sufficient margin to accommodate natural sequence diversification during B-cell affinity maturation.
- **Lineage Specificity:** It effectively prevents the over-clustering of unrelated sequences, ensuring that each identified clone originates from a single common ancestor.
- **Data Robustness:** It accounts for potential sequencing or PCR-induced errors while maintaining a high signal-to-noise ratio in defining clonal boundaries.

Length Uniformity Constraint: To model transitions as discrete amino acid substitutions, we exclude any sequence within a lineage that deviates from the consensus length. This step removes rare insertions and deletions (indels), ensuring that each transition $\mathbf{s}_t \rightarrow \mathbf{s}_{t+1}$ can be represented as a single or multiple amino acid substitution.

A.2.3 Validation of Clustering Integrity

To justify the clustering threshold ($d = 6$), we analyzed pairwise edit distances within V-J-CDR3L groups, before the $d = 6$ cut. We sampled up to 100 such groups per individual and at most 1,000 members per group. The aggregated distribution (Figure S2) is unimodal, with a sharp peak between $d = 1$ and $d = 2$, and more than 92% of pairwise distances fall at or below $d = 6$. This unimodal signature is consistent with a single genetic origin; inter-clonal mixtures typically appear multimodal with higher distance peaks. The $d = 6$ cutoff therefore retains most intra-group maturation-related variation while excluding distant non-lineage pairs.

A.2.4 Mapping Redundancy to Clonal Abundance

Finally, for each unique sequence $\mathbf{s}$ in the resulting lineage $\mathcal{C}$, we map the OAS "redundancy" count to its *clonal abundance* $A(\mathbf{s})$. In the context of repertoire sequencing, redundancy reflects the number of times a specific sequence (or highly similar reads) was detected. Under the principle of clonal selection, higher abundance is treated as a coarse proxy for clonal expansion and survival under antigenic stimulation, not as a validated experimental fitness assay. These abundance values provide the selection-derived supervision used by the alignment algorithms in Section 3.

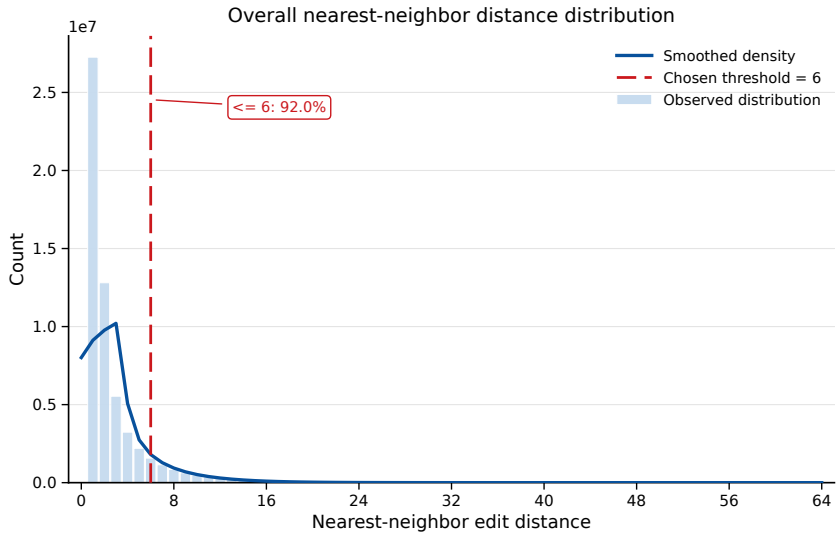

Figure S2: **Intra-group edit-distance distribution.** Levenshtein distances between sequences in the same V-J-CDR3L group, before the $d = 6$ cut. The distribution is unimodal (centered at $d \approx 2$). **More than** 92% **of pairwise distances satisfy** $d \leq 6$, as indicated by the cumulative density. The red dashed line marks the clustering threshold $d = 6$.

A.3 Detailed Methodology for Model Architectures

In this section, we provide a comprehensive technical description of the alignment objectives and training protocols employed in our study. We first detail RW-MLE endpoint alignment (Section 3.3.1), then GFlowNet trajectory alignment (Section 3.3.2), including state and action space definitions, neural backbones used for policy parameterization, and the objective functions that govern learning.

A.3.1 Abundance-Guided Endpoint Alignment via RW-MLE

Figure S3 illustrates the RW-MLE endpoint-alignment architecture. RW-MLE integrates abundance supervision at trajectory endpoints. For each lineage, terminal sequences are reweighted by their clonal-abundance reward from Eq. equation 1, so the model upweights high-abundance variants during alignment. AbEvolve optimizes Eq. equation 2 using the backbone's seed-conditioned terminal likelihood and within-lineage normalized rewards $\bar{R}(\mathbf{s})$. Unlike GFlowNet alignment, which propagates rewards along the full mutational trajectory, RW-MLE operates exclusively on terminal states, utilizing the scalar abundance signal without modeling intermediate abundance-derived selection dynamics along the path.

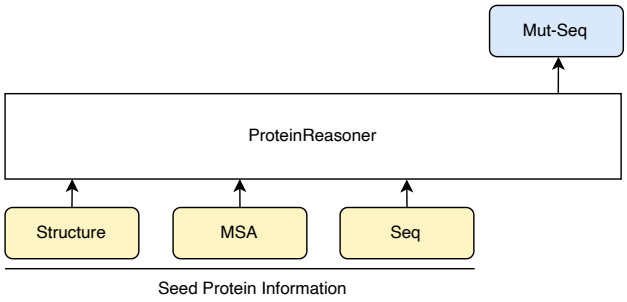

Figure S3: Schematic of abundance-guided endpoint alignment via RW-MLE.

RW-MLE is instantiated only on ProteinReasoner. The seed-conditioned term $\pi_\theta(\mathbf{s} \mid \mathbf{s}_{\text{seed}})$ in Eq. equation 2 is the backbone's native structure–MSA-conditioned likelihood: each terminal variant

s is scored in the context of its lineage seed's MSA profile and structural tokens (Figure S3), without additional policy heads beyond the pretrained likelihood model. Training and zero-shot scoring rules are summarized in Appendix A.4.

A.3.2 Abundance-Guided Trajectory Alignment via GFlowNet

Figure S4 illustrates the architecture of GFlowNet trajectory alignment (Section 3.3.2). We implement the same FGMO-supervised state paths as the main text—not free-sequence generation—by parameterizing forward and backward transition policies over an editing graph.

Definitions and State Space Following Section 3.1, let $\mathcal{V}$ denote the 20-amino-acid alphabet and $\mathbf{s} \in \mathcal{V}^L$ a fixed-length antibody sequence. Each constructed lineage path is written as in the main text:

$$\tau_i = (\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(K_i)}), \tag{4}$$

where $\mathbf{s}^{(0)} = \mathbf{s}_{\text{seed}}$ (minimum abundance within the lineage during training; the benchmark parent at evaluation), $\mathbf{s}^{(K_i)} = \mathbf{s}_i$ is a terminal variant, and intermediate states $\mathbf{s}^{(t)}$ for $t = 1, \ldots, K_i - 1$ are observed lineage sequences or FGMO virtual nodes $\hat{\mathbf{s}}^{(t)}$. Each consecutive pair $(\mathbf{s}^{(t-1)}, \mathbf{s}^{(t)})$ is a single-point mutation step.

Action parameterization of state transitions For implementation, each state transition $\mathbf{s}^{(t-1)} \to \mathbf{s}^{(t)}$ is induced by an editing action $a_t = (i, aa)$, where $i \in \{1, \ldots, L\}$ is the mutated position and $aa \in \mathcal{V}$ is the target residue; this defines the forward transition probability $P_F(\mathbf{s}^{(t)} \mid \mathbf{s}^{(t-1)}; \theta)$ used in Eq. equation 3. The path reward is the trajectory-level abundance shift $R(\tau_i)$ from Eq. equation 1, not a separate terminal-state score.

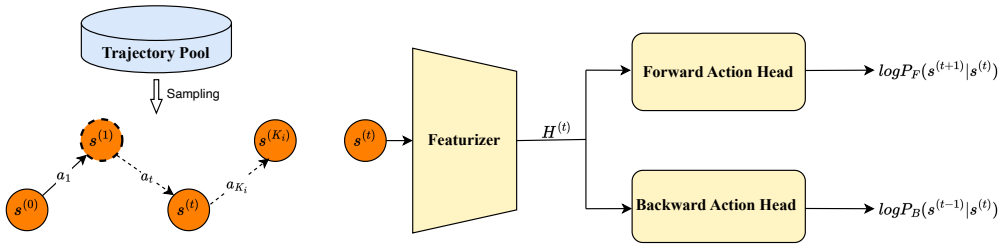

Figure S4: Schematic of our proposed architecture for AbEvolve-GFlowNet.

Sequence State Featurizer The featurizer, parameterized by a pretrained PLM (e.g., ESM2 (Lin et al., 2023), ProteinReasoner (Liu et al., 2025) in this work), defines a mapping $\Phi : \mathcal{V}^L \to \mathbb{R}^{L \times d_{model}}$, which transforms the discrete sequence into a sequence of continuous latent embeddings:

$$\mathbf{H}^{(t)} = \text{PLM}(\mathbf{s}^{(t)}) = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_L]^\top, \quad \mathbf{h}_i \in \mathbb{R}^{d_{model}} \tag{5}$$

where $\mathbf{h}_i$ denotes the residue-level embedding for the i-th amino acid, capturing both its intrinsic physicochemical identity and its contextual relationship within the antibody variable region.

Forward Policy Network Architectures The forward transition policy $P_F(\mathbf{s}^{(t)} \mid \mathbf{s}^{(t-1)}; \theta)$ from Eq. equation 3 is implemented by factorizing the inducing action $a_t = (i, aa)$ into site selection and substitution. This decomposition reduces the complexity of the action space from $L \times 20$ to $L + 20$, while allowing the model to leverage both local residue properties and global sequence context. Both heads condition on the predecessor state $\mathbf{s}^{(t-1)}$, with $\mathbf{H}^{(t-1)} = \text{PLM}(\mathbf{s}^{(t-1)})$.

- **Position Head:** We aggregate predecessor-state information into a global context vector $\mathbf{g}^{(t-1)} = \text{MeanPooling}(\mathbf{H}^{(t-1)}) \in \mathbb{R}^{d_{model}}$, form an augmented matrix $\tilde{\mathbf{H}}^{(t-1)} \in \mathbb{R}^{L \times 2d_{\text{model}}}$ by concatenating $\mathbf{g}^{(t-1)}$ to each row of $\mathbf{H}^{(t-1)}$, apply LayerNorm, and pass the result through a two-layer MLP (f_{pos}):

$$P(i \mid \mathbf{s}^{(t-1)}; \theta) = \text{softmax}\Big(f_{\text{pos}}(\text{LayerNorm}(\tilde{\mathbf{H}}^{(t-1)}))\Big)_i \tag{6}$$

where each index $i \in \{1, \ldots, L\}$ corresponds to a candidate mutation position on the FGMO path.

- **Amino Acid Substitution Head:** In parallel, the predecessor-state embeddings $\mathbf{H}^{(t-1)}$ are processed by an amino acid head f_{aa} (a two-layer MLP):

$$P(aa \mid i, \mathbf{s}^{(t-1)}; \theta) = \text{softmax}\Big(f_{\text{aa}}(\mathbf{h}_i^{(t-1)})\Big)_{aa} \tag{7}$$

The log-probability of the FGMO transition $\mathbf{s}^{(t-1)} \to \mathbf{s}^{(t)}$ induced by $a_t = (i, aa)$ is

$$\log P_F(\mathbf{s}^{(t)} \mid \mathbf{s}^{(t-1)}; \theta) = \log P(i \mid \mathbf{s}^{(t-1)}; \theta) + \log P(aa \mid i, \mathbf{s}^{(t-1)}; \theta). \tag{8}$$

This hierarchical structure conditions substitution on the chosen mutation site; LayerNorm stabilizes feature integration across PLM scales.

Backward Policy Network Architecture The backward transition policy $P_B(\mathbf{s}^{(t-1)} \mid \mathbf{s}^{(t)}; \theta)$ in Eq. equation 3 uses the same dual-head architecture and PLM featurizer as P_F, with shared alignment parameters θ over lightweight action heads on the pre-trained backbone.

Implementation Details of Featurizer As shown in Figure S5, for ESM2 (Lin et al., 2023), the pretrained model naturally provides sequential embeddings and can be used directly as a featurizer. For ProteinReasoner (Liu et al., 2025), the featurizer processes the current state $\mathbf{s}^{(t)}$ in the context of the clonal lineage's shared seed sequence $\mathbf{s}_{\text{seed}}$, including its MSA profile and structural tokens, to generate a context-aware feature representation. The backbone processes these inputs to produce feature representations that capture both sequence context and structural constraints, enabling trajectory-aware modeling.

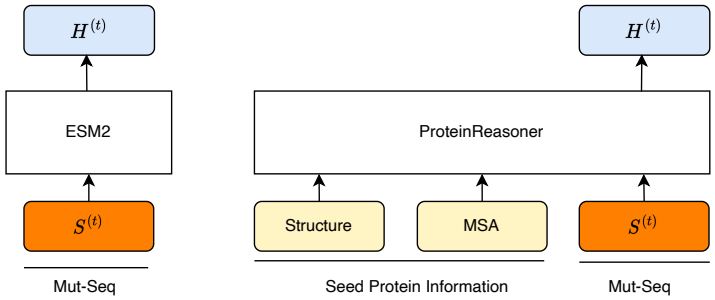

Figure S5: Schematic of the featurizer in GFlowNet alignment.

Pseudo-code The training process of AbEvolve-GFlowNet with Trajectory Balance (TB) optimization is detailed in Algorithm 1

A.4 Detailed Model Scoring Protocols

This section summarizes how each model variant assigns a scalar score to a candidate antibody sequence during zero-shot evaluation. All downstream metrics are computed by ranking variants according to these scores.

Vanilla and CP baselines (ESM2). For Vanilla ESM2 and ESM2-CP, we use the backbone's native pseudo-log-likelihood:

$$S_{\text{ESM2}}(\mathbf{s}) = \sum_{i=1}^{L} \log p_\theta(s_i \mid \mathbf{s}_{\setminus i}). \tag{9}$$

ESM2-CP uses the same scoring rule as Vanilla ESM2. ProteinReasoner scoring is defined separately below.

Algorithm 1 GFlowNet alignment training with Trajectory Balance

Input: FGMO trajectories $\mathcal{D} = \{\tau_i\}$, PLM featurizer Φ, alignment parameters θ (including Z_θ)
Output: Updated θ

for each batch $\mathcal{B} \subset \mathcal{D}$ **do**
Forward / Backward Probabilities
  for each $\tau_i = (\mathbf{s}^{(0)}, \ldots, \mathbf{s}^{(K_i)}) \in \mathcal{B}$ with $\mathbf{s}^{(0)} = \mathbf{s}_{\text{seed}}$ **do**
    $\log P_F(\tau_i) \leftarrow \sum_{t=1}^{K_i} \log P_F(\mathbf{s}^{(t)} \mid \mathbf{s}^{(t-1)}; \theta)$
    $\log P_B(\tau_i) \leftarrow \sum_{t=1}^{K_i} \log P_B(\mathbf{s}^{(t-1)} \mid \mathbf{s}^{(t)}; \theta)$
  each P_F, P_B term is computed via the dual-head action parameterization above
  end for
Trajectory Balance Objective
  $\mathcal{L}_{\text{TB}} \leftarrow \frac{1}{|\mathcal{B}|} \sum_{\tau_i \in \mathcal{B}} \left(\log Z_\theta + \log P_F(\tau_i) - \log P_B(\tau_i) - \log R(\tau_i)\right)^2$
  where $R(\tau_i)$ is the 0-1-normalized abundance-shift reward from Eq. equation 1 (i.e., $\log A(\mathbf{s}^{(K_i)}) - \log A(\mathbf{s}_{\text{seed}})$, min–max normalized across training trajectories so that $R \in (0, 1]$)
Parameter Update
  $\theta \leftarrow \theta - \eta \nabla \mathcal{L}_{\text{TB}}$
end for

ProteinReasoner variants (Vanilla, CP, and RW-MLE). For Vanilla ProteinReasoner, ProteinReasoner-CP, and ProteinReasoner-RW-MLE, we score each candidate $\mathbf{s}$ with the backbone's sequence-level structure–MSA-conditioned log-likelihood. For RW-MLE, the model also conditions on the seed sequence $\mathbf{s}_{\text{seed}}$, together with its MSA profile and structural tokens, and evaluates the full candidate sequence $\mathbf{s}$. At zero-shot evaluation, $\mathbf{s}_{\text{seed}}$ is the benchmark parent (wild-type). The score is

$$S_{\text{PR-RW-MLE}}(\mathbf{s}) = \sum_{i=1}^{L} \log p_\theta(s_i \mid \mathbf{s}_{<i}, \mathbf{s}_{\text{seed}}, \mathbf{MSA}_{\text{seed}}, \mathbf{Structure}_{\text{seed}}), \tag{10}$$

where each conditional term is computed under ProteinReasoner's native structure–MSA-conditioned likelihood rule. ProteinReasoner-CP and ProteinReasoner-RW-MLE use the same scoring rule at evaluation; alignment only changes the training objective in Eq. equation 2.

GFlowNet based models. For GFlowNet alignment, we score each candidate by the likelihood of its constructed mutational path under the learned forward policy. At evaluation the seed $\mathbf{s}_{\text{seed}}$ is the benchmark parent (wild-type), not an OAS training seed. Given a candidate sequence $\mathbf{s}$, we first identify the set of substitutions that transform $\mathbf{s}_{\text{seed}}$ into $\mathbf{s}$. For multi-mutant candidates, these substitutions are ordered using the same Frequency-Guided Mutation Ordering (FGMO) rule used during training. Although marginalizing over all possible mutation orders is conceptually possible, it scales factorially with the number of mutations and is not aligned with the FGMO-ordered trajectories used during training. We therefore use the same FGMO ordering at inference as a deterministic approximation. This yields a deterministic path

$$\tau(\mathbf{s}) = (\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(K)}), \quad \mathbf{s}^{(0)} = \mathbf{s}_{\text{seed}}, \quad \mathbf{s}^{(K)} = \mathbf{s}.$$

The GFlowNet alignment score is then defined as

$$S_{\text{GFN}}(\mathbf{s}) = \sum_{t=1}^{K} \log P_F\left(\mathbf{s}^{(t)} \mid \mathbf{s}^{(t-1)}; \theta\right), \tag{11}$$

where P_F is the learned forward policy. This score measures the evolutionary accessibility of the candidate along the FGMO-reconstructed trajectory.

A.5 Dataset Statistics

A.5.1 Trajectory Statistics

To characterize the constructed pseudo-trajectories, we report statistics of trajectory length, mutation count, and abundance shift in Table S6. These statistics quantify the scale and diversity of the supervision signals used by RW-MLE and GFlowNet alignment.

Table S6: Statistics of abundance-informed trajectories.

Statistic	Mean	Median	Std.	Min	Max
Trajectories per lineage	72.8	30.0	237.3	1.0	10791.0
Mutations per trajectory	3.5	3.0	1.6	0.0	6.0
Reward (0-1 Normalized)	0.16	0.09	0.214	7.18e-08	1.0

A.5.2 Zero-shot Benchmark Statistics

SKEMPI-V2. SKEMPI-V2 is a curated benchmark of mutation-induced changes in protein–protein binding affinity for structurally resolved complexes (Jankauskaitė et al., 2019). From SKEMPI-V2, we selected antibody-related single-point mutation data covering 18 targets, yielding 343 variants in total. Each target contains more than five variants, enabling seed-level zero-shot ranking evaluation. We use this subset to assess whether model scores can rank affinity-altering antibody variants within each target.

AB-Bind. AB-Bind is an antibody–antigen binding mutational database designed for computational affinity prediction (Sirin et al., 2016). The original database collects experimentally measured changes in binding free energy across antibody–antigen complexes. We curated 11 antibody seeds from AB-Bind, containing 257 single-point variants in total. Each seed includes mutations on both heavy and light chains and is used to evaluate zero-shot ranking of binding-affinity changes.

IgDesign. IgDesign provides experimentally evaluated antibody designs generated for antigen-specific binding optimization (Shanehsazzadeh et al., 2023). We use its binding-related variant set as an additional antibody design benchmark. After filtering, the benchmark contains 7 seeds and 310 variants, with mutation counts ranging from single mutations to variants carrying up to 7 mutations. This dataset complements SKEMPI-V2 and AB-Bind by evaluating model performance on designed antibody variants rather than only naturally or experimentally curated mutation scans.

Hutchinson2023. Hutchinson2023 contains experimentally characterized antibody variants designed to improve thermostability and related developability properties (Hutchinson et al., 2024). We use 2 seeds and 216 variants, covering both heavy- and light-chain mutations with up to 7 mutations per variant. This benchmark provides a larger thermostability evaluation set for testing whether alignment affects antibody stability prediction.

BC-T07. BC-T07 is an internal antibody developability benchmark that is not publicly released. We use two subsets from this benchmark: a thermostability subset with 1 seed and 140 variants, and an expression subset with 1 seed and 48 variants. Both subsets contain mutations on heavy and light chains, with up to 9 mutations in the thermostability subset and up to 10 mutations in the expression subset.

Koenig2017. Koenig2017 is a deep mutational scanning benchmark of antibody variable-domain variants, measuring how mutations affect antibody expression and related functional properties (Koenig et al., 2017). We use its expression subset, which contains 1 seed and 4,274 variants with mutations on both heavy and light chains. Due to its large number of variants, Koenig2017 serves as the main public benchmark for zero-shot antibody expression prediction.

Adam2017. Adam2017 is an antibody expression benchmark based on a large mutational library (Adams et al., 2016). We include 1 seed with 10,857 variants, covering both heavy- and light-chain mutations with up to 3 mutations per variant. This dataset provides an additional large-scale expression benchmark for evaluating whether model scores can rank sequence variants by expression-related fitness.

A.6 Chain Handling in Zero-shot Evaluation

For zero-shot evaluation, paired heavy- and light-chain variable regions are concatenated and scored as a single sequence, so mutations on either chain are evaluated in the full paired context. For chain-

specific mutations, only the mutated positions change and the partner chain stays fixed. For datasets with only one available chain, we score that variable region alone. Chain boundaries are preserved in the input representation.

A.7 Ablation Studies

Impact of continuous pre-training on ESM2. The main-text CP ablation uses ProteinReasoner as the base model (Table 4). As a cross-backbone robustness check, Table S7 repeats the same comparison with ESM2, which is continuously pre-trained on the same OAS-derived lineages as AbEvolve while keeping the masked-language-modeling objective. ESM2-GFlowNet trained at the same scale improves all eight benchmarks over CP.

Table S7: Continuous pre-training ablation for ESM2. CP continues the masked-language-modeling objective on the same OAS-derived lineages used by AbEvolve, without trajectory or abundance supervision. Bold indicates the best of the three variants on each benchmark. Spearman ρ is reported.

	Binding			Thermostability		Expression		
Method	SKEMPI-V2	AB-Bind	IgDesign	Hutchinson2023	BC-T07†	Koenig2017	BC-T07†	Adam2017
ESM2-Vanilla	0.075	0.122	0.004	0.260	0.164	0.360	0.418	0.147
ESM2-CP	0.114	0.108	0.023	0.343	0.394	0.259	0.416	0.232
ESM2-GFlowNet	**0.214**	**0.196**	**0.250**	**0.418**	**0.447**	**0.399**	**0.453**	**0.304**

Note. † denotes an internal benchmark not publicly released.

Impact of lineage scale. Table S8 reports the lineage-scale ablation. All rows use GFlowNet alignment with FGMO ordering and correct abundance labels; only the number of largest lineages used for alignment varies. Scaling from 2.6K to 260K raises the binding average (0.216 vs. 0.263) and the expression average (0.331 vs. 0.369); thermostability stays at 0.520 (and dips to 0.507 at 26K). Per-benchmark changes are not strictly monotonic (e.g., AB-Bind is highest at 26K).

Table S8: Lineage-scale ablation. All rows use GFlowNet alignment with FGMO ordering and correct abundance labels; only the number of largest lineages used for alignment varies. Thermostability and expression are averages over their benchmarks. Spearman ρ is reported.

	Binding			Average	
Scale	SKEMPI-V2	AB-Bind	IgDesign	Thermostability	Expression
2.6K	0.208	0.256	0.183	**0.520**	0.331
26K	0.214	**0.277**	0.202	0.507	0.339
260K	**0.257**	0.273	**0.258**	**0.520**	**0.369**

Note. Thermostability and Expression columns report category averages (Hutchinson2023 and BC-T07; Koenig2017, BC-T07, and Adam2017, respectively). Bold indicates the best result per benchmark.

A.8 Redundancy as a Structured Weak Label: Start-Sequence Sensitivity

The main-text shuffle ablation (Section 4.3, Table 5) reports seed re-selection after permuting redundancy labels. Here we record the remaining construction details. From the same OAS processed lineage pool (19,255 filtered CSVs), the two datasets use identical filtering, CSV pool, and global ranking (top-2,600 by `num_members`); they differ only in whether Redundancy(R) is original or permuted within each lineage.

Under canonical min-R, 2,600 trajectory seeds collapse to 593 unique amino-acid sequences, despite 19k independent lineage CSVs. After R permutation, the same pipeline yields 2,579 / 2,600 (99.2%) unique starts, a 4.4× increase in seed diversity, and the two datasets share only 25 of 2,600 seeds (Jaccard 0.0076). Trajectory viability is nevertheless preserved: 85,507 candidate payloads were available before the top-2,600 cut, and the final shuffled-R dataset contains 1,468,136 trajectory members across 2,600 seeds.

A.9 Hyperparameters

Tables S9–S11 summarize the hyperparameters used for alignment. Shared settings across all objectives are listed in Table S9, continuous pre-training specifics in Table S10, and GFlowNet-specific settings in Table S11.

Table S9: Shared hyperparameters for model alignment.

Setting	Value	Description
Optimizer	AdamW	Optimizer for all alignment objectives (Loshchilov & Hutter, 2017)
β_1	0.9	first moment decay rate
β_2	0.99	second moment decay rate
Learning rate	1e-4	backbone learning rate
Head learning rate	1e-4	action-head learning rate
Batch size	16	number of sequences or trajectories per batch
Epochs	20	maximum training epochs
Weight decay	0.01	AdamW weight decay
Gradient clipping	0.1	maximum gradient norm

Table S10: Hyperparameters for continuous pre-training.

Hyperparameter	Value
Training objective	MLM / Autoregressive
Learning rate	1e-5
Batch size	32
Epochs	20
Masking ratio	0.15
Weight decay	0.01

Table S11: Hyperparameters for GFlowNet-based models.

Hyperparameter	Value
Training objective	trajectory balance
Reward	$\log A(\mathbf{s}_i) - \log A(\mathbf{s}_{\text{seed}})$ (0-1 normalized)
Learning rate	1e-4
Batch size	16
Epochs	20
Reward clipping / normalization	0-1 Normalization
Trajectory sampling	FGMO paths
Inference score	forward path likelihood