MULTI-OBJECTIVE AMORTIZED CDRH3 SEQUENCE SAMPLING DIVERSIFIES PROMISING ANTIBODY CANDIDATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Antibody therapeutics have transformed treatment across diverse therapeutic areas and have become central to modern medicine. While recent machine learning approaches aim to accelerate antibody discovery through structure-based design and target-affinity optimization, successful candidates must also satisfy broad developability constraints, including structural compatibility, low immunogenicity, and favorable biophysical properties, while remaining sufficiently diverse to provide sufficient coverage of design space to mitigate later-stage development failures. We introduce a preference-conditioned multi-objective generative flow network (GFN) fine-tuning pipeline for fixed-framework CDRH3 loop redesign that fine-tunes a biological sequence model against many inexpensive proxies covering backbone compatibility, sequence likelihood, immunogenicity risk, novelty, and developability liabilities. It combines up to fourteen predictive rewards in a single policy trained to sample CDRH3 loops across 623 SAbDab antibody structures, leveraging a reward reduction strategy to preserve Pareto dominance under this learned map. We also introduce a novel autoregressive span-in-filling variant of AMPLIFY-350M that serves both as a scaffold-conditioned CDRH3 generation prior and as a direct conditional likelihood scorer. We evaluated reward scores, diversity, and design quality on loops sampled on 45 held-out antibody complexes to assess structural consistency, confidence, aggregation propensity, stability, and antigen binding. Compared with the inverse-folding methods ProteinMPNN and ESM-IF, our approach produces roughly $10\times$ more distinct acceptable candidates per design at 50% sequence identity, while holding 10% of the 14-objective Pareto front pooled across methods, trading per-sample quality for greater exploration of novel possible structures. The results suggest that by relying on a more diverse data distribution and reward distribution, we are able to generate a wider array of promising candidates. All code and data will be made publicly available upon publication.

1 INTRODUCTION

Antibody therapeutics are a major drug modality across therapeutic areas, with more than 200 antibodies and antibody-like biologics now approved (Chan et al., 2025). Current antibody discovery still relies heavily on experimental screening of large libraries, followed by extensive engineering and optimization (Laustsen et al., 2021). Yet such libraries sample only a small fraction of the vast antibody sequence space; a typical 13 residue loop alone allows $20^{13} \approx 8.2 \times 10^{16}$ possible sequences. Computational and machine learning-driven approaches increasingly aim to direct this search in high dimensional space (Laustsen et al., 2021; Bennett et al., 2026; Shanehsazzadeh et al., 2024; Nabla Bio & Biswas, 2025; Agarwal et al., 2025), focusing primarily on affinity, structural plausibility, or sequence likelihood, with developability usually addressed later in development. However, binding rank-based methods can perform at parity with or worse than random mutational screens (Erasmus et al., 2026), motivating a broader exploration of sequence design space. Moreover, antibody design is inherently multi-objective: candidates must balance structural compatibility with low immunogenicity and favorable biophysical and developability properties (Khetan et al., 2022; Xu et al., 2019; Zhang et al., 2023b; Raybould et al., 2019). These objectives can conflict (Adams et al., 2001) and maintaining diversity provides alternatives when otherwise promising candidates fail later

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.

in development. This is particularly important, as validation studies repeatedly show that once in-silico structural or binding scores exceed an acceptable quality threshold, further score improvements do not necessarily translate into a better candidate ranking (Chu et al., 2026; Cotet et al., 2025), motivating the consideration of a broader suite of design criteria when selecting candidates to progress to the lab for further therapeutic development (Overath et al., 2025).

This paper approaches fixed-framework redesign of antibody heavy-chain complementarity-determining region 3 (CDRH3) loops as a sequence generation problem in the context of a fixed structural scaffold. CDRH3 is a major determinant of antigen binding and specificity due to its central position in the binding interface and high sequence diversity (Xu & Davis, 2000; McConnell & Casadevall, 2025)[1], Each design replaces CDRH3 within the fixed framework of a known antibody-antigen crystal complex, avoiding the need for antibody-antigen pose search required in many *de novo* design settings. We adopt Multi Objective Generative Flow Networks (MOGFNs; Jain et al. (2023)) to train a (multi-) preference conditioned CDRH3 sequence sampler across 623 experimentally-derived antibody-antigen complexes, enabling the sampling of diverse, high-reward designs representing different trade-offs along the Pareto front. We use this approach to fine-tune a new span-in-filling variant of AMPLIFY-350M (Fournier et al., 2024), conditioned on both scaffold and preference. Our oracle-agnostic framework supports additional or alternative oracles without changing the generator architecture, and we demonstrate this flexibility by training with up to $M = 14$ predictive rewards. To avoid the combinatorial training and deployment burden of sampling preferences across a growing M-simplex, we learn a dominance-preserving, strictly positive, row-stochastic projection from the M normalized rewards onto $D \ll M$ preference slots (Park & Sung, 2025).

We evaluate reward performance, diversity, and structural quality on 45 held-out antibody complexes. In sum, we contribute:

1. A preference-conditioned MOGFN fine-tuning pipeline for fixed-framework CDRH3 redesign, optionally incorporating a learned preference projection to support many rewards.
2. A modular suite of fourteen antibody-design oracles covering inverse-folding backbone compatibility, general and antibody-specific sequence likelihood, immunogenicity risk, novelty, and developability liabilities.
3. An autoregressive span-in-filling version of AMPLIFY-350M, serving as a fast and memory-efficient scaffold-conditioned CDRH3 generation prior.
4. A structural quality-diversity evaluation pipeline designed to produce a Chai-1 (Chai Discovery, 2024) and Rosetta (Alford et al., 2017) ranked shortlist of promising CDRH3 loop candidates using estimator scores weakly correlated with the training rewards.

2 Related Work

Antibody and binder design. Computational antibody design spans structure-based de novo design (Bennett et al., 2026; Stark et al., 2025), antigen-specific generators (Shanehsazzadeh et al., 2024; Nabla Bio & Biswas, 2025; Levine et al., 2026; Mille-Fragoso et al., 2026; Wasdin et al., 2025), affinity prediction and optimization (Agarwal et al., 2025), sequence infilling (Shuai et al., 2023), and antibody-specific foundation and inverse-folding models (Guloglu et al., 2025; Dreyer et al., 2023). These systems typically optimize binding, sequence likelihood, or structure recovery and often rely on structural information, which covers only a small fraction of antibody sequence space: SAbDab contains $\approx 10^4$ antibody–antigen structures (Schneider et al., 2022), compared with $> 10^9$ observed antibody sequences in OAS (Olsen et al., 2022). In contrast, we train a single CDRH3 generator across many scaffolds using structural priors, sequence likelihood, developability, immunogenicity, and novelty as rewards. The resulting model generalizes to new scaffolds at test time, unlike antigen-specific generators that treat each target as a separate design problem (Shanehsazzadeh et al., 2024; Nabla Bio & Biswas, 2025; Levine et al., 2026; Mille-Fragoso et al., 2026; Wasdin et al., 2025). Previous work has also cast antibody design as a multiobjective problem. Makowski et al. (2022) co-optimized affinity against non-specific binding across a CDR mutational library; Tagasovska et al. (2025) affinity and expression from limited assay data; others decomposed interface-energy terms

[1] Antibody variable domains are organized around conserved framework regions, which provide the structural scaffold for six hypervariable complementarity-determining regions (CDRs) responsible for antigen recognition and binding. Here, we focus on CDRH3, which typically makes a major contribution to antigen binding.

using diffusion-based sequence-structure co-design (Wen et al., 2026) or sequential design (Feng et al., 2025); and Ren et al. (2025) Rosetta binding energy and evolutionary plausibility while treating non-specific binding, self-association and stability as constraints. These approaches use at most a handful (2-5) and largely emphasize binding energy, whose predictive accuracy remains limited for mutated antibody-antigen interfaces (Sirin et al., 2016), and out-of-domain antigens (Hummer et al., 2025). We instead train a single multi-scaffold (623 bound SAbDab complexes), multi-task sampler to *explore* the sequence space of promising CDRH3 variants over a broader suite of up to fourteen objectives.

GFlowNets. Generative Flow Networks learn policies whose terminal states are sampled proportional to a positive reward (Bengio et al., 2021), preserving diversity among high-reward candidates. Multi-objective GFNs (Jain et al., 2023) and HN-GFN (Zhu et al., 2023) extend this idea to Pareto-optimal generation under competing objectives. Yin et al. (2025) leveraged a GFN to sample high likelihood sequences under a product of experts (ProGen2 fromNijkamp et al. (2023), and a Potts model, respectively providing global protein sequence distribution and local antibody-domain distributions) to generate diverse CDR candidates. Instead, we condition on preferences over up to fourteen rewards to take advantage of the same diversity-seeking behaviour of GFNs.

Objective reduction. Multi-objective optimization often relies on objective reduction to accommodate many rewards, via methods such as PCA (Deb & Saxena, 2006) or dominance-preserving reduction with an explicit δ-error (Brockhoff & Zitzler, 2006; 2009), motivated by the observation that Pareto dominance degenerates as the objective count grows. Recent work in multi-objective reinforcement learning (MORL; Park & Sung (2025)) showed that learning to sample from the Pareto front of an affine, all-positive, row-stochastic reward reduction provably produces a policy that samples from regions of the non-reduced objective set Pareto front. We adopt this method and transfer it from value-based MORL to a preference-conditioned MOGFN which targets the whole reward distribution, and reward compression therefore trades against both diversity and optimality.

Structural filtering of designs. Forward-folding self-consistency and co-folding are standard practices for validating designed sequences; we rely on both an antibody-specific structure predictor ABodyBuilder3 (Kenlay et al., 2024) and the general structure prediction model Chai-1 (Chai Discovery, 2024), which does not require multiple sequence alignment. Fold confidence and empirical, energy-based binding estimators like FoldX (Delgado et al., 2025) and Rosetta (Alford et al., 2017) are useful as binary gates for selecting promising binders for experimental validation, although their ranking performance remains limited (Pacesa et al., 2025; Bennett et al., 2023; Sirin et al., 2016; Hummer et al., 2025).

3 METHODS

3.1 TASK OVERVIEW

We denote a candidate CDRH3 amino acid sequence by $\mathbf{x} = (x_1, \ldots, x_L)$, where L is the CDRH3 loop length. For a fixed antibody scaffold s, we redesign only the CDRH3 loop while fixing the surrounding framework context. Each design task is a multi-objective optimization problem which seeks feasible candidates $\mathbf{x}^\star \in \mathcal{X}$ that jointly maximize M objectives $R(\mathbf{x}) = [R_1(\mathbf{x}), \ldots, R_M(\mathbf{x})]$ i.e., $\max_{\mathbf{x}\in\mathcal{X}} R(\mathbf{x})$. When objectives conflict, no single candidate simultaneously maximizes all objectives, so we seek Pareto-optimal candidates where improving one objective dimension would compromise another one.

Our method introduces four components: 1) a modular reward suite for antibody design (Section 3.2), 2) a learned reward projection onto a fixed-size preference simplex (Section 3.2), 3) a scaffold-conditioned span-in-filling policy that proposes CDRH3 sequences in framework context (Section 3.3), and 4) a preference-conditioned GFN objective that samples diverse high-reward sequences under different tradeoffs (Section 3.4). Generated pools are then sharpened by a structure-based filtering pipeline (Section 3.5). Figure 1 shows the training framework.

3.2 REWARDS

We optimize up to $M = 14$ independent reward terms in six categories: structural stability, immunogenicity, developability, liability, likelihood, and novelty (see Table 11). All variance in the rewards

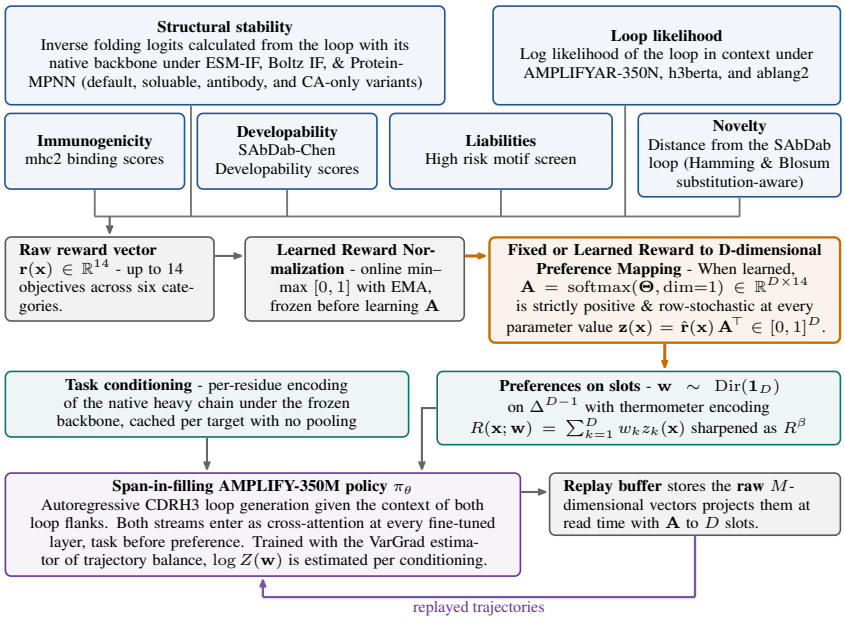


Figure 1: Training pipeline. The **14 rewards** in six categories feed one reward vector. Blue: oracles, orange: reward dimension reduction, teal: conditioning, purple: policy. Table 11 lists every reward and Section 3.2 derives the projection.

is driven by changes to the CDRH3 loop, regardless of whether the oracles required additional framework-specific context. **Structural stability** guides the sampler to produce loops with high likelihood under the common inverse-folding models ESM-IF, Boltz-IF, and ProteinMPNN (Hsu et al., 2022; Dauparas et al., 2022; Stark et al., 2025). **Immunogenicty** is measured by an MHC-II binding predictor built on a frozen ESM-2 8M backbone (Lin et al., 2023) to detect antibody designs that could trigger immune responses. **Developability** captures whether an antibody can be practically manufactured, and is measured using a fine-tuned ESM-2 650M sequence classifier (Lin et al., 2023) on SAbDab-derived binary labels (Chen et al., 2020). **Sequence liability** (chemical and post-translational) was measured using a motif scan adapted from the Liability Antibody Profiler (Satława et al., 2024). **Sequence likelihood** was estimated under our AMPLIFY-AR model, a CDRH3-specific masked language model `h3berta` (Rodella & Lemmin, 2026), and `ablang2` (Olsen et al., 2024), which scores the loop inside its target's paired variable domain and is less germline-biased. **Novelty** was incentivized using the minimum normalized Levenshtein distance and BLOSUM62 (Henikoff & Henikoff, 1992) substitution score to the training CDRH3 set, minimized over training loops of the same length. All rewards were normalized using an online min-max normalizer to facilitate multiobjective training. See appendix B.4 for details for more details.

Reduction to Slots. Covering a M-simplex to resolution ϵ requires $\binom{1/\epsilon+M-1}{M-1}$ cells, so conditioning the policy on the native preference simplex is not tractable with larger values of M. We therefore use a learned reward projection with a fixed number of *preference slots* $D \ll M$ (Park & Sung, 2025). Let $\hat{\mathbf{r}}(\mathbf{x}) \in [0,1]^M$ be the normalized reward vector. The free parameter matrix $\boldsymbol{\Theta} \in \mathbb{R}^{D\times M}$ induces $\mathbf{A} = \mathrm{softmax}(\boldsymbol{\Theta}, \mathrm{dim} = 1) \in \mathbb{R}^{D\times M}$, which allows us to calculate the reduced $\mathbf{z}(\mathbf{x}) = \hat{\mathbf{r}}(\mathbf{x})\,\mathbf{A}^\top \in [0,1]^D$, ensuring $\mathbf{A}$ is strictly positive and row-stochastic regardless of convergence[2]. The influence of this reduction on the Pareto front is analyzed in Section 4.1. Preferences are then drawn on the D-simplex, $\mathbf{w} \sim \mathrm{Dir}(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} = (1,\ldots,1) \in \mathbb{R}^D$, and the scalarized reward is a convex combination of the slot rewards $R(\mathbf{x};\mathbf{w}) = \sum_{k=1}^{D} w_k\, z_k(\mathbf{x})$. $\mathbf{A}$ is trained by reconstruction on a decoder h_ψ reward batches[3]: $\mathcal{L}_{\mathrm{rec}} = \frac{1}{BM}\sum_{b,j}\frac{1}{\sigma_j^2}\Big(h_\psi(\mathbf{z}_b)_j - \big(\hat{r}_{bj} - \mu_j\big)\Big)^2$,

[2] Since every entry of $\mathbf{A}$ is strictly positive, $\hat{\mathbf{r}}' \succeq \hat{\mathbf{r}}$ with $\hat{\mathbf{r}}' \neq \hat{\mathbf{r}}$ implies $\mathbf{A}(\hat{\mathbf{r}}' - \hat{\mathbf{r}}) > 0$ in every coordinate, so $\mathbf{z}'$ strictly dominates $\mathbf{z}$. This preserves dominance and ensures that no $\mathbf{z}$-Pareto-optimal sample is $\hat{\mathbf{r}}$-dominated.

[3] A stop gradient prevents the projection from being trained on the trajectory-balance loss to avoid a constant reward degenerate optimum.

where **A** is frozen before MOGFN training terminates so the the GFN can train against a static reward during the final stages of training (see appendix C for details).

Reward sharpening. The scalarized reward is sharpened as $R(\mathbf{x};\mathbf{w})^{\beta}$ before entering the GFN objective to set the exploitation–controllability trade-off. Reported experiments use $\beta = 8$, see appendix G.2 for details.

3.3 AMPLIFY-AR SPAN-IN-FILLING POLICY

We trained an autoregressive span-in-filling variant of AMPLIFY-350M (Fournier et al., 2024) from scratch on the original UniRef100 (Suzek et al., 2015), OAS, and SCOP (Chandonia et al., 2022) data (1,050,000 steps; 1.77×10^{11} in-filling tokens; appendix A.1) to initialize π_θ. Sequences were rearranged in FIM format:

<bos> prefix <mask> suffix <sep> infill <eos>,

and the model was trained with causal cross-entropy on the missing span. Span lengths were sampled uniformly from 1–50 residues, covering the 5–24-residue CDRH3 range. This enables teacher-forced log-probability calculation in one forward pass rather than the L passes required for masked-model pseudo-likelihood, yielding a 10.7× throughput improvement on held-out CDRH3 loops (see appendix A.1). **Conditioning** used sequential task and preference cross-attention after self-attention and before each conditioned layer's feed-forward network. Task conditioning used unpooled per-residue native heavy-chain representations from the frozen backbone, computed once per target. Preference weights w_d were thermometer-encoded, combined with learned slot embeddings, and projected to the AMPLIFY hidden dimension. This ordering allows the same preference trade-off to act differently across target-conditioned hidden states. Each cross-attention output projection was initialized to zero. During sampling, the model predicts each CDRH3 residue conditioned on the framework prefix and suffix, previously generated CDRH3 residues, and scaffold/preference conditioning using two sequential cross-attention modules per conditioned layer. The resulting trajectory log-probability is: $\log P_F(\tau \mid \mathbf{w}) = \sum_{t=1}^{L} \log \pi_\theta(x_t \mid \mathbf{x}_{<t}, \mathbf{x}_{\text{frame}}, \mathbf{w})$. Full conditioning and cross-attention details are provided in appendix A.4.

3.4 GFN TRAINING OBJECTIVE

We train a preference-conditioned MOGFN (Jain et al., 2023) to sample CDRH3 sequences along the Pareto frontier of the design objectives. For a fixed preference vector $\mathbf{w}$, the GFN target distribution is proportional to the scalarized reward, $R(\mathbf{x};\mathbf{w})/Z(\mathbf{w})$, where $Z(\mathbf{w})$ is the preference and context-dependent normalizing partition function. The FIM policy above defines the forward transition probability P_F: we use π_θ when emphasizing the per-step action distribution and P_F when referring to trajectory-level probabilities.

We train with the trajectory balance (TB) objective (Malkin et al., 2022). Let $P_F(\tau)$ and $P_B(\tau)$ denote the forward and backward probabilities of a complete trajectory τ that generates sequence $\mathbf{x}$. TB requires $\log Z(\mathbf{w}) + \log P_F(\tau \mid \mathbf{w}) = \log R(\mathbf{x};\mathbf{w}) + \log P_B(\tau)$, but as each autoregressive trajectory has a unique backward decomposition, $P_B(\tau) = 1$ and can be eliminated. Rather than learning a separate $\log Z(\mathbf{w})$ for each conditioning vector, we use a VarGrad-style centered-residual estimator (Richter et al., 2020; Malkin et al., 2023; Zhang et al., 2023a). With the per-sample residual $\delta_n = \log P_F(\tau_n \mid \mathbf{w}) - \log R(\mathbf{x}_n;\mathbf{w})$, the batch mean estimates $-\log Z(\mathbf{w})$ and the VarGrad loss minimizes residual variance over a batch of N trajectories:

$$\mathcal{L}_{\text{VarGrad}} = \frac{1}{N}\sum_{n=1}^{N} \ell\big(\delta_n - \bar{\delta}\big), \quad \bar{\delta} = \frac{1}{N}\sum_{n=1}^{N} \delta_n. \tag{1}$$

At the fixed point all residuals are equal and the loss is zero[4]. We take ℓ to be a log-cosh penalty, $\ell(\delta) = \log\cosh(\beta_{\text{loss}}\delta)/\beta_{\text{loss}}$, rather than a squared error, which bounds the influence of the large residuals that a sharpened ($\beta = 8$) reward and structure-to-structure variance produces during training. To further reduce structure-to-structure gradient variance, each optimizer step averages $K = 4$ independent single-structure updates. Training details are provided in appendix A.

[4] The resulting gradient can be viewed as a REINFORCE estimator (Williams, 1992) with the batch mean residual as a control variate (Malkin et al., 2023).

3.5 POST-FILTER CASCADE, BENCHMARK, & EVALUATION

Post-Filter Cascade. We calculate scores on the distribution of designs from each sampler with oracles of interest to the drug development pipeline, including liability, developability, structural plausibility, and structural compatibility. Each stage of the funnel is increasingly more computationally expensive to run, and is performed on a smaller subset of the previous stage's designs. Briefly the sampled pool undergoes a Stage 0 sequence liability motif scan, followed by a Stage 1 forward fold under ABodyBuilder3 (collecting the CDRH3 pLDDT confidence scores and scRMSD structural fidelity scores), a Stage 2 measure of self-aggregation propensity, a Stage 3 forward fold under Chai-1 (pLDDT & RMSD), followed by a Stage 4 estimation of relative binding and stability energies as measured by FoldX and Rosetta. Stage 3's budget is 1000 per arm due to computational constraints, and Stage 4's is 250 (budgets referred to as B3 and B4 in text). See appendix D for all details of the evaluation.

Dataset. Antibody-antigen complexes were curated from the Structural Antibody Database (SAbDab; Dunbar et al. (2014)) and split by heavy-chain scaffold to evaluate generalization of CDRH3 to novel antibody frameworks, yielding 623 training, 10 validation, and 55 held-out test targets. Antigen chains were used only to confirm that held-out test targets represent distinct binding problems. Each arm was evaluated at a matched sampling budget (32,768 designs per arm per target for the ranked pipeline, 100 designs per arm for the uniform pipeline). We compare three GFlowNet arms (a pre-defined 9/14-reward mapping and two learned objective-slot models with $D = 4$ and $D = 5$), the AMPLIFY-AR policy provided the loop design performance of the prior. We also evaluated a uniform random policy, and compared against two structural inverse-folding baselines: ProteinMPNN & ESM-IF. Full details are provided in appendix B.

Evaluation. We evaluate all arms both for their ability to produce high quality antibodies on the test scaffolds (as defined by both the reward set and the held out postfilter pipeline) conditions and sample diversity. On the held out scaffolds, we report oracle scores, postfilter score distribtions, distinct clusters among the acceptable designs (50% loop identity), pareto front enrichment by the postfilter pipeline, and sequence diversity in both sequence and embedding space.

4 RESULTS

4.1 LEARNED PROJECTION

An analysis of the reward landscape can be found in Section F.1. Briefly, our learned projections recover expected groupings in the data. Fitted on the native-PDB reward matrix at $D = 4$, the slots resolve to an inverse-folding stability cluster, a novelty cluster, the prior likelihood clusters with the Antibody ProteinMPNN and BoltzIF inverse fold logits, and developability & immunogenicity comprise the final axis. These groupings were not found to vary with other training hyperparameters. See Appendix G.1 for more details. We also report results for a GFN trained on 9 objectives with a fixed reduction mapping: all stability oracles collapsed into one, developability, AMPLIFY-AR likelihood, and immunogenicity.

4.2 QUALITY-DIVERSITY: MORE DISTINCT ACCEPTABLE CANDIDATES PER DESIGN

An analysis of pool-wise rewards across all arms (Appendix B.4) reveals that the GFNs sample designs far from the AMPLIFY-AR prior, and score higher on novelty, immunogeniciy, and developability than any trained model. The liability filter removes $\approx 30\%$ of sampled designs from the $D = 4$ and $D = 5$ GFNs, as well as the inverse folding baselines, compared to $\approx 50\%$ for the random baseline, AMPLIFY-AR prior, and Fixed GFN (which did not train on the liability reward). An analysis of sequence diversity at this stage of the filter reveals that the GFNs, AMPLIFY-AR prior, and random baseline retain substantially more sequence diversity as we sweep cluster thresholds until $0.8 - 0.9$. In addition, the Vendi score (Friedman & Dieng, 2023) — the effective number of distinct designs in a set, which equals 1 when every design is identical and n when all n are mutually dissimilar — shows that these methods maintain substantially more diversity in the tails of the distribution, where the strongest final designs reside. Figure 4 reports all three measures — acceptability, the cluster sweep and the Vendi score — with bootstrap intervals over targets. Each sampler's pre-and-post filter scores can be found in Figure 11.

Figure 2 shows how the diversity of our samples, measured by the Vendi score, varies with increasing the quality threshold (see also Figure 7 for baseline diversity measures at the beginning of the postfilter funnel). A clear pattern emerges: while the GFN models have lower quality scores and lower survival rates, the sequences which do survive are still more diverse than the higher quality samples from our baseline methods. This relationship between quality and diversity is made most clear in Figure 3, which shows not only that the Vendi scores are highest for the GFlowNets at the end of the postfilter funnel (even higher than the random baseline), but the distribution of survivors from the GFNs contains very promising binders. Figure 8 demonstrates that the diversity gap between the GFN and structural approaches grows as more designs are drawn after the Chai-1 and Rosetta stages. Figure 9 shows where each arm's surviving designs sit on a shared landscape, showing the GFN samples overlapping the design space occupied by structual models, our random baseline, and the AMPLIFY-AR prior.

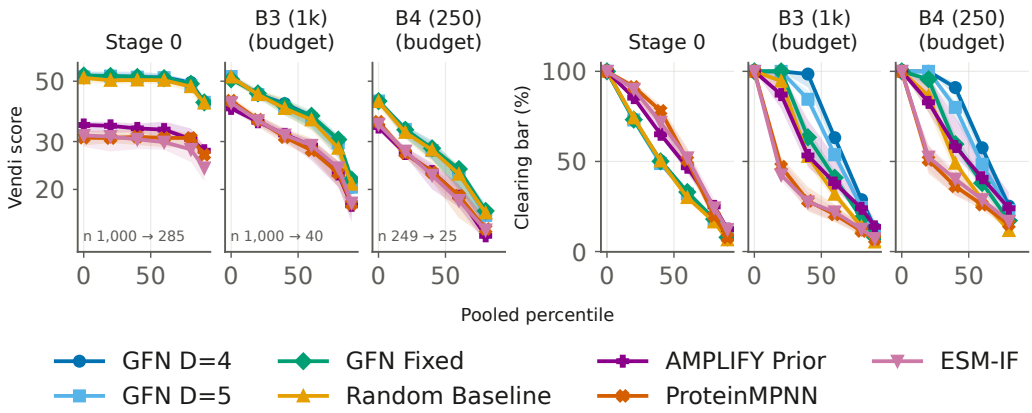


Figure 2: Diversity at matched quality per arm throughout the postfilter. Left: The Vendi score over ESM-2 650M embeddings of the loop in its own heavy-chain context, after subtracting each target's mean embedding, computed over percentile quality thresholds. Right: the fraction of each arm's designs over a percentile threshold. "Bar" represents the percentile of ABodyBuilder3 CDR-H3 pLDDT pooled over the arms within a target, Bands represent 95% bootstrap CIs resampling targets. The GFN models show increased diversity relative to the other trained baselines across thresholds.

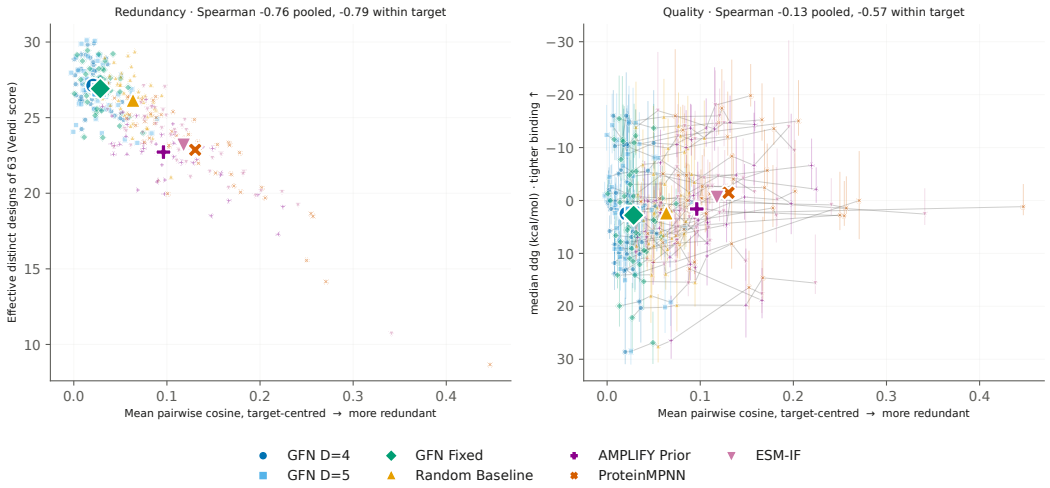


Figure 3: Sample redundancy (left) and final quality (right) among postfilter survivors ($n = 250$ per arm). Sequences embedded in ESM-2 650M. Cosine distances taken after removing each target's mean. One dot per test-target. Large markers denote the arm's mean (right). Grey lines connect the designs across all arms for a single target, and whiskers denote within-arm 5-95% CI. GFN-based models produce substantially more unique designs than all baselines in exchange for a reduction in median binding scores.

4.3 Reward Dominance

Table 1: Joint Pareto front pooled over all seven arms, per target, summed over 45 held-out targets.

Objectives	Front	% of pool	ESM-IF	ProteinMPNN	GFN D=4	GFN D=5	GFN Fixed	AMPLIFY	Random
4 (collapsed)	19,501	0.19%	27.4%	12.4%	3.8%	2.8%	6.7%	46.5%	0.4%
9 (common)	86,972	0.85%	40.6%	25.0%	2.0%	1.5%	3.5%	27.1%	0.2%
14 (full)	2,115,506	20.71%	21.9%	22.1%	9.8%	8.2%	9.7%	24.3%	4.0%

Pooled Pareto front. The number of non-dominated samples naturally grows as M grows, so we report three reductions of the Pareto front over rewards in Table 1. The GFN models claim less of the front than the prior. The prior and inverse folding models hold roughly similar proportions of it. Preference controllability — whether a design's realized rewards actually track the preference vector it was conditioned on — is reported separately in Appendix C.

4.4 Structural Validation Through the Post-Filter

The GFNs match or modestly outperform the random baseline, while trailing structural models on end-funnel quality metrics. Recall that the GFNs outperform random baselines on the other outcome metrics and pipeline diversity (Appendix B.4). Among non-dominated designs, stability oracles correlate most strongly with structural (as expected) and likelihood endpoints; novelty is most strongly anticorrelated, immunogenicity mildly so, while developability and liability are largely uncorrelated (Figure 5). CDRH3 spatial aggregation propensity (SAP) (Chennamsetty et al., 2009) across Stage-1 folds is lower for three GFN arms than all baselines in both the distribution bulk and tail (Table 2). Figure 4 further shows that structural models perform best on training-aligned rewards, whereas GFNs match or outperform baselines on novelty, immunogenicity, developability, and sequence liability.

Table 2: Design performance through the funnel across 45 held-out targets and seven arms. Stage 2 is CDRH3 spatial aggregation propensity (Chennamsetty et al., 2009) over the full pre-funnel Stage-1 population (ABodyBuilder3), where lower is better. Stage 3 reports forward fold RMSD to the crystal structure, and the CDRH3 agreement RMSD in Å from ABodyBuilder3↔ and Chai-1 (lower is beter). CDRH3 pLDDT *Retained* is the fraction of each target's pLDDT top 250 surviving the best-25% agreement threshold. *Stage 4* reports Top-1 $\Delta\Delta G$ (kcal/mol for FoldX, REU for Rosetta) and IMP, the fraction with $\Delta\Delta G < 0$.

	Stage 0 (45)	Stage 2 — CDR-H3 SAP (45)			Stage 3 — fold agreement (45)				Stage 4 — binding $\Delta\Delta G$ (45)				
Arm	Surv.	Median	Best 10%	Won	RMSD (Å)	pLDDT	ρ	Retained	FoldX	Rosetta	FoldX IMP	Rosetta IMP	ρ
GFN D=4	71.1%	−0.0088	−0.0877	0/45	3.54	77.7	−0.351	40.6%	−4.96	−8.88	23.1%	17.2%	+0.47
GFN D=5	68.3%	−0.0116	−0.0886	0/45	3.65	77.1	−0.340	40.2%	−5.50	−9.73	24.0%	17.5%	+0.46
GFN Fixed	47.1%	−0.0200	−0.0959	40/45	3.55	77.6	−0.359	40.1%	−4.78	−9.31	22.7%	16.8%	+0.46
Random baseline	46.0%	+0.0177	−0.0619	0/45	3.87	76.0	−0.283	39.6%	−5.48	−9.34	24.0%	18.0%	+0.47
AMPLIFY prior	80.1%	+0.0271	−0.0443	0/45	2.94	80.4	−0.337	36.5%	−5.09	−9.34	24.2%	19.6%	+0.47
ProteinMPNN	68.7%	+0.0138	−0.0514	2/45	3.02	82.0	−0.331	35.0%	−6.20	−14.52	30.4%	24.6%	+0.52
ESM-IF	72.1%	+0.0096	−0.0587	3/45	3.18	80.6	−0.297	35.9%	−6.09	−13.07	27.7%	23.3%	+0.50

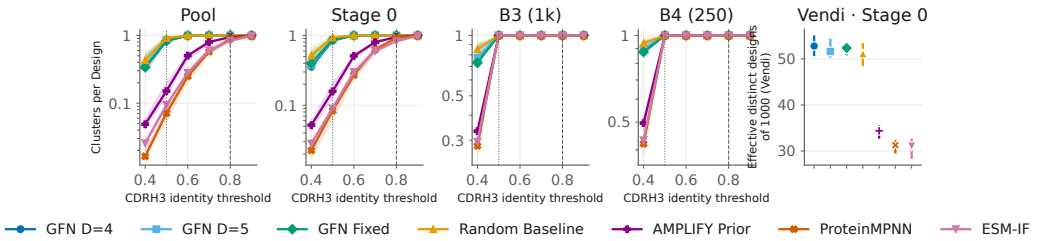


Figure 4: Design diversity throughout the post-filter pipeline across clustering thresholds. Dotted line at 0.50 marks the Stage-3 ranking threshold, & dashed line at ≥0.80 marks the clonotyping reference for shared B-cell lineage. At earlier funnel stages, the three GFN arms and the random baseline retain more clusters at permissive thresholds up to the clonotyping threshold, while the AMPLIFY-AR prior tracks the inverse-folding baselines. Rightmost panel: Vendi score over ESM-2 650M embeddings of CDRH3 in heavy-chain context after subtracting each target's mean embedding. Bands show 95% bootstrap CIs over resampling targets.

Both inverse-folding baselines and the AMPLIFY prior exceed all three GFN arms in forward-fold loop confidence, while all arms exceed the random baseline (Figure 12). It is important to note that

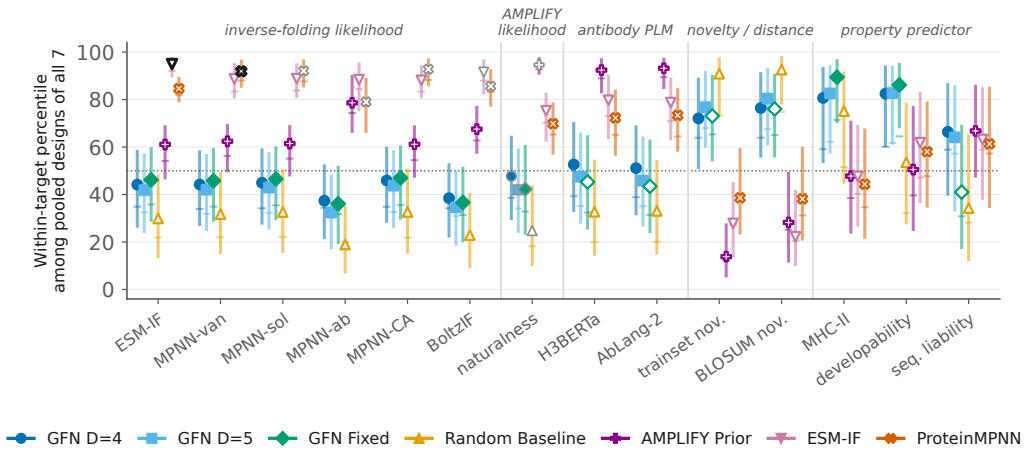


Figure 5: Non-dominated designs per generator. Markers show the median within-target percentile of each arm's front members per objective, bars span the interquartile range. Inverse-folding models favor non-dominated structural quality, whereas GFNs favor novelty while satisfying liability, developability, and immunogenicity objectives. Filled markers denote objectives used to train that arm; black rings indicate evaluation on a method's own outputs. Ticks denote whole-pool medians, with the tick-to-marker gap indicating distance to the front.

forward-fold confidence is useful as a binary binder-rejection gate but not as a binding ranker (Bennett et al., 2026). Inverse-folding baselines also lead on Rosetta and FoldX endpoints (Table 2). Rosetta Top-1 $\Delta\Delta G$ is -14.52 for ProteinMPNN and -13.07 for ESM-IF, versus -9.73 to -8.88 for the three GFN arms. The FoldX margin is narrower (-6.20 and -6.09 versus -4.78 to -5.50) but consistent; the GFN arms do not outperform random on this measure. This advantage is expected because the energy functions assess sequence–backbone compatibility, closely aligned with the inverse-folding training objective, whereas the GFNs' inverse-folding logits provide only a rough proxy. The GFNs instead distinguish themselves in diversity at matched acceptability (Section 4.2). The Rosetta and FoldX pipelines largely corroborate each other.

5 Limitations & Conclusions

All experiments are computational and rely on learned or sequence and structure-derived oracles which may not reflect experimental outcomes. Generated candidates therefore require experimental validation before their potential therapeutic relevance can be assessed. Our sampler is also restricted to a fixed CDRH3 length per scaffold to support the inverse-folding oracles, substantially reducing the accessible sequence space. Future work could enable variable-length generation using alternative structural priors and potentially more accurate, albeit considerably more expensive, oracle evaluations.

Our method frames fixed-framework CDRH3 redesign as preference-conditioned multi-objective generation followed by structural filtering. By fine-tuning a span-in-filling AMPLIFY backbone against fourteen modular oracles, we generate substantially more distinct acceptable candidates than inverse-folding baselines and explore regions of protein design space not recovered by structure-based approaches, at the cost of lower average per-sample quality. The practical value of this quality-diversity trade-off requires experimental validation. However, given that many functional proteins exist outside the sequence space represented in the UniRef90 and MGnifyp protein sequence databases (Diplock et al., 2026), sequence-based generators may provide a complementary approach for exploring novel functional protein sequence space. Our framework can incorporate additional objectives, subject to oracle and computational cost, while the projection mechanism prevents the cost of covering preference space from growing with the number of objectives.

6 Ethics Statement

This work uses public protein sequence and structure data and does not involve human subjects, private personal data, or clinical deployment. The remaining risk is dual-use: generated binders should not be synthesized, tested, or deployed without appropriate biosafety review, target-risk assessment, experimental validation, and regulatory oversight. Our models inherit biases from SAbDab, UniRef100, and the learned reward models, including biases toward well-studied antibody families, antigens, and assay conditions. This work should not be considered safe to use in any autonomous discovery program.

7 Reproducibility Statement

We will release all code, including dataset curation scripts, model training code, oracle training code, and the postfilter pipeline scripts, as a open-source monorepo upon publication. We will additionally release all oracles, the FIM AMPLIFY-AR model, and fine-tuned GFNs as open source checkpoints. The release will also include the SAbDab train and held-out splits, oracle preprocessing code, and method for generating all cached training artifacts (e.g., inverse-folding logits) in a reproducible manner. We will additionally release training configurations, random seeds, evaluation scripts, and figure-generation scripts. These artifacts are intended to allow the community to inspect the work, improve the framework within the domain of antibodies, as well as adapt it to other biological applications.

8 AI Assistance Disclosure

Large language models (LLMs) were used to assist with manuscript drafting (primarily consistency checks, accuracy of claims validation, and LaTeX automation), and for code generation for all model evaluations. Many elements of this codebase predate the advent of human-level coding agents, so most core methods and oracles were implemented entirely by hand, with some assistance on individual methodological improvements. LLMs were used under heavy supervision for figure and table generation. They were not used for research ideation, method design, or experiment design. The authors reviewed all code underlying reported results and re-wrote any generated manuscript text by hand.

REFERENCES

Gregory P Adams, Robert Schier, Adrian M McCall, Heidi H Simmons, Eva M Horak, R Katherine Alpaugh, James D Marks, and Louis M Weiner. High affinity restricts the localization and tumor penetration of single-chain fv antibody molecules. *Cancer research*, 61(12):4750–4755, 2001.

Jared Adolf-Bryfogle, Oleks Kalyuzhniy, Michael Kubitz, Brian D. Weitzner, Xiaozhen Hu, Yumiko Adachi, William R. Schief, and Roland L. Dunbrack. RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLOS Computational Biology*, 14(4):e1006112, 2018. doi: 10.1371/journal.pcbi.1006112.

Aditya A Agarwal, James Harrang, David Noble, Kerry L McGowan, Adrian W Lange, Emily Engelhart, Miranda C Lahman, Jeffrey Adamo, Xin Yu, Oliver Serang, Kyle J Minch, Kimberly Y Wellman, David A Younger, Randolph M Lopez, and Ryan O Emerson. AlphaBind, a domain-specific model to predict and optimize antibody-antigen binding affinity. *MAbs*, 17(1):2534626, 2025. doi: 10.1080/19420862.2025.2534626.

Bissan Al-Lazikani, Arthur M. Lesk, and Cyrus Chothia. Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology*, 273(4):927–948, 1997. doi: 10.1006/jmbi.1997.1354.

Rebecca F Alford, Andrew Leaver-Fay, Jeliazko R Jeliazkov, Matthew J O'Meara, Frank P DiMaio, Hahnbeom Park, Maxim V Shapovalov, P Douglas Renfrew, Vikram K Mulligan, Kalli Kappel, Jason W Labonte, Michael S Pacella, Richard Bonneau, Philip Bradley, Roland L Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J Gray. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.*, 13(6): 3031–3048, 2017. doi: 10.1021/acs.jctc.7b00125.

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Nathaniel R. Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, Frank DiMaio, Steven De Munck, Savvas N. Savvides, and David Baker. Improving de novo protein binder design with deep learning. *Nature Communications*, 14(1):2625, 2023. doi: 10.1038/s41467-023-38328-5.

Nathaniel R Bennett, Joseph L Watson, Robert J Ragotte, Andrew J Borst, Déjenaé L See, Connor Weidle, Riti Biswas, Yutong Yu, Ellen L Shrock, Russell Ault, Philip J Y Leung, Buwei Huang, Inna Goreshnik, John Tam, Kenneth D Carr, Benedikt Singer, Cameron Criswell, Basile I M Wicky, Dionne Vafeados, Mariana Garcia Sanchez, Ho Min Kim, Susana Vázquez Torres, Sidney Chan, Shirley M Sun, Timothy T Spear, Yi Sun, Keelan O'Reilly, John M Maris, Nikolaos G Sgourakis, Roman A Melnyk, Chang C Liu, and David Baker. Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*, 649(8095):183–193, 2026. doi: 10.1038/s41586-025-09721-5.

Dimo Brockhoff and Eckart Zitzler. Are all objectives necessary? on dimensionality reduction in evolutionary multiobjective optimization. In *Parallel Problem Solving from Nature (PPSN IX)*, volume 4193 of *Lecture Notes in Computer Science*, pp. 533–542. Springer, 2006. doi: 10.1007/11844297_54.

Dimo Brockhoff and Eckart Zitzler. Objective reduction in evolutionary multiobjective optimization: Theory and applications. *Evolutionary Computation*, 17(2):135–166, 2009. doi: 10.1162/evco.2009.17.2.135.

Chai Discovery. Chai-1: Decoding the molecular interactions of life. *bioRxiv*, 2024. doi: 10.1101/2024.10.10.615955.

Andrew C Chan, Greg D Martyn, and Paul J Carter. Fifty years of monoclonals: the past, present and future of antibody therapeutics. *Nat. Rev. Immunol.*, 25(10):745–765, 2025. doi: 10.1038/s41577-025-01207-9.

John-Marc Chandonia, Lindsey Guan, Shiangyi Lin, Changhua Yu, Naomi K. Fox, and Steven E. Brenner. SCOPe: improvements to the structural classification of proteins – extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Research*, 50(D1):D553–D559, 2022. doi: 10.1093/nar/gkab1054.

Lin Chen, Samuel Drapeau, Fanghao Shao, Xuekai Zhu, Bo Xue, Yunchong Song, Mathieu Laurière, and Zhouhan Lin. Controlling exploration–exploitation in GFlowNets via markov chain perspectives, 2026. URL https://arxiv.org/abs/2602.01749.

Xingyao Chen, Thomas Dougherty, Chan Hong, Rachel Schibler, Yi Cong Zhao, Reza Sadeghi, Naim Matasci, Yi-Chieh Wu, and Ian Kerman. Predicting antibody developability from sequence using machine learning. bioRxiv 2020.06.18.159798, 2020.

Naresh Chennamsetty, Vladimir Voynov, Veysel Kayser, Bernhard Helk, and Bernhardt L Trout. Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences*, 106(29):11937–11942, 2009.

Cyrus Chothia and Arthur M. Lesk. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology*, 196(4):901–917, 1987. doi: 10.1016/0022-2836(87)90412-8.

Lee-Shin Chu, Jeff Vogt, Michael Chungyoun, and Jeffrey J. Gray. Protaff: Protein binding affinity prediction via lora-finetuned esm-2. *bioRxiv*, 2026. doi: 10.64898/2026.06.13.732058. URL https://www.biorxiv.org/content/early/2026/06/13/2026.06.13.732058.

Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J. L. de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009. doi: 10.1093/bioinformatics/btp163.

Tudor-Stefan Cotet, Igor Krawczuk, Filippo Stocco, Noelia Ferruz, Anthony Gitter, Yoichi Kurumida, Lucas de Almeida Machado, Francesco Paesani, Cianna N. Calia, Chance A. Challacombe, Nikhil Haas, Ahmad Qamar, Bruno E. Correia, Martin Pacesa, Lennart Nickel, Kartic Subr, Leonardo V. Castorina, Maxwell J. Campbell, Constance Ferragu, Patrick Kidger, Logan Hallee, Christopher W. Wood, Michael J. Stam, Tadas Kluonis, Süleyman Mert Ünal, Elian Belot, Alexander Naka, and Adaptyv Competition Organizers. Crowdsourced protein design: Lessons from the adaptyv egfr binder competition. *bioRxiv*, 2025. doi: 10.1101/2025.04.17.648362. URL https://www.biorxiv.org/content/early/2025/04/24/2025.04.17.648362.

Tri Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. FlashAttention: Fast and memory-efficient exact attention with IO-awareness. In *Advances in Neural Information Processing Systems 35*, volume 35, pp. 16344–16359. Curran Associates, Inc., 2022. doi: 10.52202/068431-1189.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Alexis Courbet, Rob J. de Haas, Neville Bethel, Philip J. Y. Leung, Timothy F. Huddy, Sam Pellock, Doug Tischer, Frederick Chan, Brian Koepnick, Hannah Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Neil P. King, and David Baker. Robust deep learning–based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022. doi: 10.1126/science.add2187.

Kalyanmoy Deb and Dhish Kumar Saxena. Searching for pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)*, pp. 3352–3360, 2006.

Javier Delgado, Raul Reche, Damiano Cianferoni, Gabriele Orlando, Rob van der Kant, Frederic Rousseau, Joost Schymkowitz, and Luis Serrano. FoldX force field revisited, an improved version. *Bioinformatics*, 41(2):btaf064, 2025. doi: 10.1093/bioinformatics/btaf064.

Nathan Diplock, Thomas Desautels, Riyaz Haque, Micaela Finney, Daniel J. Wackelin, Christina S. Kang-Yun, Dan M. Park, and Sean P. Leonard. Unobserved sequence space has many functional proteins. *bioRxiv*, 2026. doi: 10.64898/2026.08.24.746828. URL https://www.biorxiv.org/content/early/2026/08/26/2026.08.24.746828.

Frédéric A. Dreyer, Daniel Cutting, Constantin Schneider, Henry Kenlay, and Charlotte M. Deane. Inverse folding for antibody sequence design using deep learning. In *ICML Workshop on Computational Biology*, 2023.

James Dunbar and Charlotte M. Deane. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*, 32(2):298–300, 2016. doi: 10.1093/bioinformatics/btv552.

James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M. Deane. SAbDab: the structural antibody database. *Nucleic Acids Research*, 42(D1):D1140–D1146, 2014. doi: 10.1093/nar/gkt1043.

M Frank Erasmus, Daniel Bedinger, Elizabeth Hopkins, Ginger Ferguson, Justine Strickler, Christilyn P Graff, Samantha R Summers, Stacy L Capehart, Joshua D Slocum, Crystal Richardson, et al. A blinded, prospective benchmark of in silico antibody discovery anchored to experimental affinity and developability. *Nature Biotechnology*, 2026. doi: 10.1038/s41587-026-03238-6.

Jake Fawkes and Jason Hartford. f-Trajectory Balance: A Loss Family for Tuning GFlowNets, Generative Models, and LLMs with Off- and On-Policy Data. In *Proceedings of the 43rd International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2026. URL https://arxiv.org/abs/2605.15417.

Hanqi Feng, Peng Qiu, Mengchun Zhang, Yiran Tao, You Fan, Jingtao Xu, and Barnabas Poczos. Learning from B cell evolution: Adaptive multi-expert diffusion for antibody design via online optimization. *arXiv [cs.LG]*, 2025. doi: 10.48550/arXiv.2508.02834.

Quentin Fournier, Robert M. Vernon, Almer van der Sloot, Benjamin Schulz, Sarath Chandar, and Christopher James Langmead. Protein language models: Is scaling necessary? bioRxiv 2024.09.23.614603, 2024.

Dan Friedman and Adji Bousso Dieng. The Vendi score: A diversity evaluation metric for machine learning. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=g97OHbQyk1.

Jason Greenbaum, John Sidney, Jolan Chung, Christian Brander, Bjoern Peters, and Alessandro Sette. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics*, 63(6):325–335, 2011. doi: 10.1007/s00251-011-0513-0.

Bora Guloglu, Miguel Bragança, Alex Graves, Scott Cameron, Timothy Atkinson, Liviu Copoiu, Alexandre Laterre, and Thomas D Barrett. AbBFN2: A flexible antibody foundation model based on bayesian flow networks. *bioRxiv*, 2025. doi: 10.1101/2025.04.29.651170.

Steven Henikoff and Jorja G. Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, 89(22):10915–10919, 1992. doi: 10.1073/pnas.89.22.10915.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8946–8970. PMLR, 2022.

Rui Hu, Yifan Zhang, Zhuoran Li, and Longbo Huang. Beyond squared error: Exploring loss design for enhanced training of generative flow networks. In *International Conference on Learning Representations*, 2025.

Alissa M. Hummer, Constantin Schneider, Lewis Chinery, and Charlotte M. Deane. Investigating the volume and diversity of data needed for generalizable antibody–antigen $\delta\delta$g prediction. *Nature Computational Science*, 5(8):635–647, 2025. doi: 10.1038/s43588-025-00823-8.

Moksh Jain, Sharath Chandra Raparthy, Alex Hernández-García, Jarrid Rector-Brooks, Yoshua Bengio, Santiago Miret, and Emmanuel Bengio. Multi-objective GFlowNets. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 14631–14653. PMLR, 2023.

Henry Kenlay, Frédéric A. Dreyer, Daniel Cutting, Daniel Nissley, and Charlotte M. Deane. ABodyBuilder3: improved and scalable antibody structure predictions. *Bioinformatics*, 40(10):btae576, 2024. doi: 10.1093/bioinformatics/btae576.

Rahul Khetan, Robin Curtis, Charlotte M Deane, Johannes Thorling Hadsund, Uddipan Kar, Konrad Krawczyk, Daisuke Kuroda, Sarah A Robinson, Pietro Sormanni, Kouhei Tsumoto, Jim Warwicker, and Andrew C R Martin. Current advances in biopharmaceutical informatics: guidelines, impact and challenges in the computational developability assessment of antibody therapeutics. *MAbs*, 14 (1), 2022. doi: 10.1080/19420862.2021.2020082.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

Timothy M. Lauer, Neeraj J. Agrawal, Naresh Chennamsetty, Kamal Egodage, Bernhard Helk, and Bernhardt L. Trout. Developability index: A rapid in silico tool for the screening of antibody aggregation propensity. *Journal of Pharmaceutical Sciences*, 101(1):102–115, 2012. doi: 10.1002/jps.22758.

Andreas H Laustsen, Victor Greiff, Aneesh Karatt-Vellatt, Serge Muyldermans, and Timothy P Jenkins. Animal immunization, in vitro display technologies, and machine learning for antibody discovery. *Trends Biotechnol.*, 39(12):1263–1273, 2021. doi: 10.1016/j.tibtech.2021.03.003.

Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, and Gérard Lefranc. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental and Comparative Immunology*, 27(1):55–77, 2003. doi: 10.1016/S0145-305X(02)00039-3.

Simon Levine, Jonathan Edward King, Jacob Stern, David Grayson, Raymond Wang, Rui Yin, Umberto Lupo, Paulina Kulyte, Ryan Matthew Brand, Tristan Bertin, Robert Pfingsten, Jovan Cejovic, Chelsea Chung, Breanna K. Luton, Andrew Hagemann, Robel Haile, Elliot Medina, Pankaj Panwar, Oleksii Dubrovskyi, Chase LaCombe, Zahra Anderson, Derrik Mildh, Scott Benjamin, Joe Kaiser, Joseph Ferron, Marta Sarrico, Alexandria Kershner, Apurva Mishra, Kai R. Ejan, Emily K. Marsh, Paul Bringas, Phetsamay Vilaychack, Kyra Chapman, Jacob Ripley, Muttappa Gowda, Kathryn M. Collins, Cailen M. McCloskey, Jeremiah S. Joseph, Rylee Ripley, Shaheed A. Abdulhaqq, David Spencer, Tiffany DeVine, Audree Feltner, Michael Guerin, Jeffrey Goby, Jesse Hendricks, Danielle Castillo, Sean McClain, Douglas Ganini, Derek Shpiel, James Mategko, Eder Cruz Garcia, Masoud Zabet-Moghaddam, John M. Sutton, Zheyuan Guo, Sean M. West, Janani S. Iyer, and Amir Shanehsazzadeh. Origin-1: a generative ai platform for de novo antibody design against novel epitopes. *bioRxiv*, 2026. doi: 10.64898/2026.01.14.699389. URL https://www.biorxiv.org/content/early/2026/05/05/2026.01.14.699389.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, 2017. doi: 10.1109/ICCV.2017.324.

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. doi: 10.1126/science.ade2574.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Emily K Makowski, Patrick C Kinnunen, Jie Huang, Lina Wu, Matthew D Smith, Tiexin Wang, Alec A Desai, Craig N Streu, Yulei Zhang, Jennifer M Zupancic, John S Schardt, Jennifer J Linderman, and Peter M Tessier. Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nat. Commun.*, 13(1), 2022. doi: 10.1038/s41467-022-31457-3.

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

Nikolay Malkin, Salem Lahlou, Tristan Deleu, Xu Ji, Edward Hu, Katie Everett, Dinghuai Zhang, and Yoshua Bengio. GFlowNets and variational inference. In *International Conference on Learning Representations*, 2023.

Scott A McConnell and Arturo Casadevall. New insights into antibody structure with implications for specificity, variable region restriction and isotype choice. *Nat. Rev. Immunol.*, 25(8):621–632, 2025. doi: 10.1038/s41577-025-01150-9.

Luis S Mille-Fragoso, Claudia L Driscoll, John N Wang, Haoyu Dai, Talal Widatalla, Jim L Zhang, Xiaowei Zhang, Bing Rao, Liang Feng, Brian L Hie, and Xiaojing J Gao. Efficient generation of epitope-targeted antibodies with Germinal. *Nature Biotechnology*, 2026. doi: 10.1038/s41587-026-03187-0.

Nabla Bio and Surojit Biswas. De novo design of epitope-specific antibodies against soluble and multipass membrane proteins with high specificity, developability, and function. *bioRxiv*, 2025. doi: 10.1101/2025.01.21.633066.

Erik Nijkamp, Jeffrey A. Ruffolo, Eli N. Weinstein, Nikhil Naik, and Ali Madani. ProGen2: Exploring the boundaries of protein language models. *Cell Systems*, 14(11):968–978.e3, 2023. doi: 10.1016/j.cels.2023.10.002.

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. doi: 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. doi: 10.1093/bioinformatics/btae618.

Max D. Overath, Andreas S. H. Rygaard, Christian P. Jacobsen, Valentas Brasas, Oliver Morell, Pietro Sormanni, and Timothy P. Jenkins. Predicting experimental success in de novo binder design: A meta-analysis of 3,766 experimentally characterised binders. *bioRxiv*, 2025. doi: 10.1101/2025.08.14.670059. URL https://www.biorxiv.org/content/early/2025/09/17/2025.08.14.670059.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, Yehlin Cho, Kourosh H. Ghamary, Laura Vinué, Brahm J. Yachnin, Andrew M. Wollacott, Stephen Buckley, Adrie H. Westphal, Simon Lindhoud, Sandrine Georgeon, Casper A. Goverde, Georgios N. Hatzopoulos, Pierre Gönczy, Yannick D. Muller, Gerald Schwank, Daan C. Swarts, Alex J. Vecchio, Bernard L. Schneider, Sergey Ovchinnikov, and Bruno E. Correia. One-shot design of functional protein binders with BindCraft. *Nature*, 646(8084):483–492, 2025. doi: 10.1038/s41586-025-09429-6.

Giseung Park and Youngchul Sung. Reward dimension reduction for scalable multi-objective reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2025.

Matthew IJ Raybould, Claire Marks, Konrad Krawczyk, Bruck Taddese, Jaroslaw Nowak, Alan P Lewis, Alexander Bujotzek, Jiye Shi, and Charlotte M Deane. Five computational developability guidelines for therapeutic antibody profiling. *Proceedings of the National Academy of Sciences*, 116(10):4025–4030, 2019.

Milong Ren, Zaikai He, and Haicang Zhang. Multi-objective antibody design with constrained preference optimization. In *International Conference on Learning Representations*, 2025.

Birkir Reynisson, Bruno Alvarez, Sinu Paul, Bjoern Peters, and Morten Nielsen. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, 48(W1): W449–W454, 2020. doi: 10.1093/nar/gkaa379.

Lorenz Richter, Ayman Boustati, Nikolas Nüsken, Francisco J. R. Ruiz, and Ömer Deniz Akyildiz. VarGrad: A low-variance gradient estimator for variational inference. In *Advances in Neural Information Processing Systems 33*, 2020.

Chiara Rodella and Thomas Lemmin. H3BERTa: A CDR-H3-specific language model for antibody repertoire analysis. *Patterns*, 7(7):101561, 2026. doi: 10.1016/j.patter.2026.101561.

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

Tadeusz Satława, Mateusz Tarkowski, Sonia Wróbel, Paweł Dudzic, Tomasz Gawłowski, Tomasz Klaus, Marek Orłowski, Anna Kostyn, Sandeep Kumar, Andrew Buchanan, and Konrad Krawczyk. LAP: Liability antibody profiler by sequence & structural mapping of natural and therapeutic antibodies. *PLOS Computational Biology*, 20(3):e1011881, 2024. doi: 10.1371/journal.pcbi.1011881.

Constantin Schneider, Matthew I. J. Raybould, and Charlotte M. Deane. SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker. *Nucleic Acids Research*, 50(D1):D1368–D1372, 2022. doi: 10.1093/nar/gkab1050.

Amir Shanehsazzadeh, Matt McPartlon, George Kasun, Andrea K. Steiger, John M. Sutton, Edriss Yassine, Cailen McCloskey, Robel Haile, Richard Shuai, Julian Alverio, Goran Rakocevic, Simon Levine, Jovan Cejovic, Jahir M. Gutierrez, Alex Morehead, Oleksii Dubrovskyi, Chelsea Chung, Breanna K. Luton, Nicolas Diaz, Christa Kohnert, Rebecca Consbruck, Hayley Carter, Chase LaCombe, Itti Bist, Phetsamay Vilaychack, Zahra Anderson, Lichen Xiu, Paul Bringas, Kimberly Alarcon, Bailey Knight, Macey Radach, Katherine Bateman, Gaelin Kopec-Belliveau, Dalton Chapman, Joshua Bennett, Abigail B. Ventura, Gustavo M. Canales, Muttappa Gowda, Kerianne A. Jackson, Rodante Caguiat, Amber Brown, Douglas Ganini da Silva, Zheyuan Guo, Shaheed Abdulhaqq, Lillian R. Klug, Miles Gander, Engin Yapici, Joshua Meier, and Sharrol Bachas. Unlocking de novo antibody design with generative artificial intelligence. *bioRxiv*, 2024. doi: 10.1101/2023.01.08.523187. URL https://www.biorxiv.org/content/early/2024/01/07/2023.01.08.523187.

Richard W Shuai, Jeffrey A Ruffolo, and Jeffrey J Gray. IgLM: Infilling language modeling for antibody sequence design. *Cell Systems*, 14(11):979–989.e4, 2023. doi: 10.1016/j.cels.2023.10.001.

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. doi: 10.1002/pro.2829.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tally Portnoi, Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein, Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang, Samuel J. Snyder, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Haoyu Fan, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M. Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi Jaakkola. BoltzGen: Toward universal binder design. *bioRxiv*, 2025. doi: 10.1101/2025.11.20.689494.

Baris E. Suzek, Yuqi Wang, Hongzhan Huang, Peter B. McGarvey, Cathy H. Wu, and the UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, 31(6):926–932, 2015. doi: 10.1093/bioinformatics/btu739.

Nataša Tagasovska, Jan Ludwiczak, Andreas Loukas, Emily K Makowski, Homa Mohammadi-Peyhani, Sai Pooja Mahajan, Karina Zadorozhny, Donald Lee, Allen Goodman, Joshua Yao-Yu Lin, Ryan Kelly, Isidro Hötzel, Jack Bevers, III, Tamica D'Souza, James T Koerber, Wei-Ching Liang, Julien Lafrance-Vanasse, Yongmei Chen, Andrew M Watkins, Henri Dwyer, Stephen Ra, Richard Bonneau, Kyunghyun Cho, and Vladimir Gligorijević. Property enhancer – a data

efficient multi-objective approach for functional antibody optimization. *bioRxiv*, 2025. doi: 10.1101/2025.10.08.681101.

Perry T Wasdin, Nicole V Johnson, Alexis K Janke, Sofia Held, Toma M Marinov, Gwen Jordaan, Rebecca A Gillespie, Léna Vandenabeele, Fani Pantouli, Olivia C Powers, Matthew J Vukovich, Clinton M Holt, Jeongryeol Kim, Grant Hansman, Jennifer Logue, Helen Y Chu, Sarah F Andrews, Masaru Kanekiyo, Giuseppe A Sautto, Ted M Ross, Daniel J Sheward, Jason S McLellan, Alexandra A Abu-Shmais, and Ivelin S Georgiev. Generation of antigen-specific paired-chain antibodies using large language models. *Cell*, 188(25):7206–7221.e16, 2025. doi: 10.1016/j.cell.2025.10.006.

Yibo Wen, Chenwei Xu, Jerry Yao-Chieh Hu, Kaize Ding, and Han Liu. Pareto-optimal energy alignment for designing nature-like antibodies. *Advances in Neural Information Processing Systems*, 38:46294–46321, 2026.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992. doi: 10.1007/BF00992696.

John L Xu and Mark M Davis. Diversity in the cdr3 region of vh is sufficient for most antibody specificities. *Immunity*, 13(1):37–45, 2000.

Yingda Xu, Dongdong Wang, Bruce Mason, Tony Rossomando, Ning Li, Dingjiang Liu, Jason K Cheung, Wei Xu, Smita Raghava, Amit Katiyar, Christine Nowak, Tao Xiang, Diane D Dong, Joanne Sun, Alain Beck, and Hongcheng Liu. Structure, heterogeneity and developability assessment of therapeutic antibodies. *MAbs*, 11(2):239–264, 2019. doi: 10.1080/19420862.2018.1553476.

Mingze Yin, Hanjing Zhou, Yiheng Zhu, Jialu Wu, Wei Wu, Mingyang Li, Kun Fu, Zheng Wang, Chang-Yu Hsieh, Tingjun Hou, and Jian Wu. Synergy of GFlowNet and protein language model makes a diverse antibody designer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22164–22172. Association for the Advancement of Artificial Intelligence (AAAI), 2025. doi: 10.1609/aaai.v39i21.34370.

David W. Zhang, Corrado Rainone, Markus Peschl, and Roberto Bondesan. Robust scheduling with GFlowNets. In *International Conference on Learning Representations*, 2023a.

Weijie Zhang, Hao Wang, Nan Feng, Yifeng Li, Jijie Gu, and Zhuozhi Wang. Developability assessment at early-stage discovery to enable development of antibody-derived therapeutics. *Antib. Ther.*, 6(1):13–29, 2023b. doi: 10.1093/abt/tbac029.

Yiheng Zhu, Jialu Wu, Chaowen Hu, Jiahuan Yan, Chang-Yu Hsieh, Tingjun Hou, and Jian Wu. Sample-efficient multi-objective molecular optimization with GFlowNets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

CONTENTS

A Implementation Details **19**

A.1 AMPLIFY-350M Autoregressive Span-In-Filling 20

A.2 MOGFN Fine-Tuning 20

A.3 Log-Cosh Residual Penalty 21

A.4 Policy Conditioning and Sampling 21

A.5 KV-Cache Sampling 22

B Data Pipeline Details **22**

B.1 SAbDab Curation 22

B.2 Inverse-Folding Logit Alignment 24

B.3 Train, Valid, and Test Splits 24

B.4 Reward Metrics Through the Funnel 25

B.5 Diversity Metrics 26

C Conditioning, Controllability, and Learned Preference Slots **26**

D Post-Filter Details **28**

E Span-In-Filling vs. Masked-Language-Model Backbone **32**

E.1 Pretraining Validation 32

F Oracle Details **35**

F.1 The Reward Landscape 35

F.2 MHC-II Binding Oracle 36

F.3 Developability Oracle 37

F.4 Naturalness and Antibody-Repertoire Oracles 38

F.5 Sequence Liability Motifs 38

F.6 Novelty Oracles 38

F.7 Stability Oracle 38

F.8 Length Dependence of Rewards 39

G Hyperparameter Tuning **39**

G.1 Learning Rate, on the Objective-Slots Arms 39

G.2 Reward Sharpening: The β Sweep 40

A Implementation Details

Structural stability. We guide our sequence sampler to produce loops with high likelihood under common inverse-folding models conditioned on the native backbone coordinates. For a candidate sequence $\mathbf{x}$ and structure s, the single-predictor stability score is the mean conditional log-probability $\widetilde{R}^{(m)}_{\text{stab}}(\mathbf{x}, s) = \frac{1}{L}\sum_{i=1}^{L} \log p_m(x_i \mid \mathbf{X}^{\text{bb}}_s)$, where $\mathbf{X}^{\text{bb}}_s$ denotes the backbone coordinates of structure s and p_m is the amino acid distribution predicted by a single inverse folding predictor. These are pre-computed from six inverse-folding model families as independent objectives and used as lookup tables during training B). See appendix F.7 for model details and validation.

Immunogenicity. We train an MHC-II binding predictor built on frozen ESM-2 8M backbone (Lin et al., 2023) to detect antibody designs that could trigger immune responses. Five models were trained via cross-validation, which we ensembled as our oracle. Each candidate CDRH3 is scored against a reference panel of 27 human leukocyte antigen (HLA) class II alleles (Greenbaum et al., 2011), and the reward penalizes the maximum predicted immunogenicity against any target $R_{\text{mhc2}}(\mathbf{x}) = 1 - \max_{h\in\mathcal{H}} f_{\text{mhc2}}(\mathbf{x}, h)$, where f_{mhc2} is the ensemble predictor, $\mathcal{H}$ is the 27-allele HLA-II panel. Architecture, training, and evaluation details are in appendix F.2.

Developability. Developability captures whether an antibody can be practically manufactured. We fine-tune an ESM-2 650M sequence classifier (Lin et al., 2023) on SAbDab-derived binary labels (Chen et al., 2020) based on the Developability Index (Lauer et al., 2012), which combines CDR spatial aggregation propensity with variable-region net charge, computed by Chen et al. (2020) over the variable region. The heavy chain with the candidate loop spliced in is scored as the developable-class probability $R_{\text{dev}}(\mathbf{x}) = \text{softmax}(f_{\text{dev}}(\mathbf{x}^{-}_{\text{frame}} \oplus \mathbf{x} \oplus \mathbf{x}^{+}_{\text{frame}}))_1$, where $\mathbf{x}^{\pm}_{\text{frame}}$ are the framework prefix and suffix and index 1 is that class. Training and cross-validation details are in appendix F.3.

Sequence liabilities. Chemical and post-translational liabilities are a distinct failure mode from aggregation-driven developability, and can be estimated directly from sequence. We score the designed loop with a motif scan adapted from the Liability Antibody Profiler (Satława et al., 2024), restricted to the designed span and covering deamidation (`NG`), isomerization (`DG`), N-linked glycosylation sequons (`N-x-[S/T]`), unpaired cysteine, oxidation-prone Met/Trp, and `DP` fragmentation $R_{\text{liab}}(\mathbf{x}) = -\sum_{u\in\mathcal{U}} \kappa_u\, n_u(\mathbf{x})$, where n_u counts occurrences of motif u and κ_u is its severity weight (hard-flagged motifs carry larger weight). See appendix F.5 for motif definitions and the full derivation.

Sequence likelihood. Three likelihood oracles score each candidate loop at distinct scopes. Our AMPLIFY-AR model is used to measure how likely a generated loop is under the pretrained span-in-filling AMPLIFY prior (Section 3.3). The candidate loop is teacher-forced after `<sep>`, and the reward is the length-normalized conditional log-likelihood $R_{\text{nat}}(\mathbf{x}) = \frac{1}{L}\sum_{i=1}^{L} \log p_{\text{FIM}}(x_i \mid \mathbf{x}_{<i}, \mathbf{x}_{\text{frame}})$, where p_{FIM} is the frozen span-in-filling model, $\mathbf{x}_{\text{frame}}$ denotes the framework prefix and suffix, and $\mathbf{x}_{<i}$ are previously generated CDRH3 residues. `h3berta` (Rodella & Lemmin, 2026) is a RoBERTa-base masked language model pretrained on over 18M isolated CDRH3 strings from healthy-donor repertoires context-free, serving as a generic CDRH3 prior. `ablang2` (Olsen et al., 2024) scores the loop inside its target's paired variable domain. Most antibody protein language models (PLMs) predict the germline residue at $> 80\%$ of somatically-mutated positions (Olsen et al., 2024), in conflict with our novelty and exploration objectives – `ablang-2` is trained with a focal loss (Lin et al., 2017) on rare non-germline predictions to avoid this. Validation details for all three are in appendix F.4.

Novelty. Two rewards incentivize moving away from the SAbDab distribution. Training set novelty is the minimum normalized Levenshtein distance to the training CDRH3 set, Blosum novelty is the analogous distance under a BLOSUM62 (Henikoff & Henikoff, 1992) substitution score, minimized over training loops of the same length. PLM embeddings are intentionally not used for this reward so that the downstream diversity analysis which depends them is not a training signal.

Normalization. As rewards span heterogeneous scales[5], we normalize each reward with a learned min-max bounds that adapt by EMA before freezing during the final training phase. The clipped normalized reward is $\hat{R}_k(\mathbf{x}) = \mathrm{clip}\Big(\frac{R_k(\mathbf{x})-b_k^-}{b_k^+-b_k^-},\, 0,\, 1\Big)$, where $b_k^- = r_k^{\min} - \epsilon_{\mathrm{m}}(r_k^{\max} - r_k^{\min})$ and $b_k^+ = r_k^{\max} + \epsilon_{\mathrm{m}}(r_k^{\max} - r_k^{\min})$[6]. The warmup, EMA, and freeze updates are specified in appendix A.4.

This appendix records implementation settings for the AMPLIFY span-in-filling backbone and AbGFN fine-tuning.

A.1 AMPLIFY-350M Autoregressive Span-In-Filling

The GFN backbone is an AMPLIFY-350M-scale transformer (Fournier et al., 2024) with hidden size $d_h = 960$ and 32 transformer layers, trained on the same data as the AMPLIFY-350M checkpoint: UniRef100 (Suzek et al., 2015), Observed Antibody Space (OAS), and SCOP (Chandonia et al., 2022), with the original AMPLIFY data versions and filtering rules, from the identical preprocessed files (2024-09-25 snapshot; 386 shards; 3,611,708 OAS and 24,495 SCOP training sequences, and ≈394M UniRef100 sequences per epoch). Every training sequence contained one interior span with a length uniformly sampled between 1-50, clamped by sequence length.This range covers the 5–24-residue CDRH3 lengths used in our experiments. The span start location was sampled uniformly from valid positions. In contrast to IgLM, no CDR-annotated or otherwise guided spans were used. The input format was `<bos> prefix <mask> suffix <sep> infill <eos>` over raw amino-acid tokens, the objective was causal shifted-token cross-entropy over the missing span. It trained for 1,050,000 optimizer steps over 17 epochs in two stages that share an effective batch of $1{,}998{,}848$ tokens per optimizer step: the first 1,000,000 steps at context length 512 ($32 \times 62{,}464$ tokens per device, no accumulation), then the next 50,000 steps at context length 2048 ($32 \times 15{,}616$ tokens per device, 4 accumulation steps). Sequences were packed with FlashAttention (Dao et al., 2022) variable-length attention. Training comprised 1.77×10^{11} loss-bearing in-filling tokens (2.10×10^{12} tokens including padding and unmasked context). Training used the AdamW optimizer (Loshchilov & Hutter, 2019) (fused) at peak learning rate 10^{-3}, $\beta = (0.9, 0.95)$, $\epsilon = 10^{-8}$, weight decay 0.01, and gradient-norm clip 1.0, with cosine decay to a floor of 10% of peak (10^{-4}) over a 900,000-step horizon after 1,000 warmup steps. The final 150,000 steps were run at the clamped floor. Training used bf16 mixed precision with TF32 matmuls. The run used 32 × NVIDIA H100 80GB (4 nodes × 8) for 130.6 hours wall-clock (≈ 4,180 GPU-hours). Three held-out sets are evaluated every 1,000 steps: `uniprot_dev` (10,000 sequences), `oas_dev` (20,000 chains from 10,000 paired antibodies), and `scop_dev` (10,000 domains), with a final in-filling loss of 1.430 (`uniprot_dev`) and 0.238 (`oas_dev`).

For CDRH3 generation, the native CDRH3 is treated as the missing FIM span, such that the model autoregressively generates the loop following the `<sep>` token while conditioning on the framework prefix and suffix. This formulation also permits the conditional log-probability of a complete CDRH3 sequence to be calculated in a single teacher-forced forward pass, rather than the L forward passes required for leave-one-out pseudo-likelihood evaluation with a masked language model for a loop of length L. On 55 held-out CDRH3 loops, the autoregressive model evaluates 97 loops/s compared with 9.1 loops/s for masked-model leave-one-out pseudo-likelihood, a 10.7× throughput improvement. The observed speedup approximately tracks the median loop length of 10 residues, reflecting the reduction from L model evaluations to one while accounting for fixed computational overhead.

A.2 MOGFN Fine-Tuning

The policy was initialized with all but the final 4 of its 32 transformer layers frozen, and a freshly initialized output head (MLP $d_h \to 2d_h \to 2d_h \to 20$) over the 20 amino-acid action space. Each fine-tuned transformer layer was wrapped in two sequential cross-attentions (scaffold, then preference), with 8 heads, dropout 0.1, independent projections, and zero-initialized output projections. Preferences were encoded per slot by a thermometer encoding with $B = 50$ bins, joined with a learned 64-dimensional slot-identity embedding and passed through a shared MLP $B + 64 \to 2d_h \to d_h$ (SiLU, dropout 0.1) and RMSNorm, which gave D key-value tokens. The task embedding was the per-residue hidden states of the native heavy-chain scaffold with its natural CDRH3 sequence, taken after the frozen layers of the span-in-filling backbone and cached per target; it gave L_{task} key-value

[5] Log-probabilities for stability and likelihood oracles, probability-derived scores for immunogenicity risk and developability, and counts for liabilities. [6] Margin $\epsilon_{\mathrm{m}} = 0.05$ to prevent score pileup at the boundaries.

tokens with a key-padding mask, unpooled. Trajectories were per-amino acid autoregressive CDRH3 generation after <sep> employing KV-cache sampling. The optimizer was AdamW at $\mathrm{lr} = 10^{-4}$ with 200 warmup steps, cosine decay to $0.1\times$, and gradient-norm clipping, over $K = 4$ target structures per gradient step. The log-cosh scale is per-structure and annealed, $\beta_{\mathrm{loss}}(t) = c(t)/(\hat{\sigma}_k + \epsilon)$ with $c(t) = c_{\max} e^{-t/\tau_\beta} + c_{\min}(1 - e^{-t/\tau_\beta})$, where $\hat{\sigma}_k$ is a per-structure EMA of the residual standard deviation for structure k (see appendix A.3). Each scaffold update drew one distinct target structure and one preference vector $\mathbf{w} \sim \mathrm{Dir}(\alpha)$, and sampled $N = 256$ completions under that fixed scaffold and preference[7]. The distinct K structures are drawn uniformly before drawing from the replay buffer, and afterwards with a staleness priority. All $KB = 1024$ candidates were scored in one batched oracle call and each sub-task contributes proportionally to the gradient. Training trajectories were sampled using temperature scaling on the policy per-step logits, $q(\tau) = \prod_t \mathrm{softmax}(z_t/T)_{a_t}$, with T annealed linearly from 2 to 1.5 over training. The trajectory-balance loss used the untempered $\log P_F$ throughout, and all reported samples are drawn from the untempered policy (see appendix A.4 for details). The replay buffer stores raw M-dimensional reward vectors, which are re-projected using $\mathbf{A}$ per step.

A.3 Log-Cosh Residual Penalty

The log-cosh training schedule used $c_{\max} = 10$, $c_{\min} = 2$, $\sigma_0 = 10$, EMA decay $\gamma = 0.999$ and ϵ of 10^{-6}, and $\tau_\beta = 875$. Training proceeded for either 1750 train steps (9 objective GFN) or 3000 steps for the 14-objective arms. The schedule results in the loss quadratic-to-linear knee moving from $|\delta| = 1.0$ out to 3.2 and 4.4, respectively, over training. Replay had capacity 20,000, staleness 3.0, and warmup 262 steps. Reward normalization was online min–max with warmup, EMA adaptation ($\gamma_{\mathrm{ema}} = 0.998$), a freeze at step 350, and margin $\epsilon_{\mathrm{m}} = 0.05$. The projection used either $D = 4$ or $D = 5$, $\lambda_{\mathrm{dec}} = 0$, EMA $\tau = 0.99$, a freeze at $0.60\times$ train steps, a warm start of 2000 steps, and reference fraction 0.5. Evaluation during training happens every 250 steps on 32 Dirichlet-sampled preferences over 5 validation structures.

Hu et al. (2025) show that the regression loss g applied to trajectory-balance residuals determines an f-divergence between the target and the learned distribution;Fawkes & Hartford (2026) extend the correspondence to translation-invariant losses and off-policy data. They classify losses by two properties of that divergence: zero-forcing ($f(0) = \infty$, which penalises putting mass outside the target distribution, i.e., favouring exploitation) and zero-avoiding ($f'(\infty) = \infty$, which penalises missing mass, favouring exploration). The squared loss $g(t) = \frac{1}{2}t^2$ is zero-forcing and not zero-avoiding. The three losses they propose all grow faster than quadratic as $t \to +\infty$, but only Shifted-Cosh does so in both tails: Linex(1) and Linex(1/2) are linear as $t \to -\infty$, which is why neither is zero-forcing. Log-cosh sits at the opposite end from the squared loss, so it is neither zero-forcing nor zero-avoiding, and does not contain the squared loss exploitation pressure. Our annealing of $c(t)$ moves the loss back towards quadratic over the course of training, growing progressively biased toward exploitation, in an attempt to reduce the magnitude of tail residuals without discarding them; Chen et al. (2026) anneal the policy mixing weight instead. The gradient is $\tanh(\beta_{\mathrm{loss}}\delta)$ — 0.81, 0.55 and 0.38 times the squared-error gradient at $1\sigma_\delta$, $2\sigma_\delta$ and $3\sigma_\delta$ — so the penalty acts as a smooth per-sample gradient clip that relaxes as $c(t)$ decreases.

A.4 Policy Conditioning and Sampling

For AMPLIFY-AR span-in-filling pretraining, we sample a single interior span by first drawing length uniformly from 1 to 50 residues, clamped by the sequence length, and then drawing a valid start position uniformly at random. The tokenized sequence is rearranged as

<bos> prefix <mask> suffix <sep> infill <eos>.

The preference vector $\mathbf{w} \in \Delta^{D-1}$ is encoded via thermometer encoding into $\mathbb{R}^{BD}$ ($B = 50$ bins per slot). Given B linearly spaced thresholds $\xi_1, \ldots, \xi_B$ over $[0, 1]$ with gap $\Delta\xi = \xi_{j+1} - \xi_j$, the thermometer encoding of a scalar c is:

$$\mathrm{therm}(c)_j = \frac{\mathrm{clamp}(c - \xi_j,\, 0,\, \Delta\xi)}{\Delta\xi}, \quad j = 1, \ldots, B. \tag{2}$$

[7] Holding both fixed across a sub-batch allows the FIM prefix be shared through a KV cache (appendix A.5) and gives a valid $\log Z(\mathbf{w})$ estimate for that one condition.

Each slot is tokenized separately. We joint the thermometer slice $\mathrm{therm}(w_d) \in \mathbb{R}^B$ with a learned slot-identity embedding $\mathbf{a}_d \in \mathbb{R}^{64}$ and passed through a shared two-layer MLP ($B + 64 \to 2d_h \to d_h$, with SiLU and dropout 0.1) and a final RMSNorm, where $d_h = 960$ is the AMPLIFY hidden dimension. This gives D preference key-value tokens $\mathbf{E}_{\mathrm{pref}} \in \mathbb{R}^{D \times d_h}$. The MLP is shared across slots, followed by a RMSNorm.

Scaffold (or task) conditioning is handled by the key-value set $\mathbf{E}_{\mathrm{task}} \in \mathbb{R}^{L_{\mathrm{task}} \times d_h}$, i.e., the full sequence of per-residue hidden states of the native heavy chain $\mathbf{x}_{\mathrm{frame}}$ (including its natural CDRH3) after the frozen layers of the span-in-filling backbone, with a key-padding mask over padded positions. It is computed once per target and cached, with no pooling applied. The cross attention conditioning mechanism rests between the self-attention and the feed-forward network of the transformer backbone, first applying scaffold cross-attention before preference cross-attention. This ordering enabled the preferences to modulate generation according to preferences from hidden states informed by the task-scaffold, i.e., the same objective trade-off can be differently applied for different targets. Each cross-attention output projection was initialized to zero. Given input token representations $\mathbf{H}^{(\ell)} \in \mathbb{R}^{L_{\mathrm{seq}} \times d_h}$ at layer ℓ, and writing RMS for RMS normalization:

$$\mathbf{H}' = \mathbf{H}^{(\ell)} + \mathrm{SelfAttn}\Big(\mathrm{RMS}_1(\mathbf{H}^{(\ell)})\Big) \tag{3}$$

$$\mathbf{H}'' = \mathbf{H}' + \mathrm{Dropout}(\mathrm{CrossAttn}_{\mathrm{task}}(\mathrm{RMS}_2(\mathbf{H}'),\, \mathbf{E}_{\mathrm{task}},\, \mathbf{E}_{\mathrm{task}})) \tag{4}$$

$$\mathbf{H}''' = \mathbf{H}'' + \mathrm{Dropout}(\mathrm{CrossAttn}_{\mathrm{pref}}(\mathrm{RMS}_3(\mathbf{H}''),\, \mathbf{E}_{\mathrm{pref}},\, \mathbf{E}_{\mathrm{pref}})) \tag{5}$$

$$\mathbf{H}^{(\ell+1)} = \mathbf{H}''' + \mathrm{FFN}(\mathrm{RMS}_4(\mathbf{H}''')) \,. \tag{6}$$

Each $\mathrm{CrossAttn}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ is multi-head cross-attention with 8 heads and dropout 0.1. The two cross-attention mechanisms have independent query, key, value, and output projections, with no weight sharing, and both output projections are initialized to zero.

Replay buffer. The buffer is partitioned by target structure. Each entry holds one trajectory's actions, its per-step log-probabilities as written, and the raw M-dimensional reward vector. When a partition is consumed, its entries are re-scored under the current policy and filtered on the log importance ratio $\rho = \log P_F^{\mathrm{now}}(\tau) - \log P_F^{\mathrm{stored}}(\tau)$. Entries with $|\rho| > c_\rho$ are discarded, with $c_\rho = 3$. The K structures per-step are drawn from the buffer without replacement with probability proportional to $t - t_s + 1$, where t_s is the last step for structure s.

Reward normalization. Three phases: 1) During warmup, a running min and max for each reward expands monotonically using an expanding window: $r_k^{\min} \leftarrow \min(r_k^{\min}, \min_{\mathrm{batch}} R_k)$ and $r_k^{\max} \leftarrow \max(r_k^{\max}, \max_{\mathrm{batch}} R_k)$. 2) During EMA adaptation, these bounds update with decay γ_{ema}: $r_k^{\min} \leftarrow \gamma_{\mathrm{ema}}\, r_k^{\min} + (1 - \gamma_{\mathrm{ema}}) \min_{\mathrm{batch}} R_k$ (likewise for the max). 3) During the frozen phase, these bounds lock so the multi-objective GFN can learn from a stationary reward distribution.

A.5 KV-Cache Sampling

To avoid sampling t times during the FIM pass $[B, P + t]$ where $P \approx 120$ is the fill-in-the-middle prompt and $L \approx 5$–25, we use a standard KV-cache prefill-plus-incremental-decode under `no_grad`. Prefill runs the full prompt through the frozen and conditioned layers capturing each layer's K and V, and each decode step embeds one new token, applies rotary embeddings at the token's absolute position, appends to the cache, and attends over it. Conditioned layers insert the same two cross-attentions against $\mathbf{E}_{\mathrm{task}}$ and $\mathbf{E}_{\mathrm{pref}}$, which are fixed for the whole step and never cached: $\mathbf{E}_{\mathrm{task}}$ is computed once per target and $\mathbf{E}_{\mathrm{pref}}$ once per preference vector. Within a sub-batch every sequence shares the same P and L, and the K sub-batches are sampled separately.

B Data Pipeline Details

B.1 SAbDab Curation

The SAbDab (Dunbar et al., 2014) dataset comprises paired heavy/light chain sequences with antigen co-crystal structures, associated metadata (resolution, species, affinity where available), and PDB

coordinate files. We removed entries missing heavy-chain, light-chain, antigen-chain, or antigen-type annotations, excluded multi-chain antigens, and retained only protein and peptide antigens. Entries were deduplicated by PDB code and heavy-chain identifier. Heavy-chain, light-chain, and antigen sequences were parsed from PDB ATOM records using Biopython (Cock et al., 2009), with non-standard residues mapped to canonical one-letter amino acid codes where possible. CDRH3 boundaries were identified using ANARCI (Dunbar & Deane, 2016) under the IMGT numbering scheme (Lefranc et al., 2003) (positions 105–117), and gap characters at insertion positions were removed. Structures with CDRH3 length < 5 were excluded because their sequence space is sufficiently small to enumerate exhaustively.

Structures resolved better than 2.3 Å were retained, leaving 873. The retained set spans 1.2–2.3 Å (median 2.0) over 469 protein and 315 peptide antigens. We additionally remove 72 samples due to parser disagreements between the structures and inverse-folding models: 17 entries are excluded by a maintained list, and the rest fail the per-structure consistency check of the loop logits.

Train, validation, and test sets were defined based on the heavy-chain scaffold to evaluate whether the loop-design policy generalizes to novel antibody frameworks rather than memorizing CDRH3 sequences associated with nearly identical scaffolds. Because antibody frameworks are germline-conserved (median scaffold sequence identity ≈ 0.85 across unrelated antibodies in our dataset), we used a near-identity threshold of ≥ 0.95 to identify potential leakage. This removes exact and closely related affinity-matured scaffold duplicates across splits while retaining members of the same germline family as distinct design problems. Antigen information was not used to define the primary split, but was used to verify that held-out targets represented distinct antigen-binding problems. After curation, filtering, and leakage control, the final dataset comprised 623 training, 10 validation, and 55 held-out test targets from an initial universe of 784 structures, with 161 structures excluded during processing.

Step	Rule	Removed	Retained
SAbDab entries after annotation cleaning	complete H/L/antigen annotation; single-chain protein or peptide antigen; one entry per PDB	—	1,905
Resolution	$0 <$ resolution ≤ 2.3 Å	1,032	873
Parser blacklist	17 PDB entries with known structure/inverse-folding parser mismatch	17	856
Per-structure checks inside the inverse-folding loop			
Chain residue-count check	parsed residue count must equal the heavy-chain sequence length	67	
BoltzIF resolved-residue check	BoltzIF resolved residues must equal the sequence length	4	
ProteinMPNN parse	structure must parse under every ProteinMPNN variant	1	784
Cross-model CDRH3 length agreement	all 24 models must agree on CDRH3 length, under Chothia and under IMGT	0	784
Recorded per structure; never a cull			
Sequence recovery	argmax agreement with the native CDRH3	0	784
Boundary-shift test	native CDRH3 must score above every ± 1.. ± 5 shift and above chance	0	784

Table 3: SAbDab curation, from the cleaned metadata table to the structures carrying precomputed inverse-folding logits. The first three steps act on the metadata; the next four are per-structure checks inside the inverse-folding loop, which drop a structure when the parsers disagree with each other or with the deposited sequence. Sequence recovery and the boundary-shift test are recorded per (structure, model) and never remove a structure, so a structure may be admitted carrying a failing flag. The per-structure drops are recovered from the run log, since the stored artifact records only the admitted set.

B.2 INVERSE-FOLDING LOGIT ALIGNMENT

We computed inverse folding logit matrices across the full curated SAbDab dataset for the stability oracles as follows:

1. Each PDB structure is processed independently by ESM-IF, four ProteinMPNN variant families, and BoltzIF.
2. For each model family m, stored logits are indexed as $\mathbf{g}_{v,e}^{(m)}(s) \in \mathbb{R}^{|\mathcal{A}| \times L_{\text{chain}}}$, where $v \in \mathcal{V}_m$ indexes weight variants, $e \in \{1, \ldots, E\}$ indexes ensemble members, $|\mathcal{A}|$ is the vocabulary size, and L_{chain} is the chain length.
3. Because models parse PDB files differently, each model records per-position PDB residue numbers. Cross-model alignment is performed at inference time via these residue identifiers.
4. Full-chain logits are sliced to the CDRH3 region using both Chothia (Al-Lazikani et al., 1997; Chothia & Lesk, 1987) and IMGT boundary definitions.
5. The pipeline performs sequence-recovery, boundary-shift, and cross-model CDRH3 length-agreement checks (Table 3). Separately, a maintained list excludes 17 PDB entries whose parsers are known to disagree.

Sequence recovery (fraction of CDRH3 positions where the argmax prediction matches the native residue) was performed as a correctness evaluation for each structure, variant, and ensemble member:

$$\text{Recovery}_{v,e}^{(m)}(s) = \frac{1}{L} \sum_{i=1}^{L} \mathbb{1}\left[\arg\max_{a} \mathbf{g}_{v,e}^{(m)}(s)_{a,i} = x_i^{\text{native}}\right]. \tag{7}$$

A boundary-shift test was also performed to ensure correct loop placement: the mean log-probability of the native CDRH3 sequence is computed at the stated logit columns and at $\pm 1, \ldots, \pm 5$ shifted positions. The stated position must yield a higher score than all shifted positions and must exceed uniform chance ($\log(1/20) \approx -3.0$). For cross-model agreement, structures where the 24 models disagree on CDRH3 length under either the Chothia or the IMGT scheme are excluded, though no structure in the current set triggers this. The shift test is recorded rather than enforced: all 24 models pass it on 558 of 784 structures under Chothia and on 695 under IMGT, and the per-model flags are stored so that a consumer can require agreement where it matters.

B.3 TRAIN, VALID, AND TEST SPLITS

The generator conditions only on the heavy-chain scaffold; the light chain and antigen never enter generation, and the training reward is antibody-intrinsic (binding is computed only in the post-filter). The CDRH3 loop is the model's output. We therefore generate splits using a clustering of the framework region.

Exclude scaffold identity $\geq T$	training kept	excluded
0.50	1	707
0.70	43	665
0.80	136	572
0.90	529	179
0.95	**660**	**48**
0.97	669	39
1.00	691	17

Table 4: Leakage menu on the heavy scaffold (heavy chain minus CDRH3) over the 784-antibody pool: normalized identity of each training scaffold to its nearest held-out test scaffold. The $T = 1.00$ row excludes exact duplicates only.

Since antibody VH frameworks all descend from a small set of germline genes, their scaffolds are similar even between unrelated antibodies: measured percentiles of nearest-test-scaffold identity are p50 $= 0.856$, p75 $= 0.900$, p90 $= 0.934$, p95 $= 0.981$, p99 $= 1.000$.

Since the population sits at ≈0.85 median scaffold identity, "zero leakage" is not achievable in an absolute sense, leakage on the scaffold axis is necessarily a near-identity notion we opt for 0.95 clustering to retain more training examples.

The minimum CDRH3 length of 5 removes 31 antibodies: Loops of length ≤ 4 span at most $20^4 = 160$k sequences which can be searched exhaustively, $L = 5$ jumps to 3.2M and serves as the lower length bound.

	count
Universe (structures with precomputed inverse-folding logits)	784
After minimum CDRH3 length ≥ 5	753
TRAIN	**623**
VALID (CDRH3-length-stratified)	**10**
TEST (RAbD-derived, reserved)	**55**
`exclude_from_training`	161

Table 5: Final split. The exclusion set (161) is much larger than TEST + VALID because leakage is defined on the scaffold: training antibodies whose scaffold is ≥0.95 identical to a held-out target are removed as well (48 test-leaks, 27 valid-leaks), on top of the 31 dropped by the length filter. Validation CDRH3 lengths are {5, 7, 9, 9, 10, 11, 12, 14, 16, 18}.

The scaffold split above only controls what the GFN is fine-tuned on. The backbone was pretrained on UniRef100, OAS and SCOP (appendix A.1). No held-out antibody is in the the SCOP training set (24,495 sequences), the SCOP development set (10,000), or the OAS training set (3,611,708). Four test and one validation sequence to have a matched PDB entries on the antigen side: HER2 (`1N8Z_C`), the cytochrome bc_1 complex (`3CX5`), and the TSH receptor (`2XWT_C`). The generator never conditions on these. Additionally, only five test and one valid CDRH3 loops are present in the OAS, which is a chance occurance rate. Held-out loops are not enriched relative to the rest of the 784-structure pool.

Table 6: Verbatim CDRH3 recurrence in the OAS pretraining fill with held-out targets against the rest of the structure pool, by loop length. A short loop appearing somewhere in 3.6M repertoire sequences means nothing on its own, so the interpretable quantity is the rate relative to non-held-out loops at the same length.

	held-out (TEST + VALID)		rest of pool	
CDRH3 length	entries	recur	entries	recur
≤5	2	2 (100%)	76	76 (100%)
6–7	3	1 (33%)	53	29 (55%)
8–9	12	2 (17%)	130	21 (16%)
10–12	23	1 (4%)	204	2 (1%)
total (≤12)	**40**	**6 (15.0%)**	**463**	**128 (27.6%)**

The test set is the RAbD benchmark (Adolf-Bryfogle et al., 2018) targets for which scaffolds are available and which clear the length filter (55 of 60), deduplicated on scaffold ≥0.90 and antigen ≥0.90 so every test entry is a distinct binding problem; at these thresholds 0 targets collapse. Generation covered 45 of the these 55 test targets as they are members of the original 784-structure universe.

B.4 Reward Metrics Through the Funnel

In Figure 6 we show the the rewards from all arms, shown as a percentile against the full pool. The learned GFNs sample designs with substantially different properties than the AMPLIFY-AR prior. Two broad trends can be seen in this analyses: 1) the structural models and AMPLIFY-AR prior outperform the learned GFNs and random sampler at the outset of the pool – this gap is closed via the application of the postfilter pipeline, not completely. On these metrics, the learned GFNs outperform a random sampler. 2) The learned GFNs outperform all baselines on novelty, developability, MHC-II binding, and perform at parity with respect to liability. The delta in performance between the

AMPLIFY-AR prior and the various trained objectives suggests that our training recipe was not optimal across all objectives, and a more precise training procedure which prioritizes particular reward dimensions adaptively during training might produce better tradeoffs in the final sampler.

B.5 Diversity Metrics

Pool diversity is measured over ESM-2 650M sequence embeddings, with the Vendi score (Friedman & Dieng, 2023) (the effective number of distinct modes; the exponential of the Shannon entropy of the eigenvalues of $\mathbf{K}/n$, where $\mathbf{K}_{ij}$ is the embedding similarity between designs i and j and n is the pool size, which is scale-free and mode-count-aware) and mean pairwise L_2 distance. We do not report 2D UMAP convex-hull area as a diversity metric due to volumes in this space being a projection artifact. UMAP is used for visualization only. Vendi counts effective modes and therefore grows with sample size, so every arm is subsampled to the same n before scoring. Embeddings are calculated with the loop in context of it's heavy chain. The GFN methods are close to the untrained control on both diversity measures while adding objective controllability:

A Vendi score at one sample size cannot separate many modes from a dense sample of few modes, because it is a single number at a single n. We therefore also resample each arm's pool at increasing draw sizes and report how the effective count grows.

The same populations can also be shown in the space itself rather than as a count, which says *where* in the landscape an arm still has designs once the cascade has run.

C Conditioning, Controllability, and Learned Preference Slots

Since the learned projection centres the target ($\hat{r}_{bj} - \mu_j$) and weights the reconstruction loss by $1/\sigma_j^2$, a reward with a large dynamic range cannot win a slot on scale alone. Reconstruction quality is measured as $R_j^2 = 1 - \mathrm{MSE}_j/\mathrm{Var}(\hat{r}_j)$. $\mathbf{A}$ is warm-started on 2,000 steps of offline rewards over native CDRH3 loops. $\mathbf{A}$ mixes min-max normalized rewards. Following Park & Sung (2025) the learned reduction is affine, positive and row-stochastic, $f(r) = \mathbf{A}r + b$ with $b = 0$, which keeps every slot a convex combination and therefore $\mathbf{z} \in [0, 1]^D$. An EMA target $\bar{\phi} \leftarrow \tau\bar{\phi} + (1 - \tau)\phi$ supplies the rewards during training, decoupling how fast the projection learns from how much the reward can vary across scaffolds. GFN D=4 and GFN D=5, which learn their slots over all 14 objectives, and GFN FIXED, whose four axes are fixed config groups over 9.

An analyses of the learned preference slots (Figure 10a,b) reveals that both learned groupings recover reasonable and repeatable structure: both isolate MHC-II, each gives the three sequence-likelihood oracles a single slot in near-equal thirds, and each keeps the six inverse-folding stability oracles (roughly) together. At D=5 the two novelty terms concentrate in one slot, carrying 0.387 of the total row mass; at D=4 they carry 0.022 of 4. The D=4 reduction effectively discards novelty rather than compressing it. The conditioned arms return $\rho = +0.63$–$+0.75$ Spearman's correlations between the preferences and the rewards, while the unconditioned models are centred on zero (-0.04 and +0.01); Figure 10c). Interestingly, ρ falls by roughly half among B_4 survivors, suggesting that the structural post-filter pipeline removes designs which are too extremely committed to any preference.

Selectivity (Figure 10d) measures specific modulation of rewards over the full sampled pool. It measures whether an axis moves the objectives it with higher weight under $\mathbf{A}$ than those with lower weight, and reports how reliably an objective tracks the axis weight, rather than by how much. Zero means the axis moves everything similarly. Steering (Figure 10e) is a magnitude in sequence space, and measures how much of the design distribution moves with the axis weight in general. It is a ratio against a shuffled null on that axes, so 1 means indistinguishable from no conditioning.

Selectivity is positive on every axis of all three arms, and the higher weighted axes are the top-responding objectives. The MHC-II axis is the highest-selectivity axis in all three (0.827, 0.729 and 1.476 for slot2, slot1 and mhc2). The six-way stability mixture is in the bottom half of each, which is unsuprising as each inverse folding model was trained on different data distributions. Within each arm the two measurements order the axes similarly (Spearman +1.00, +0.60 and +0.60).

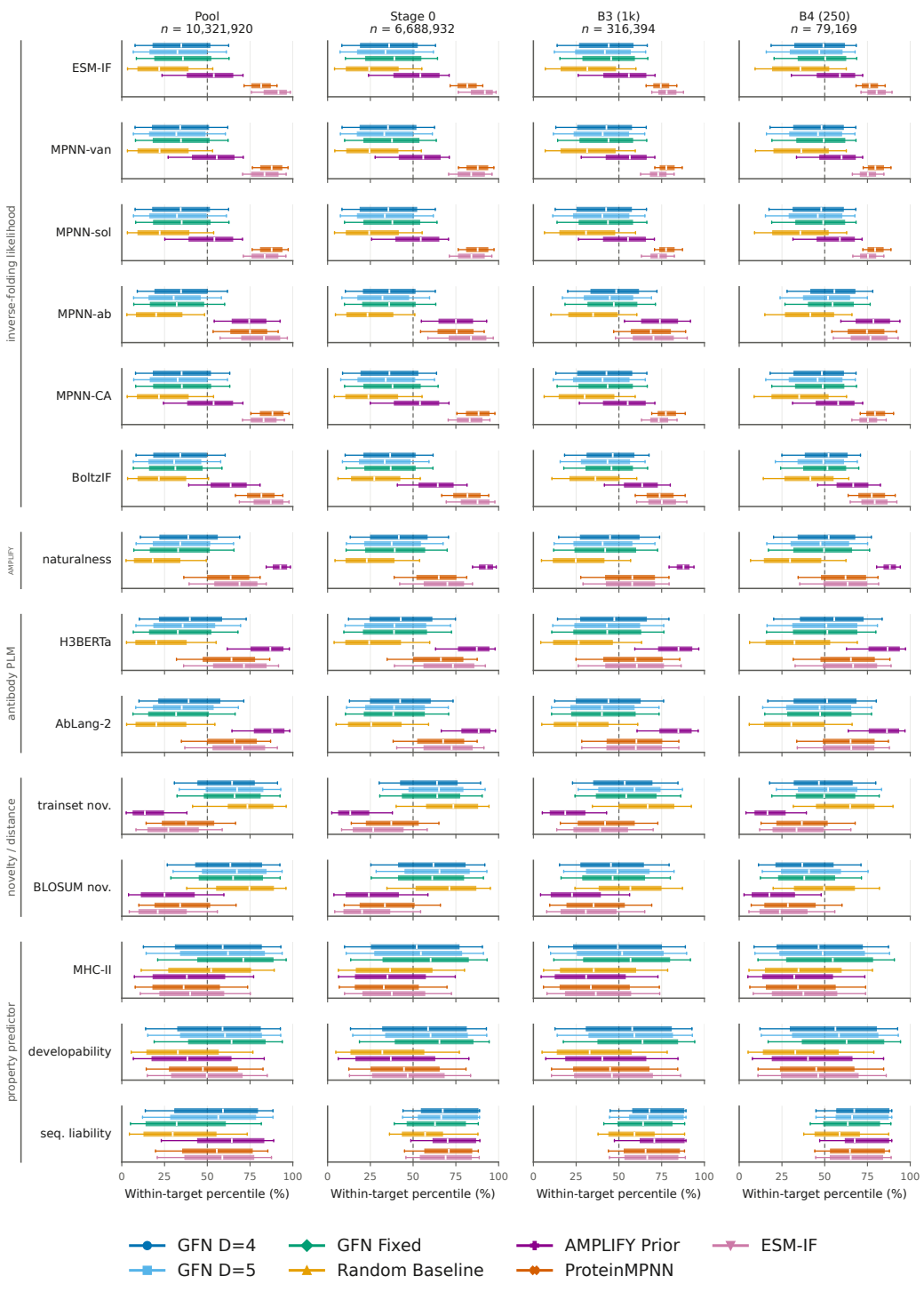


Figure 6: Distribution of per-objective scores for each sampler at each level of the postfilter funnel, relative to the pool-level statistics. Within-target percentile is calculated against the pooled pool of all arms. 50% is calculated as the pooled-pool medium, which explains why values sum to $> 100\%$ at the end of the funnel. Boxplots denote the inter-quartile range, with a tick at the medium and whiskers denoting 10% and 90%.

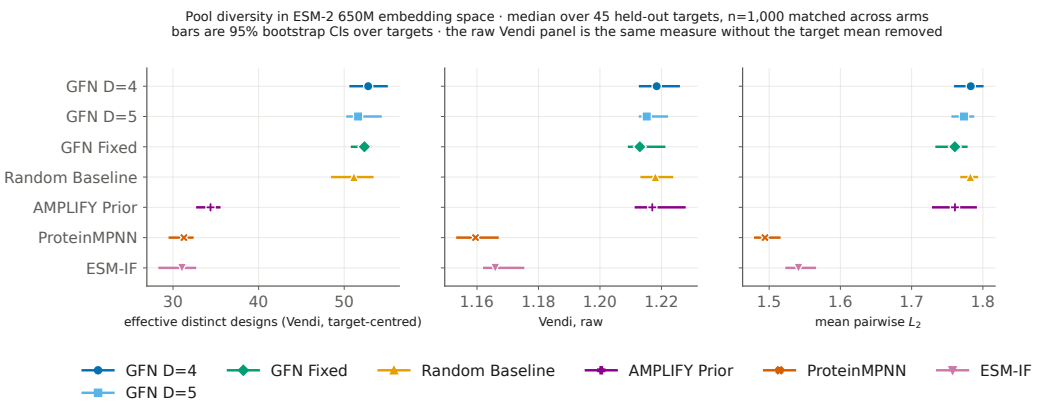


Figure 7: Sequence diversity over ESM-2 650M embeddings of the CDRH3 loop in its own heavy-chain context, on the post-filter design pools. Median over held-out targets with 95% bootstrap CIs resampling targets.

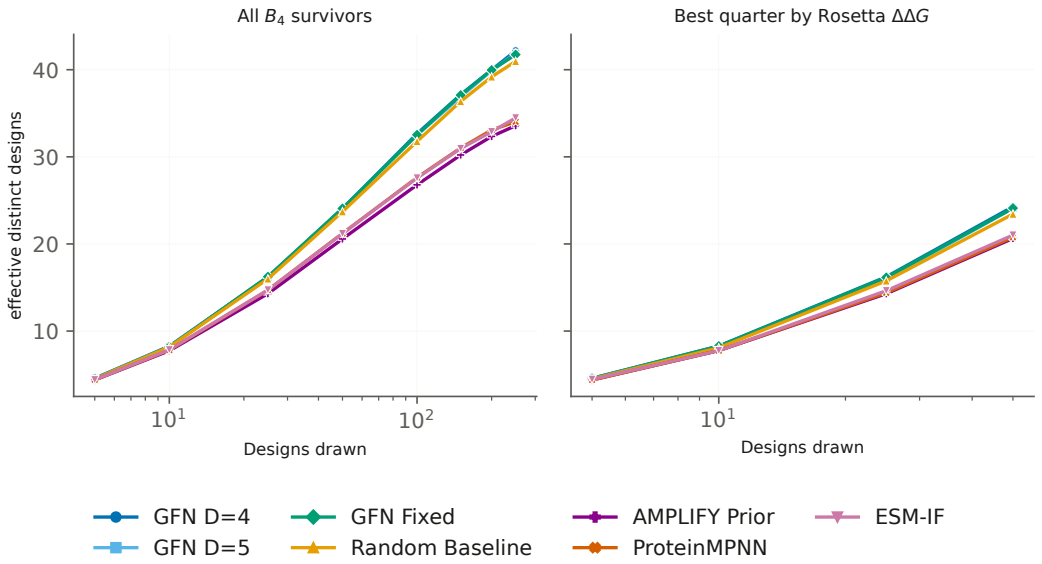


Figure 8: Rarefaction of the B_4 survivor pools, which measures the effective number of distinct designs against how many designs are drawn per pool / target, averaged over 45 held-out targets. Draws come from the same pool and are repeated at each size. The Vendi score is taken over ESM-2 650M embeddings of the loop in its own heavy-chain context, after removing each target's mean embedding, so an arm is measured against the other arms at that target rather than against the target. *Left:* all B_4 survivors. *Right:* the same measurement over the best quarter of those survivors by Rosetta $\Delta\Delta G$. Both curves clearly show that the GFN arms draw more unique designs per draw than the structural baselines.

D Post-Filter Details

This analysis' purpose is to refine the samples generated by each arm according to increasingly strict and expensive held-out oracles, in lieu of wet lab tests of sample quality. A liability gate first removes every design that fails sequence-level requirements. Afterwards, all removals are done according to a *budget* informed by our computational constraints. These designs are ranked and the top performers are carried forward to the final analysis. Stage 0 gates on clear sequence liabilities. Stage 1 folds every Stage-0 survivor with ABodyBuilder3, clusters the survivors' CDRH3 at 50% identity, and

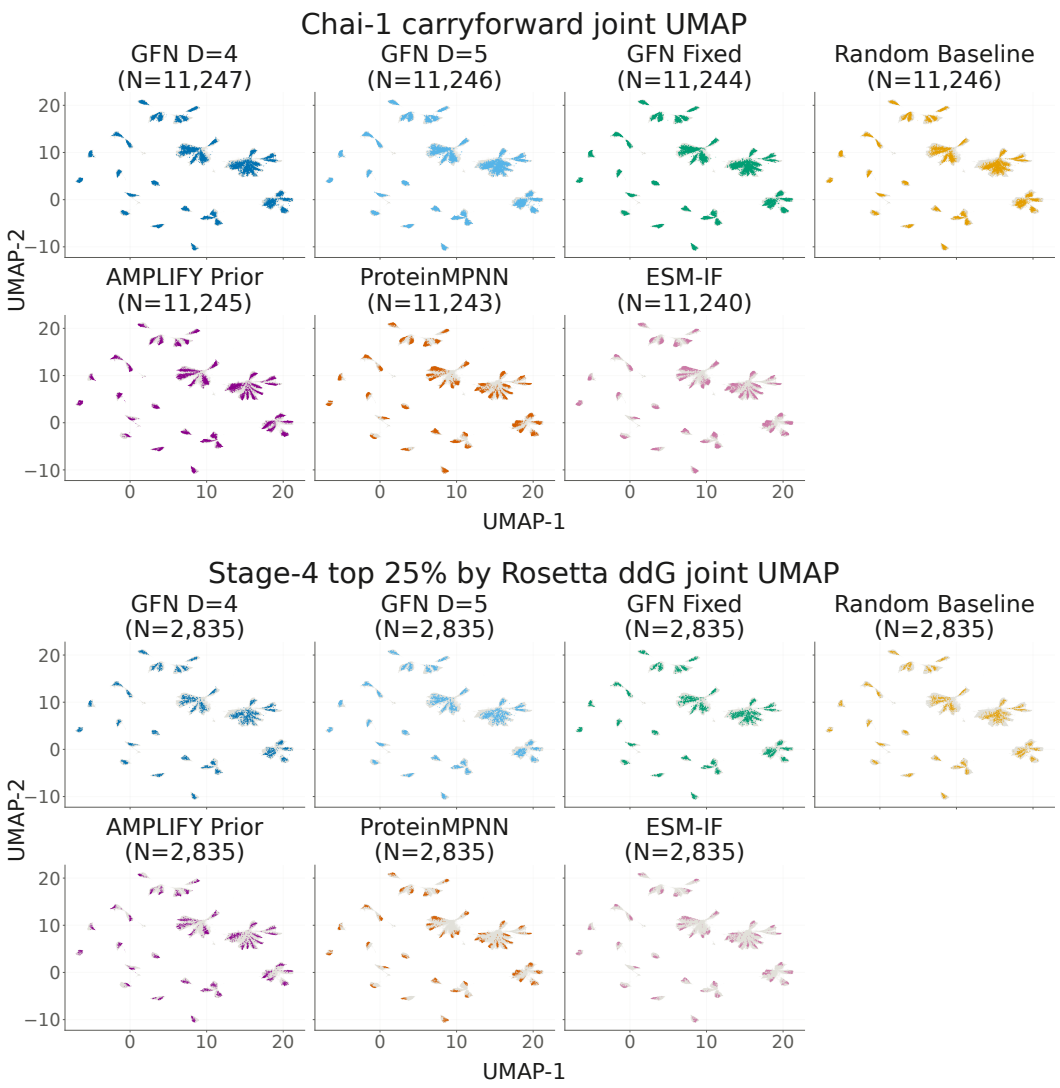


Figure 9: Where each arm's surviving designs sit in the shared sequence landscape. The UMAP is fit once over the background distribution of every embedded design of every arm and target, and each panel then plots that background distribution in grey. The coloured points are the designs of that arm which passed the post-filter thresholds. *Top:* the designs carried forward to the Chai-1 stage, B_4. *Bottom:* the best quarter of those by Rosetta $\Delta\Delta G$. Embeddings are ESM-2 650M over the CDRH3 loop in its own heavy-chain context under cosine distance. The islands largely represent individual test-set targets. The plots demonstrate qualitatively how the different methods coexist in design space – in particular, how the GFN arms reach across the boundaries of both the structural models, the prior, and the random baselines.

carries B_3 designs forward: a cluster of n members earns $\lfloor \log_{10} n \rfloor + 1$ slots and spends them on its members of highest mean CDRH3 pLDDT. Stage 2 computes spatial aggregation propensity on those same Stage-1 structures and records it for analysis only. Stage 3 forward folds the carried designs a second time under Chai-1, keeps the best B_4 designs of the target's own ABodyBuilder3–Chai agreement distribution. These are assembled onto the crystal antigen in Stage 3c and scored for interface energy by FoldX and Rosetta in Stage 4. The scores of each arm entering and exiting the post-filter funnel can be found in Figure 11.

Stage 2 SAP multiplies solvent exposure by side-chain hydrophobicity and sums the product over a 5 Å radius. It mostly measures composition and exposure. A design can score low either due to fewer

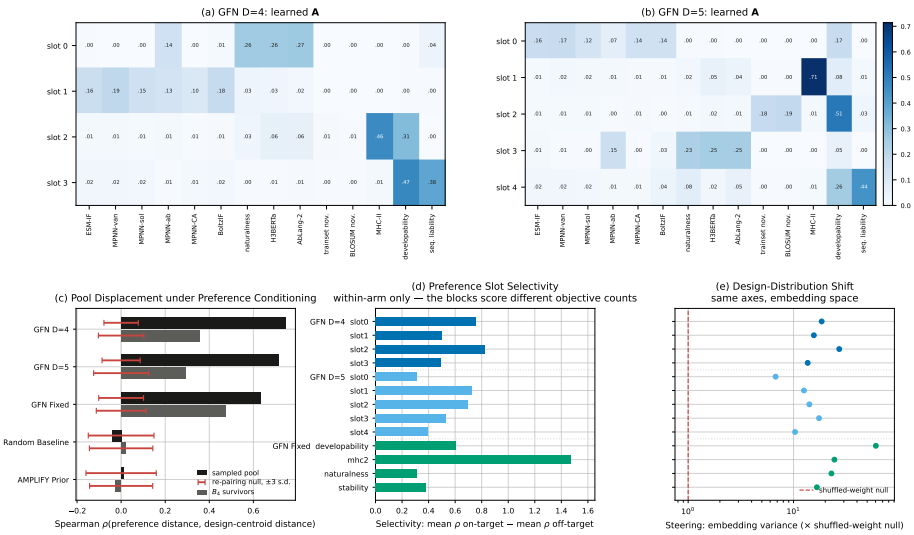


Figure 10: (a,b) Learned $\mathbf{A}$ weights over the 14 rewards for the $D = 4$ and $D = 5$ GFNs. Each row d sums to 1. While many groupings are broadly similar between the two slot counts, some parings change dramatically, such as the isolation of MHC-II in slot 1 for $D = 5$. **(c)** Spearman ρ of the Euclidean distances between every pair of an arm's preference vectors, and the distance between the matching design centroids (the mean target-centred ESM-2 embedding of a preference's designs) pooled over targets. Red bars span ± 3 s.d. Only conditioned models show design dispersion outside of this null region. **(d)** R_{dj} is the Spearman correlation, across preferences, between the weight on axis d and the per-preference mean of objective j, averaged over 45 targets (i.e., it is performed in reward space). Let w_d be the group indicator of an axis's row of $\mathbf{A}$. Selectivity is the difference between on target $\sum_j w_{dj} R_{dj} / \sum_j w_{dj}$ minus the off-target $\max_j w_{dj} - w_{dj}$. **(e)** The same axes as (d) in embedding space space which shows the degree to which each axis displaces the embedding.

Stage	Signal	Role	Cost
0	Sequence liability motif scan	Removes undesirable designs	μs, CPU
1	ABodyBuilder3: H3 pLDDT, scRMSD	Cheap structural quality filter	0.25 s/design
2	Aggregation propensity	Developability score	0.15 s/design
3	Chai-1: RMSD, lDDT, pLDDTs	Expensive quality filter	$\leq$31 s/design
3c	Assembles predicted Fv w/ antigen crystal	Assemble structures	0.27 s/design
4a	Rosetta:interface energy	Evaluate	2.66–44 min/complex
4b	FoldX: interface energy	Evaluate	1.05–2.3 min/complex

Table 7: This evaluation cascade is designed for reference-free mode recall. Stage 0 gates on clear sequence liabilities. Stage 1 folds all surviving sequences, clusters at 50% similarity, and carries forward the top designs. Stage 2 calculates aggregation propensity on the Stage-1 structures. Stage 3 forward folds top designs under Chai-1, which are analyzed both for fold confidence and structural similarity to ABB3. The best sequences are then brought forward for binding estimation with the crystal scaffold under FoldX and Rosetta.

hydrophobic residues, or buried residues. The `developability` objective is related to this due to training labels derived from the Developability Index (Lauer et al., 2012), which is built on CDR aggregation propensity (Section 3.2). The six inverse-folding likelihoods also influence this outcome variable because they score burial on the native backbone.

Stage 3 uses Chai-1, a MSA-free folding model which is trained on the general PDB rather than on SAbDab alone. Agreement between Chai-1 and ABodyBuilder3 is measured either by RMSD after superposition on the VH framework or by bidirectional H3 lDDT, which is superposition-free and

bounded. These provide a measure of structural quality in the absence of a reference. All confidence scores are evaluated over the CDRH3 loop residues.

Table 8: CDRH3 confidences over the 45 targets present in all 7 arms (78,711 designs). Values are mean Chai pLDDT read from the Stage-3c complex: over CDRH3 residues within 5 Å of the antigen, over all CDRH3 residues, and over the whole Fv. *Contacting H3* is the fraction of loop residues meeting the 5 Å criterion. The final row is the between-arm range: the whole-Fv column is nearly invariant, so an effect measured on it is compressed relative to the same effect measured on the loop. Budgets rank on CDRH3 pLDDT, and the whole-Fv column is reported only as the comparison scale.

Arm	Targets	H3 contact	H3 all	Fv	Contacting H3
GFN D=4	45	74.58	80.53	93.91	47.1%
GFN D=5	45	74.14	80.10	93.84	47.0%
GFN Fixed	45	74.58	80.51	93.93	47.2%
Random baseline	45	73.12	78.87	93.58	47.6%
AMPLIFY prior	45	76.01	82.31	94.23	46.4%
ProteinMPNN	45	77.18	83.55	94.41	45.1%
ESM-IF	45	75.98	82.45	94.24	45.4%
between-arm range		4.06	4.68	0.83	

Stage 3c builds each scored complex from the Chai-1 predicted antibody structure and a crystal structure-derived antigen. Chai-1 Fv is rigid-body superimposed onto the crystal structure-derived antibody using the VH framework, after which the crystallographic antibody is removed while the antigen is retained in its experimental coordinate frame. The resulting complex therefore combines the predicted antibody structure with the experimentally determined antigen pose. This construction assumes that changes in the CDRH3 loop do not substantially alter the orientation of the VH framework relative to the antigen.

Finally, Rosetta (Alford et al., 2017) performs a constrained FastRelax of the Stage-3c complex, adapting the new CDRH3 loop to the complex by optimizing its backbone and side-chain conformations while relieving local structural strain and steric clashes, and then scores the resulting interface. FoldX (Delgado et al., 2025) subsequently repairs the relaxed complex and independently scores the interface. Binding and stability energies are reported relative to a matched WT reference ($\Delta\Delta E = \Delta E_{\text{design}} - \Delta E_{\text{WT}}$). Binding scores are additionally penalized for unfavorable steric interactions by adding the FoldX intraclash or Rosetta `fa_rep` term, respectively, to the calculated binding energy. Diversity is tracked throughout the funnel using the the number of distinct CDRH3 loop sequences, the mean per-position Shannon entropy of the CDRH3 (maximum $\log_2 20 = 4.32$ bits), and the number of clusters at 70% identity.

Table 9: Estimate of liability removal criteria impacts.

Criterion	Removes	Cull	Diversity cost
Free cysteine	any loop containing C	≈46% est.	Deletes disulfide-stapled and cyclic CDRH3 entirely — a real and productive antibody class. Pairing cannot be decided from sequence, so this rule cannot distinguish a reactive free thiol from a stabilizing intra-loop disulfide.
N-glyc sequon	N-x-[S/T], $x \neq$ P	≈5% est.	A real liability, but the motif is short and its occurrence is not correlated with function.
Deamidation NG	any loop containing NG	≈3% est.	NG occurs in the CDRs of approved antibodies; fast deamidation is a formulation problem, not a binding one.
Isomerisation DG	any loop containing DG	≈3% est.	As NG.
Soft flags	NS, NN, NT, NH; DS, DD, DT, DH; M, W; DP	none	Recorded only, no cull.

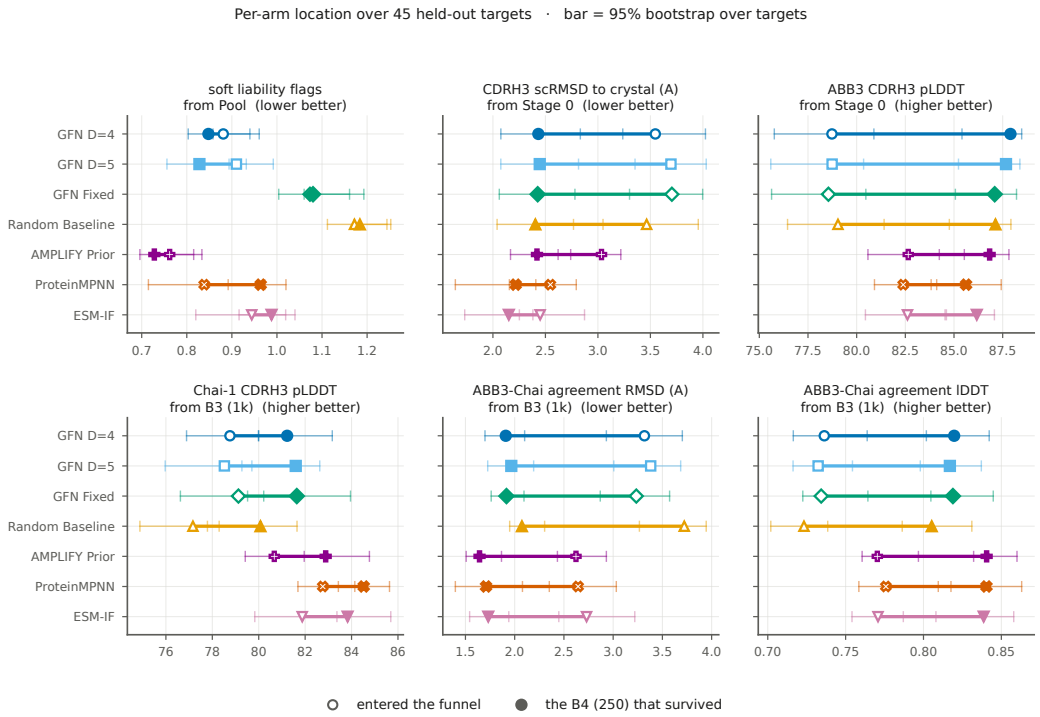


Figure 11: Scores of each arm at the beginning and end of the funnel (B4). Whiskers represent 95% bootstrap intervals over the held-out targets. Generally speaking, the improvement imposed by the funnel was greater for the random baselines and GFNs.

E Span-In-Filling vs. Masked-Language-Model Backbone

E.1 Pretraining Validation

Held-out in-filling loss, computed by separate code from the training loop, reproduces the pretraining log's own final evaluation to within 1.6% on `uniprot_dev` (1.4074 against 1.4296) and 1.8% on `scop_dev` (1.8637 against 1.8970). `oas_dev` is the one set where the two disagree (0.4902 against 0.2381), because OAS loss is bimodal — germline-conserved framework spans score near zero and CDR spans score high — so its mean summarizes a mixture. Memorization is also well controlled: on OAS, held-out data scores -0.012 nats against training data (0.4902 against 0.5022), while SCOP shows +0.737 (1.8637 against 1.1271). That SCOP gap survives length-matching and is not explained by homology: `scop_dev` entries that share a UniProt protein with `scop_train` score worse. On matched span recovery, the MLM pulls ahead in performance (Table 10, Figure 13).

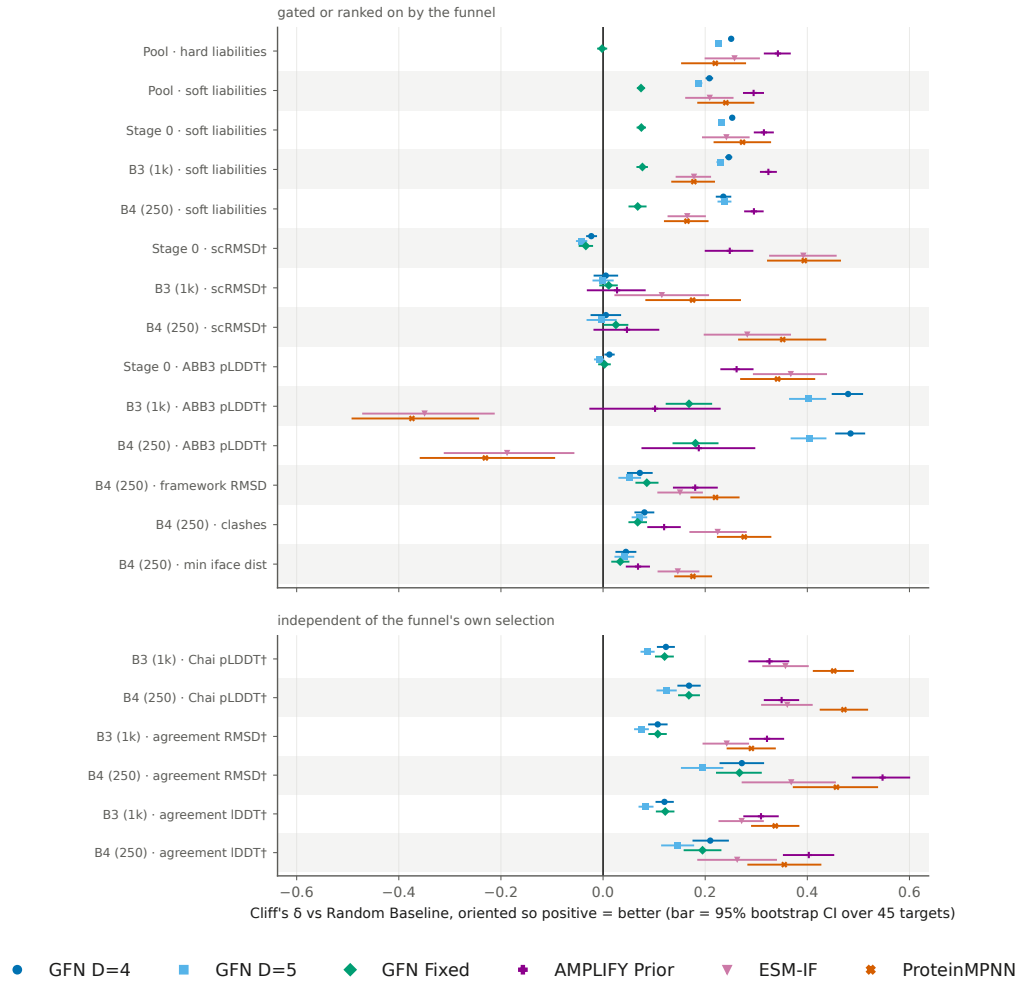


Figure 12: Every arm against the random baseline across the postfilter. Cliff's δ is computed within each target and averaged over the 45 held-out targets. Bars are 95% bootstrap CIs resampling targets. Signs are oriented so that positive is favourable. † marks metrics which reward proximity to the native structure. The lower panel holds the metrics nothing upstream in the cascade gates or ranks on. The size of the selection effect that separation exists for is visible in ABodyBuilder3 pLDDT, which moves from roughly zero at Stage 0 to $+0.5$ at B3 for the same arm.

Table 10: Span recovery against the original masked-language-model AMPLIFY-350M, at matched span length. Recovery is argmax accuracy over the 20 canonical residues. Note that the exact likelihood calculated under AMPLIFY-AR is not directly comparable to the pseudo-likelihood calculated under the original AMPLIFY. $\mathrm{AR}_{\mathrm{gen}}$ and $\mathrm{MLM}_{\mathrm{gen}}$ both generate the span conditioned only on their own predictions. $\mathrm{AR}_{\mathrm{nat}}$ and $\mathrm{MLM}_{\mathrm{nat}}$ see native context. n falls with span length because the L-forward decodings cost $O(L)$ for MLM models. We use the same sequences for all model evaluations.

			generation (self-conditioned)			native-conditioned	
Corpus	L	n	$\mathrm{AR}_{\mathrm{gen}}$	$\mathrm{MLM}_{\mathrm{gen}}$	Δ	$\mathrm{AR}_{\mathrm{nat}}$	$\mathrm{MLM}_{\mathrm{nat}}$
UniProt, ≤256 aa	1	2000	0.514	0.526	-0.012	0.514	0.526
	2	2000	0.501	0.513	-0.012	0.510	0.526
	3	2000	0.485	0.507	-0.021	0.502	0.525
	5	2000	0.481	0.488	-0.007	0.519	0.519
	8	2000	0.477	0.470	**+0.007**	0.528	0.533
	13	1230	0.454	0.424	**+0.030**	0.532	0.528
	21	761	0.450	0.386	**+0.064**	0.553	0.543
	34	470	0.400	0.304	**+0.096**	0.552	0.535
	50	400	0.373	0.226	**+0.148**	0.547	0.544
UniProt, 257–512 aa	1	2000	0.569	0.652	-0.084	0.569	0.652
	2	2000	0.555	0.638	-0.083	0.564	0.640
	3	2000	0.560	0.638	-0.078	0.576	0.652
	5	2000	0.558	0.630	-0.072	0.592	0.661
	8	2000	0.527	0.594	-0.067	0.579	0.650
	13	1230	0.544	0.567	-0.023	0.614	0.673
	21	761	0.514	0.496	**+0.018**	0.612	0.663
	34	470	0.478	0.396	**+0.081**	0.614	0.657
	50	400	0.442	0.343	**+0.099**	0.618	0.672
OAS antibody, ≤256 aa	1	2000	0.900	0.942	-0.041	0.900	0.942
	2	2000	0.885	0.932	-0.048	0.888	0.936
	3	2000	0.884	0.934	-0.050	0.896	0.938
	5	2000	0.881	0.935	-0.053	0.902	0.942
	8	2000	0.862	0.926	-0.064	0.899	0.941
	13	1230	0.846	0.915	-0.069	0.903	0.943
	21	761	0.816	0.892	-0.076	0.901	0.941
	34	470	0.778	0.837	-0.059	0.902	0.942
	50	400	0.734	0.735	-0.001	0.901	0.945
native CDRH3 (held-out)	native	55	0.121	0.314	-0.193	0.343	0.416

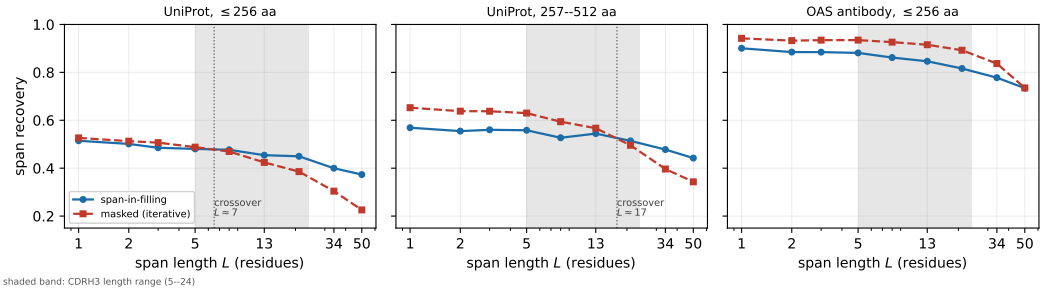


Figure 13: Span recovery against span length on the in-filling and masked-language models, both producing the span conditioned on their own predictions. On general protein sequence the masked model leads at short spans and in-filling leads at long ones, with the crossover moving from $L \approx 8$ on proteins of at most 256 residues to $L \approx 21$ on longer ones. On antibody sequences there is no crossover in range: the masked model leads marginally at every span length tested. The shaded band is the CDRH3 length range our paper's tasks operate in.

F ORACLE DETAILS

Table 11: The 14 reward objectives. Context denotes how much information in addition to the CDRH3 loop is presented to the oracle during scoring.

Category	Objective	Context	Source
Structural stability	ESM-IF	loop + backbone	ESM-IF (Hsu et al., 2022)
	ProteinMPNN	loop + backbone	ProteinMPNN (Dauparas et al., 2022)
	ProteinMPNN (Soluble)	loop + backbone	ProteinMPNN
	ProteinMPNN (Antibody)	loop + backbone	AbMPNN (Dreyer et al., 2023)
	ProteinMPNN (Ca Only)	loop + backbone	ProteinMPNN
	BoltzIF	loop + backbone	BoltzIF (Stark et al., 2025)
Naturalness	AMPLIFY-AR	loop in HC	AMPLIFY-AR (Section 3.3)
	h3berta	loop	H3BERTa (OAS CDRH3 MLM)
	ablang2	loop in Fv	AbLang-2 (Olsen et al., 2024)
Immunogenicity	MHC-II Binding	loop	MHC-II ensemble (appendix F.2)
Developability	Developability	loop in HC	Fine-tuned ESM-2 650M classifier (Lin et al., 2023)
Liabilities	Sequence Liability	loop	PTM/glycosylation motif scan (Satława et al., 2024)
Novelty	Hamming Distance	loop	edit distance to training loops
	Blosum Novelty	loop	BLOSUM62 distance to training loops

Figure 14 shows the latent substructure of the reward landscape. The strongest negative pairwise correlation on designs is -0.58, between ProteinMPNN-antibody vs. trainset novelty, which is modest, so this reward set does not contain severe *pairwise* tradeoffs — which is a caution about how much preference conditioning can be expected to do, not a defect: if objectives do not trade off, preferences have little to control. Reward sharpening does not manufacture the tensions this set lacks: it raises the response of realized rewards to preference weights without strengthening the objective-to-objective correlations (appendix G.2).

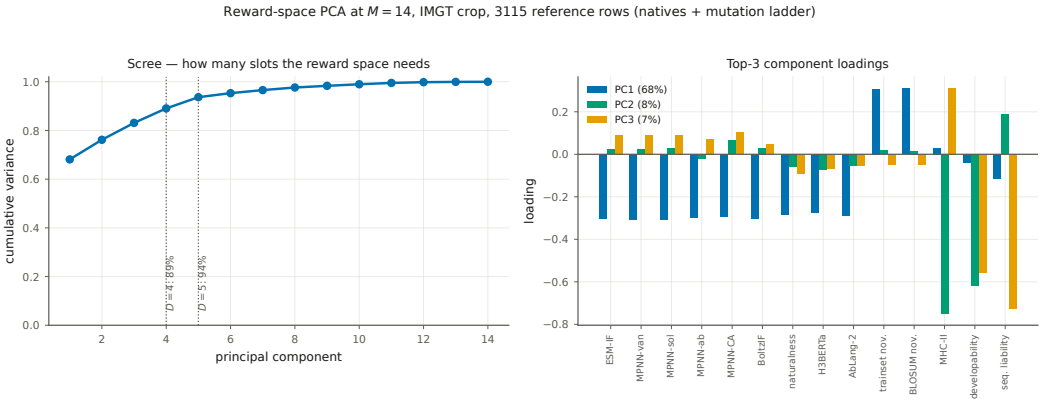


Figure 14: PCA of the reward space at $M = 14$, IMGT crop, over the natives-plus-mutation-ladder reference set ($n = 3{,}115$). Left: cumulative variance explained, with the two production slot counts marked. Right: loadings of the top three principal components on each objective. The ladder rather than natives alone, for the reason given in Figure 15: on natives both novelty terms are constant, and a PCA over them silently runs in 12 dimensions while reporting 14.

F.1 THE REWARD LANDSCAPE

The reward suite should provide complementary signals rather than redundant scores. Across the curated SAbDab set, the six stability rewards cluster together but retain some variation, while immunogenicity risk, developability, and naturalness show weak correlations with the other objectives (Figure 15). Positive correlations are stronger than negative ones throughout: 0.97 for the most aligned pair, ProteinMPNN vs. ProteinMPNN-soluble, against -0.58 for the most opposed.

Figure 20 further shows that inverse-folding recovery drops in CDRH3 relative to the framework, with higher predicted sequence variability and ensemble disagreement in the loop. This is the regime where a single stability model is least likely to be sufficient, motivating the use of multiple structure-conditioned rewards as separate objectives.

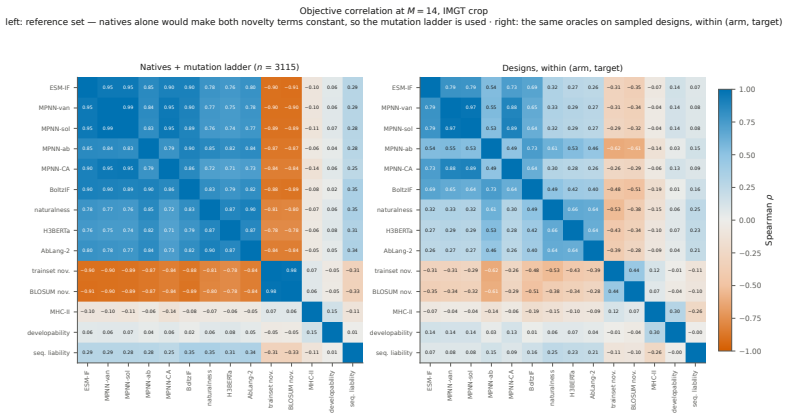


Figure 15: Spearman rank correlation between all pairs of reward objectives at $M = 14$, IMGT crop. **Left:** the reference set — natives plus a mutation ladder at rates 0.1–0.75 ($n = 3{,}115$). **Right:** the same oracles on sampled designs, computed within (arm, target) and averaged. Correlation on natural antibodies is inflated by every likelihood oracle liking real antibodies: `naturalness`, `H3BERTa` and `AbLang-2` agree at 0.87–0.90 on the reference set but only 0.64–0.66 on designs, and the two novelty terms fall from 0.98 to 0.44. On the other hand, `MPNN`/`MPNN-sol` stay at 0.97, which is genuine redundancy. The reduction in agreement of these oracles on generated designs motivates our use of this broader reward set.

Over the reference set at $M = 14$, $D = 4$ PCA components retains 89.1% of the variance and $D = 5$ retains 93.7%, so a fifth slot buys 4.6 points. D is treated as an experimental variable in Section 4.1 rather than determined by a variance target.

F.2 MHC-II Binding Oracle

The MHC-II binding oracle predicts peptide–allele binding affinity to estimate immunogenicity risk for candidate CDRH3 sequences. The model scores each candidate against a panel of 27 reference human leukocyte antigen (HLA) class II alleles covering the major HLA-DP, HLA-DQ, and HLA-DR supertypes (Greenbaum et al., 2011). The model uses ESM-2 (8M parameters, 6 layers, 320-dimensional hidden states) (Lin et al., 2023) as a frozen backbone encoder. Peptide and MHC-II pseudo-sequences are encoded independently through the backbone and mean-pooled across sequence positions (ignoring padding). The two 320-dimensional representations are concatenated into a 640-dimensional vector and passed through an MLP with two hidden layers ($640 \to 512 \to 512 \to 1$), GELU activations, dropout ($p = 0.25$), and a sigmoid output activation. The backbone is frozen throughout training. Only the MLP head (591K parameters) is optimized. Total model size including the frozen backbone is 8.4M parameters.

We train on binding affinity data from NetMHCIIpan 4.0 (Reynisson et al., 2020), which provides $\log_{50k}$-transformed binding affinities ($y = 1 - \log_{50k}(\text{IC}_{50}\ \text{nM})$, $y \in [0, 1]$) for peptide–allele pairs across 79 HLA class II alleles. We perform 5-fold cross-validation using the provided BA1–BA5 partitions ($\sim$87K training samples per fold). The model is trained for 40,000 steps with batch size 128 using AdamW (lr $= 10^{-3}$, no weight decay) and mean squared error (MSE) loss. A linear warmup schedule (500 steps) followed by linear decay is applied.

At inference time, the five fold-specific heads share the frozen backbone and their output probabilities are averaged. Each candidate CDRH3 is scored against all 27 reference alleles in parallel. Higher rewards correspond to lower predicted immunogenicity. Table 12 reports 5-fold cross-validation performance of the ensemble-averaged model evaluated on each fold's held-out test set. Binder classification uses the standard 500 nM threshold ($y > 0.426$). Figure 16 shows the label distribution

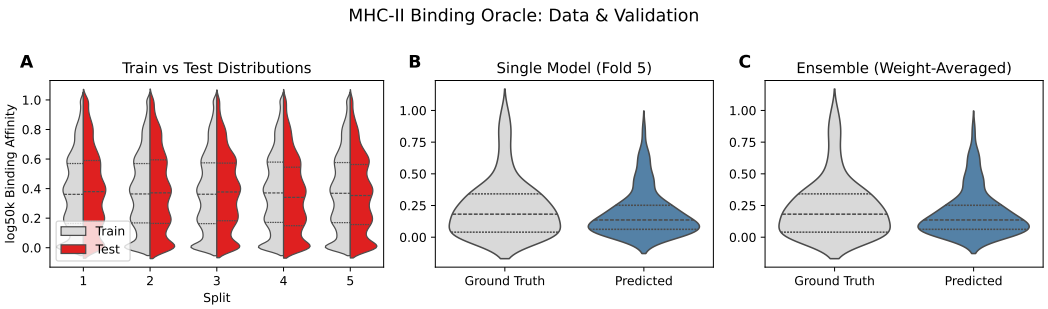


Figure 16: MHC-II binding affinity distributions across the 5 cross-validation splits, comparing train (gray) and test (red) partitions (left). MHC-II oracle validation: predicted versus ground-truth binding affinity distributions for a single-fold model (middle) and the weight-averaged ensemble (right), showing little difference across folds. The shipped oracle averages output probabilities instead.

Table 12: MHC-II binding oracle: 5-fold cross-validation results (ESM-2 8M backbone, frozen, MSE loss).

Fold	AUROC	MSE
1	0.833	0.039
2	0.879	0.033
3	0.862	0.031
4	0.868	0.030
5	0.866	0.033
Mean $\pm$ std	0.862 ± 0.015	0.033 ± 0.003

across the 5 splits and final ensemble performance against the labeled ground truth. Our oracle achieves good median performance but under-predicts extreme positive binding affinity values.

For efficiency at training-loop scale, the 27 allele pseudo-sequence embeddings are cached once, reducing molecule-side forward passes per scoring call from $B \times 27$ to zero and peptide-side from $B \times 27$ to B. The cached path is verified exactly identical to the uncached path (maximum absolute difference 0.0 on a 256-candidate batch).

F.3 Developability Oracle

We use a SAbDab-derived dataset (Chen et al., 2020) that assigns binary developability labels to 2,409 antibodies based on the Developability Index (Lauer et al., 2012). The original definition of the index combines the spatial aggregation propensity of the CDRs with net charge computed on the full-length antibody structure. Chen et al. (2020) computes it from variable-region structures alone, so the net-charge term in these labels is a variable-region quantity. Antibodies in the bottom 20% of Developability Index scores, corresponding to the lowest aggregation risk, are labeled developable. This yields an imbalanced dataset with approximately 482 positives and 1,927 negatives.

The developability classifier fine-tunes ESM-2 650M (Lin et al., 2023) with a BOS-token-pooled sequence classification head and 2-class output. The heavy-chain sequence is provided as input. Only the head and the layer-normalization parameters are fine-tuned, yielding an efficient adaptation of the pretrained backbone. Training uses 10-fold stratified cross-validation for 15 epochs with the Adam optimizer (Kingma & Ba, 2015) (learning rate 5×10^{-4}), batch size 8 via gradient accumulation, and early stopping with patience 10.

Note that this oracle is a sequence classifier of *ease of synthesis* and contains no spatial-aggregation or charge-patch descriptors. It is nontheless correlated with the the structural aggregation signal supplied in the postfilter analysis.

F.4 Naturalness and Antibody-Repertoire Oracles

The naturalness oracle is a frozen copy of the AMPLIFY-350M span-in-filling model used to initialize our GFN. A candidate CDRH3 loop is scored by rendering the heavy chain as `<bos> prefix <mask> suffix <sep> CDRH3 <eos>` and computing the length-normalized log-likelihood of the CDRH3 tokens under teacher forcing. This gives a direct estimate of $p(\mathrm{CDRH3} \mid \mathrm{framework})$ under the same factorization used by the generator. As masked models score loops through leave-one-out pseudo-likelihood (Salazar et al., 2020), we measured a $10.7\times$ throughput penalty at the median CDRH3 length compared to the autoregressive variant (Section 3.3). The two antibody-specific likelihood oracles are described in Section A.

F.5 Sequence Liability Motifs

The motif catalog is adapted from the Liability Antibody Profiler (LAP) (Satława et al., 2024), which catalogs 70 motifs and retains 30 that either recur across sources or carry strong evidence. Hard-flagged (gate-rejecting in the post-filter, maximally weighted in the reward): deamidation `NG`, isomerization `DG`, N-linked glycosylation sequon `N-x-[S/T]`, and free or unpaired cysteine, which LAP likewise treats as severe. The cysteine rule is a deliberately conservative sequence-only heuristic — every cysteine in the designed loop is hard-flagged, since pairing cannot be determined from sequence. Soft-flagged (recorded and weighted, not gating): slow deamidation and isomerization variants, methionine and tryptophan oxidation, and `DP` fragmentation. The scanner emits per-motif occurrence counts, so the same code serves as a training reward, a post-filter gate, and an evaluation statistic. LAP is a screening tool which consumes a paired variable domain, folds it with ABodyBuilder2, and returns a flag list.

Our sequence-based scan is unable to identify solvent-based scans as it flags buried liability counts identically to exposed ones. Second, we do not include the "germline" or "therapeutic" exclusions to prevent our sampler from being biased towards those distributions. Third, we only evaluate the designed region and treat this score as a binary accept/reject criteria.

F.6 Novelty Oracles

Both novelty oracles precompute their reference set once at task setup from the *resolved training split*, never from the full universe — assembling the reference by hand is only correct when training on all targets, and is silently wrong for any subset. The reference is bucketed by loop length, both for speed (distance is at least $|\Delta L|$, so most buckets prune immediately) and, for the BLOSUM variant, because distance is minimized over training loops of the same length only.

CDRH3 length is a primary determinant of loop conformation, so comparing within a length class avoids introducing gap-open and gap-extend penalties as two free parameters that would then drive most of the length-dependent behaviour. The training set spans every length from 5 to 24 with no gaps. Note that both novelty rewards are *constant on the native reference set by construction* (a native loop scored against a reference containing itself is exactly zero), so they must be scored leave-one-out when building the offline reference matrix for the projection warm start, or they vanish from the spectrum entirely.

F.7 Stability Oracle

For a model family m with weight variants $\mathcal{V}_m$ and $E = 5$ ensemble members per variant, the operational stability reward averages over both levels:

$$R^{(m)}_{\mathrm{stab}}(\mathbf{x}, s) = \frac{1}{|\mathcal{V}_m|} \sum_{v \in \mathcal{V}_m} \frac{1}{L} \sum_{i=1}^{L} \log \mathrm{softmax}\left(\frac{1}{E} \sum_{e=1}^{E} \mathbf{g}^{(m)}_{v,e}(s) \right)_{x_i, i} \tag{8}$$

where $\mathbf{g}^{(m)}_{v,e}(s) \in \mathbb{R}^{|\mathcal{A}| \times L}$ are the precomputed logits for model family m, variant v, ensemble member e, and structure s (here L is the CDRH3 length after slicing from full-chain logits). The mean over ensemble members is taken before the log-softmax.

The six model families are correlated but not interchangeable: ProteinMPNN-Vanilla and ProteinMPNN-Soluble measure $+0.97$, but ESM-IF↔Vanilla measures 0.76 while Vanilla↔CA-only

measures 0.73, so no correlation threshold separates them cleanly, and ProteinMPNN-Antibody tracks *naturalness* ($+0.60$) rather than the other structural models. Any merge is therefore an arbitrary decision about which biological signals are "the same". This is precisely the decision the learned projection in Section 3.2 makes soft, data-driven, and reversible instead.

Structural stability of possible loop changes was assessed as the likelihood of loop sequences under structure-based inverse-folding models: ESM-IF (Hsu et al., 2022), ProteinMPNN (Dauparas et al., 2022) variants, antibody ProteinMPNN (Dreyer et al., 2023), and BoltzIF (Stark et al., 2025). ESM-IF is a 142M-parameter GVP-Transformer inverse-folding model. Predictions use all-atom coordinates from ATOM records, and ensembles are generated by adding Gaussian coordinate noise with $\sigma_{\text{noise}} \in \{0, 0.0375, 0.075, 0.1125, 0.15\}$ Å. BoltzIF is the inverse-folding model from the BoltzGen framework; ensembles are generated with 5 forward passes using different random seeds.

ProteinMPNN was used with four weight variants trained on different data distributions. **Vanilla** was trained on all PDB assemblies resolved to better than 3.5 Å (clustered at 30% sequence identity), using full backbone atom coordinates (N, C_α, C, O, and a virtual C_β). **Soluble** uses the same architecture and featurization as Vanilla but excludes trans-membrane proteins from the training set, biasing surface residue preferences toward soluble compositions. **Antibody** (AbMPNN) (Dreyer et al., 2023) is fine-tuned from Vanilla weights in two stages: first on ≈108k paired variable region clusters from the Observed Antibody Space predicted by ABodyBuilder2, then on ≈1,700 experimentally resolved antigen-binding clusters from SAbDab. **CA-only** uses only C_α coordinates as input, replacing inter-atom distances with orientation features derived from consecutive C_α displacement vectors. This variant is applicable to low-resolution or coarse-grained structures where full backbone coordinates are unavailable. Each variant was trained with multiple levels of Gaussian backbone noise (0.02–0.30 Å for Vanilla and Soluble, 0.02–0.20 Å for CA-only, and 0.0–1.0 Å for Antibody), producing multiple sets of weights per base model.

We validate these models in Figure 17 by assessing sequence recovery for the CDRH3 loop and the surrounding framework under each for the full curated SAbDab set we used for our experiments. Panel A of Figure 17 shows the distribution of recovery rates per model (where "Base Model" indicates the model with no extra noise added during training, where applicable). Panel B shows the relationship between logit kurtosis (a proxy for model confidence) and recovery. We find a strong correlation here for all models. Panel C shows the progressive reduction in recovery seen for ProteinMPNN models trained with increasing levels of additive noise, and panel D shows the expected drop in recovery for these models in the more variable CDRH3 region. Figure 18 shows cross-model correlations of the recovery statistics (A), the effect of added structural noise (B), and cross-model prediction agreement between inverse folding models on the full sequence (C) and the CDRH3 (D), showing a substantial drop in agreement between models in the more variable CDRH3 region. Taken together, these results show that these inverse folding models all capture different statistics of the structural properties of the CDRH3 region, supporting our decision to use them all as unique reward models in the training of our GFN model. See Figure 19 for a more detailed view of the relationship between measures of the predicted logit distribution, recovery rate, and ensemble stability. In general, model confidence measures (KL Divergence, kurtosis) correlated strongly with the recovery rate while cross entropy did not. Ensemble stability was uncorrelated with recovery aside from the ESM-IF model.

F.8 Length Dependence of Rewards

In general, rewards did not correlate with sequence length, with some exceptions at extremely short loop lengths. See Figure 21 for a per-reward breakdown. Length dependence nonetheless dominates several *pipeline* quantities — Stage-0 liability survival spans 19–61% across targets and tracks CDRH3 length at $r = -0.97$ — which is why every comparison in this work is paired per target.

G Hyperparameter Tuning

G.1 Learning Rate, on the Objective-Slots Arms

10 arms, identical except for the policy learning rate, at each of $D = 4$ and $D = 5$ on the $M = 14$ objective set: 3000 steps, $\beta = 8$, the scaffold-leakage split, and every other setting held at the values in Appendix A. The grid brackets the learning rate $\pi_{\text{lr}} = 1 \times 10^{-4}$ on both sides as well.

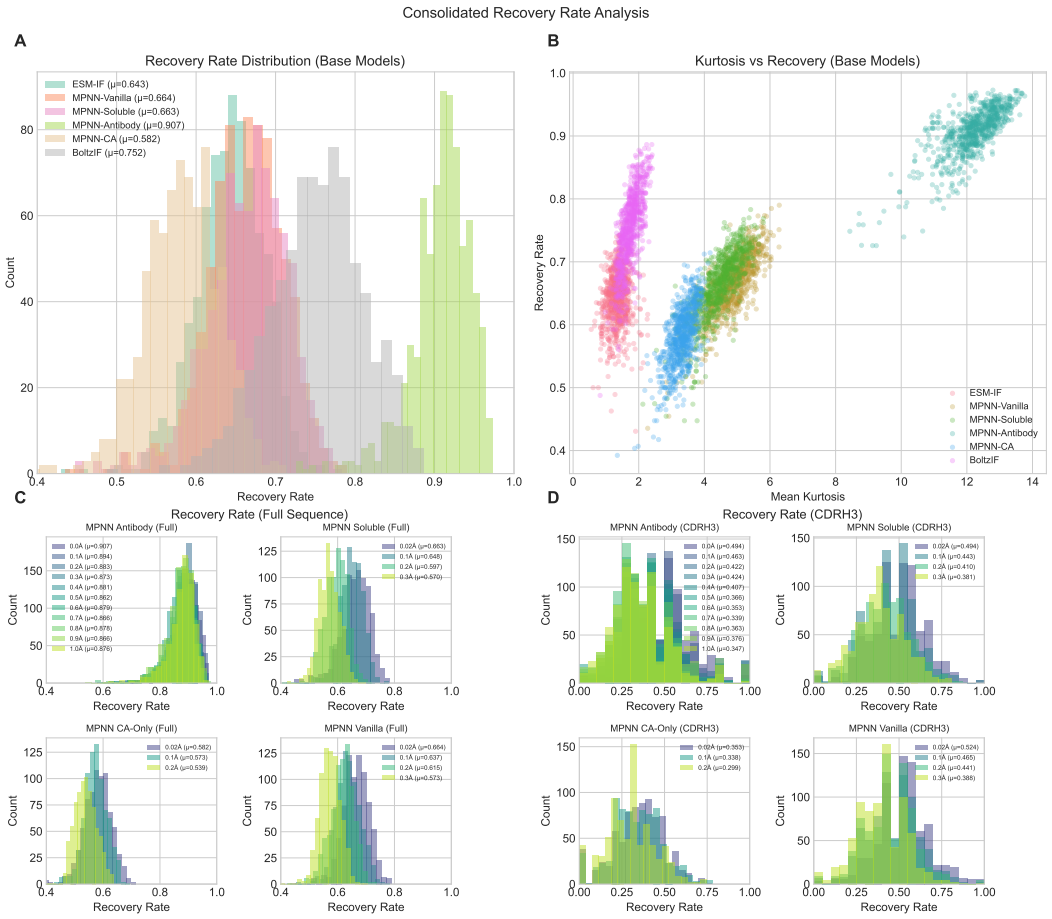


Figure 17: Native sequence recovery rates across inverse folding models. Recovery is the fraction of positions where the model's top prediction matches the native amino acid. A shows the distribution of recovery rates across all base inverse folding models (no extra noise added) on all SAbDab entries. ProteinMPNN-Antibody clearly has the highest recovery rate (this data is in-distribution for the model, so this is expected). Panel B shows the model confidence (measures via logit kurtosis) against recovery rate, which generally show strong positive correlations. Panel C shows the distribution of recovery rates for each ProteinMPNN model with varying levels of structural noise added across the full sequence. Additional noise produces a leftward shift in the distribution of recovery rates. Panel D shows the same analysis in just the CDRH3 loop region, showing a large and expected drop in recovery rate due to the inherent variability of this structure.

$\pi_{\text{lr}} = 1 \times 10^{-4}$ is selected at both D, by at least 3.5 block standard errors at $D = 4$ and 3.4 at $D = 5$, and the ordering is a clean single optimum degrading in both directions. Both production arms therefore use the anchor value: the sweep confirmed it rather than displacing it. Note that the two D columns are not comparable to each other — reward is measured in each arm's own slot space, and D changes that space.

G.2 Reward Sharpening: The β Sweep

Three arms identical except for the reward-sharpening exponent β, one seed, 64 Dirichlet preferences $\times$ 512 samples $\times$ 20 evaluation targets = 655,360 designs per arm. Controllability is the per-axis diagonal Spearman correlation between preference weight and realized reward.

Controllability rises monotonically with β on every axis. Mean |diagonal| 0.013 $\rightarrow$ 0.110 $\rightarrow$ 0.204 — so near-zero controllability at $\beta = 1$ is substantially a β effect rather than a conditioning-architecture

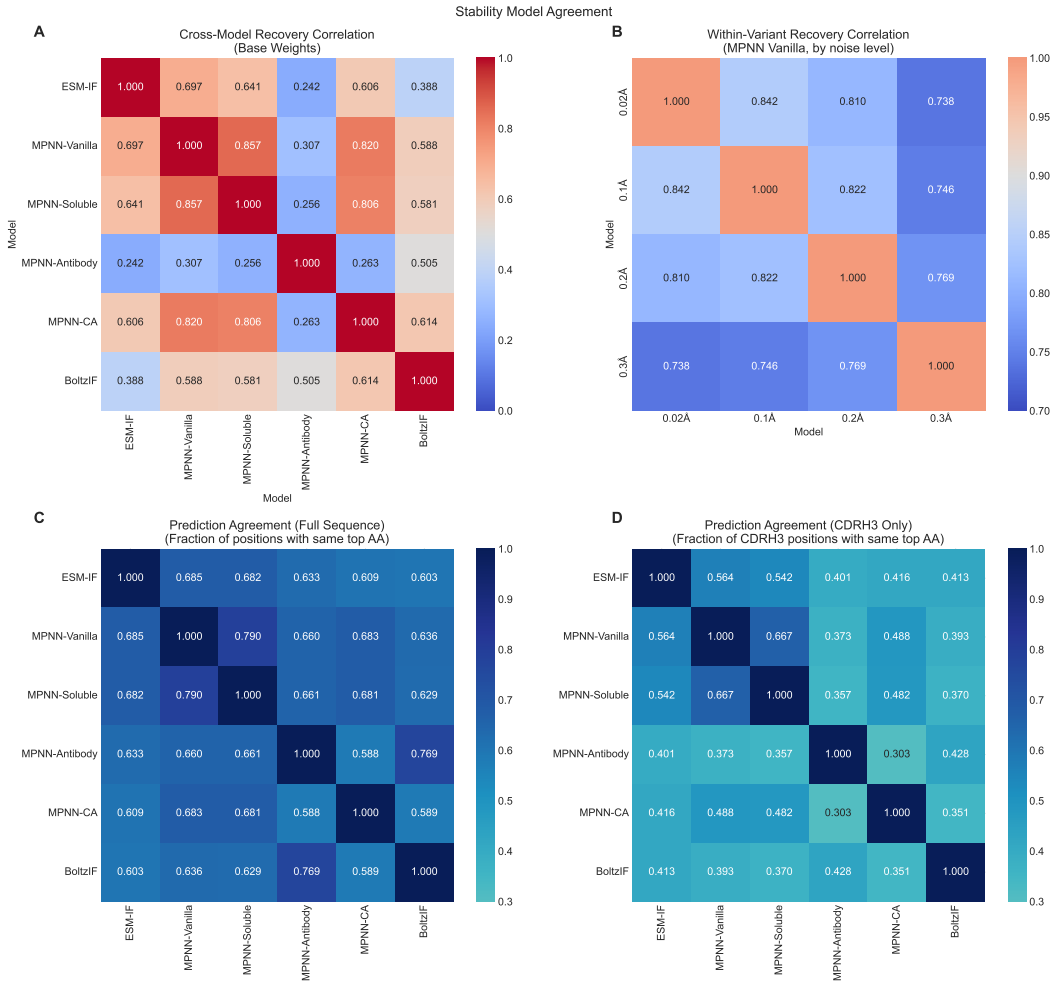


Figure 18: Cross-stability model correlation and agreement. Pairwise Spearman correlations of per-structure recovery rates and per-position prediction agreement across base models. Panel A compares the base model of each major class of inverse folding model, showing their general disagreement (while correlations are modest, each model clearly captures different signal). Panel B shows the gradual drop in recovery agreement with additional structural noise added (this pattern is consistent across ProteinMPNN models, data not shown). Panel C shows prediction agreement (% of positions with the same top amino acid predicted) across all base models for the entire sequence, where agreement is generally high. Panel D shows the same analysis restricted to the CDRH3 loop, showing these models capture different statistics in this functionally more variable region.

limit. The effect is large relative to the split difference between the arms, but it is not independent of it, and the measurement deserves repeating on the leakage-free split.

There is no mode collapse at high β, though we began to observe some diversity loss. At $\beta = 32$, 0.997 of designs are unique and 0.935 are distinct modes at 70% CDRH3 identity, against 1.000 and 0.999 at $\beta = 1$; global objective coverage *rises* (0.179 $\rightarrow$ 0.207) while within-preference local spread falls 9.2%. Increasing β did not have an adverse effect on training dynamics. While $\beta = 32$'s trajectory-balance residual spread is far larger in absolute terms (4.72 against 0.86), in its scale-free form, the fitted trend over the last 40% of training was roughly the same between $\beta = 8$ (-10.9%) and $\beta = 32$ (-11.5%).

Our data does not show that objectives develop genuine tensions at high β. Sharpening does raise the off-diagonals of the preference-to-reward matrix: at $\beta = 32$, preference on immunogenicity against realized naturalness is -0.366 and against realized stability -0.272, from -0.019 and -0.024 at

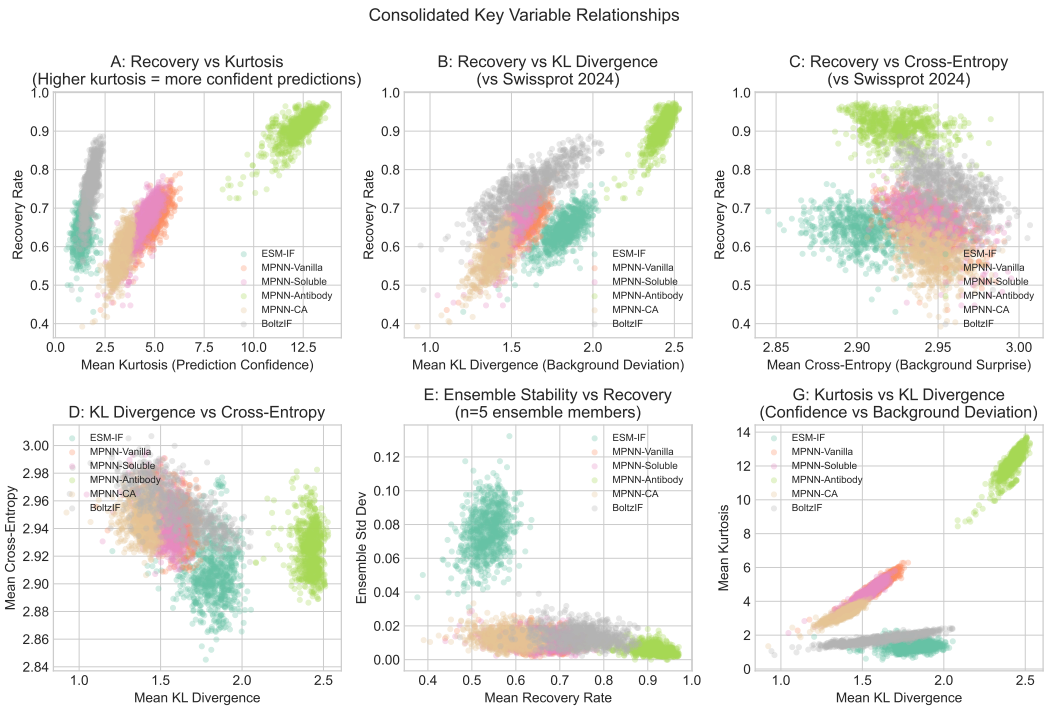


Figure 19: Key variable relationships across inverse folding models: recovery versus KL divergence, entropy, kurtosis, and ensemble variability.

D	π_{lr}	Mean reward	Block SE	TB loss	Δ vs. best (σ)
4	1×10^{-5}	-0.8027	0.0037	0.1155	0.0224 (4.8)
	3×10^{-5}	-0.7950	0.0030	0.0930	0.0146 (3.5)
	$\mathbf{1 \times 10^{-4}}$	**-0.7803**	0.0028	**0.0792**	—
	3×10^{-4}	-0.8115	0.0040	0.1330	0.0312 (6.3)
	1×10^{-3}	-0.8623	0.0039	0.1906	0.0820 (16.9)
5	1×10^{-5}	-0.7608	0.0028	0.0954	0.0232 (5.1)
	3×10^{-5}	-0.7545	0.0035	0.0826	0.0169 (3.4)
	$\mathbf{1 \times 10^{-4}}$	**-0.7376**	0.0036	**0.0727**	—
	3×10^{-4}	-0.7593	0.0032	0.0908	0.0217 (4.5)
	1×10^{-3}	-0.7956	0.0040	0.1454	0.0580 (10.8)

Table 13: Learning-rate sweep on the $M = 14$ objective-slots arms. Mean reward and trajectory-balance loss over the final 250 training steps; Δ vs. best is the reward gap to that D's selected arm, in units of the pooled block standard error.

$\beta = 1$. But those off-diagonals carry the simplex confound of appendix C — raising one preference weight necessarily lowers the others — so any policy that responds to preferences at all must produce them, and they say nothing about the objectives. The objective-to-objective correlations, where a real trade-off would appear, move the other way: immunogenicity against naturalness reads -0.169 at $\beta = 1$ and -0.112 at $\beta = 32$, weaker rather than stronger, and immunogenicity against stability -0.057 $\rightarrow$ -0.107. Sharpening makes the policy *steerable*; it does not make the objectives *conflict*.

On trained-objective metrics alone, this table selects $\beta = 32$. However, as our goal was to emphasize the value of diversity in building loop samplers, we opted to use $\beta = 8$. Future work should explore the performance benefits of using sharpened rewards in this setting for improved steerability and, with the right optimization methods, coverage of the reward landscape. For this work, it was not selected largely because the generate-then-filter workflow's value depends on *pool breadth reaching*

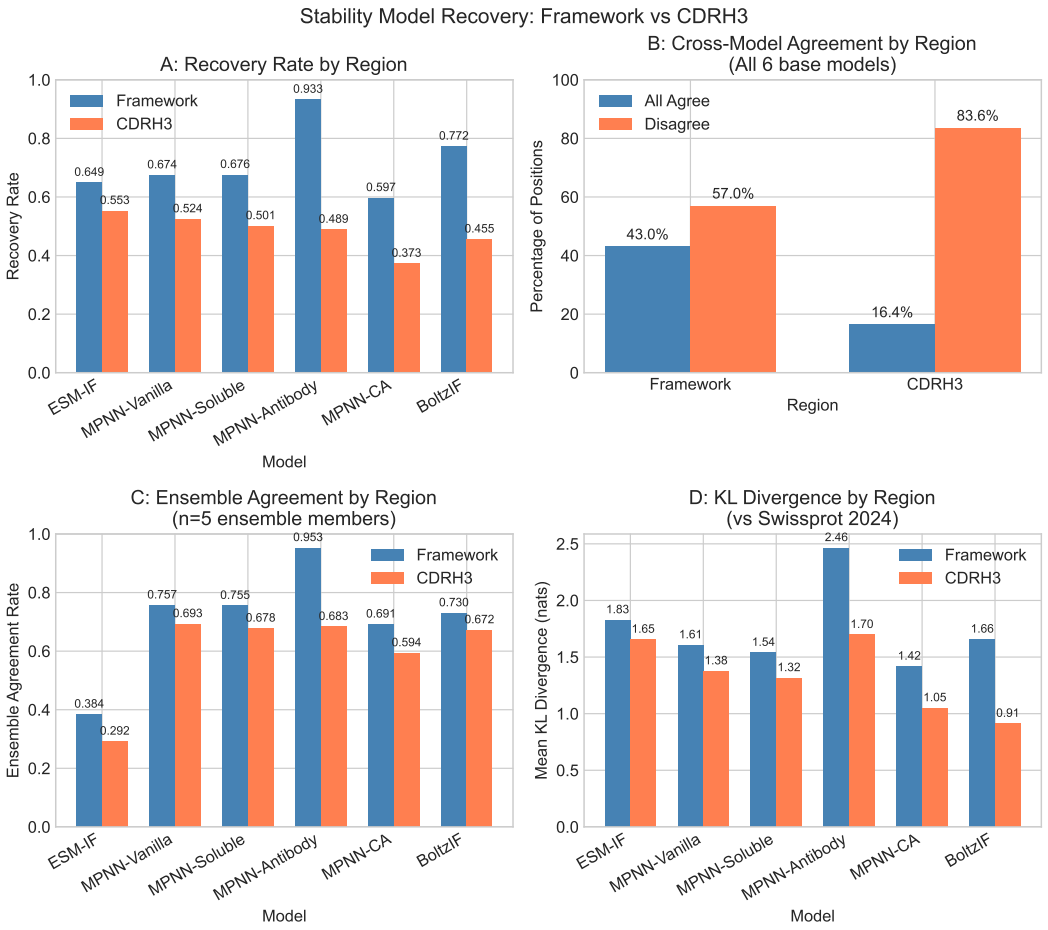


Figure 20: Regional comparison of inverse folding predictions: framework versus CDRH3. Models show consistently lower recovery, higher entropy, and higher KL divergence in the CDRH3 region.

axis / statistic	$\beta = 1$	$\beta = 8$	$\beta = 32$
immunogenicity	0.021	0.127	0.247
developability	0.019	0.176	0.266
naturalness	0.004	0.061	0.200
stability	0.009	0.076	0.103
mean |diagonal|	**0.013**	**0.110**	**0.204**
signal-to-noise (diagonal / off-diagonal)	1.49	1.40	1.57
mean pool reward	0.435	0.521	0.656
top-decile pool reward	0.569	0.669	0.764
unique-sequence fraction	1.000	1.000	0.997
within-preference pairwise Hamming	12.1	12.0	11.0
distinct-mode fraction @70% identity	0.999	0.996	0.935
global objective coverage (mean σ)	0.179	0.201	0.207
within-preference local spread	0.146	0.160	0.145
hypervolume (training, final)	0.503	0.875	0.904
trajectory-balance residual spread (final)	0.13	0.86	4.72
relative change, last 40% of training	-4.5%	-10.9%	-11.5%

Table 14: Effect of reward sharpening β on controllability, pool quality and diversity.

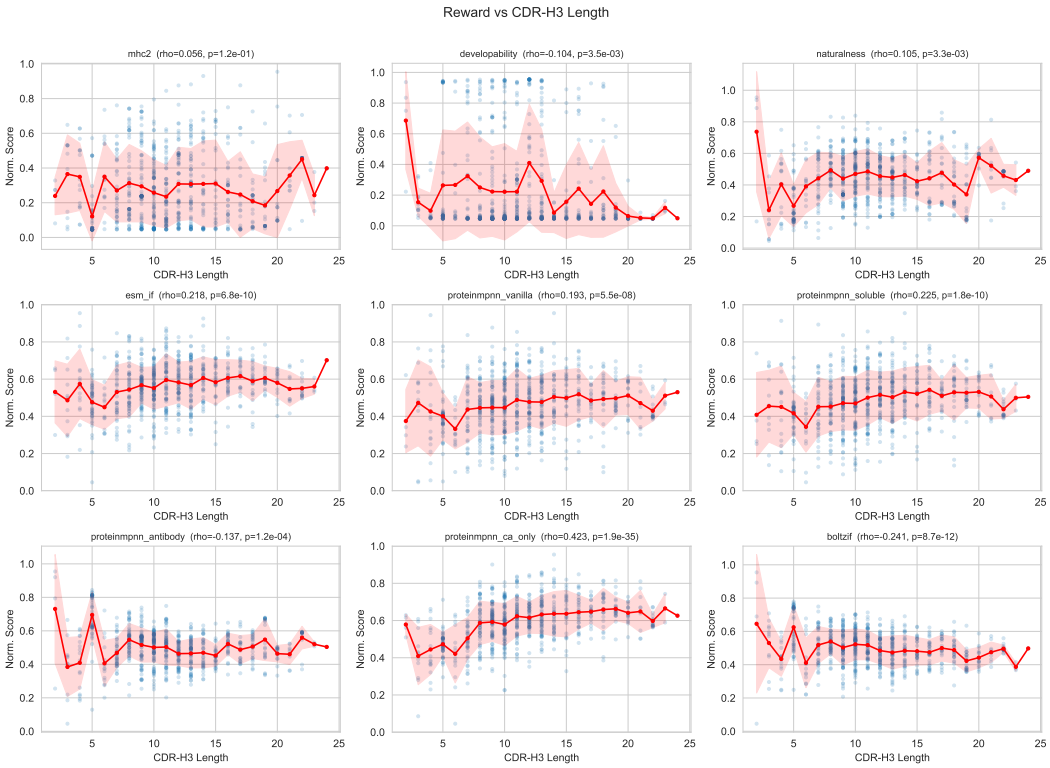


Figure 21: Normalized reward scores as a function of CDRH3 loop length. Red lines show mean per length with ± 1 standard deviation bands. Spearman ρ and p-values are shown per objective.

the post-filter, and $\beta = 8$ was seen to be on the operating point of the controllability↔exploration trade-off for this workflow. In addition, due to compute constraints, many promising avenues were not able to be explored in this work - the reader should not interpret the lack of these experiments as the authors believing they are not worth performing.

Absolute controllability remains modest even at $\beta = 32$ (mean 0.204; the stability axis alone 0.103), so while these analyses shows *β-dependence of controllability*, it does not show strict controlability over the generations. All metrics are on trained objectives only. The lack of training collapse is conditional on the dual-cross-attention span-in-filling architecture adopted here – earlier architectures *did* collapse at $\beta = 32$ during training.