DIFFANTISEQ: TARGET-STEERED DIFFUSION IN LATENT SEQUENCE SPACE FOR ANTIBODY LIBRARY DESIGN

Fang WU
Department of Computer Science
Stanford University
fangwu97@stanford.edu

ABSTRACT

Antibodies are among the most versatile molecules in therapeutic discovery, yet computational antibody library design remains challenging when evolutionary signals from multiple sequence alignments are sparse or unreliable. We present DiffAntiSeq, a controllable diffusion-based generative framework for efficient, target-specific antibody sequence design. DiffAntiSeq performs non-autoregressive denoising in a continuous latent residue embedding space, enabling global sequence refinement beyond the limitations of autoregressive or discrete diffusion models. To steer generation toward desired functional outcomes, we incorporate gradient-based classifier guidance derived from protein language models trained to predict antibody–antigen binding affinity and specificity. We evaluate DiffAntiSeq using large-scale antibody sequence and binding data from the AlphaSeq platform, and apply it to the design of thousands of single-chain variable fragment antibodies targeting a SARS-CoV-2 peptide. Across extensive in silico analyses and structure-based validation, DiffAntiSeq consistently outperforms state-of-the-art machine-learning-driven evolution methods, producing antibody libraries with substantially stronger binding while maintaining meaningful sequence diversity. These results demonstrate that controllable diffusion in continuous latent sequence space provides an effective and scalable paradigm for antibody library design in data-sparse and structure-limited settings.

1 INTRODUCTION

Therapeutic antibodies are among the most successful classes of biologics due to their high binding specificity, favorable safety profiles, and broad applicability across oncology, immunology, and infectious diseases. However, the space of possible antibody sequences is astronomically large, rendering exhaustive experimental screening infeasible. As a result, efficient computational methods for antibody library design are critical for accelerating early-stage discovery by prioritizing diverse yet functionally optimized candidates [18, 31].

A central challenge in antibody design is modeling the complementarity-determining regions (CDRs), which govern antigen recognition and exhibit high variability in both length and sequence [40]. Traditional approaches often rely on multiple sequence alignments (MSAs) to infer positional constraints and co-evolutionary signals. While effective for conserved protein families, alignment-based methods struggle in antibody settings where evolutionary signals are sparse, CDRs are hypermutated, and insertions and deletions are common. These limitations substantially restrict their ability to guide target-specific optimization.

Recent advances in deep learning have opened new avenues for molecular, antibody and protein design [14, 22, 51, 42, 52, 43, 45, 44, 48, 41, 39, 50, 47, 49, 5, 1, 37]. Structure-aware models leverage geometric deep learning to co-design antibody sequences and structures or optimize binding interfaces directly [12, 23, 15, 17]. Although powerful, these methods typically require high-quality antigen or antibody-antigen complex structures, limiting their scalability and applicability in early discovery stages where such structures are unavailable or uncertain. In contrast, sequence-only

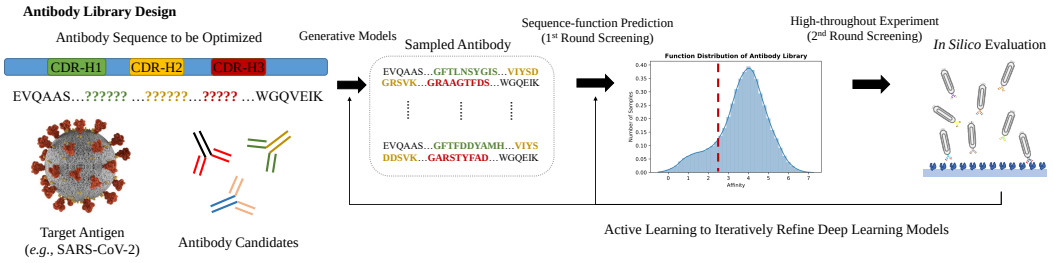

Figure 1: Illustration of the antibody library design task. Given a target antigen, the goal is to generate a diverse set of antibody sequences that can bind to the epitope with high affinity. This involves designing the complementarity-determining regions (CDRs) of the antibody while ensuring structural stability and manufacturability. Subsequently, active learning is employed iteratively to refine the antibody library by selecting high-quality candidates for further optimization.

approaches based on protein language models (PLMs) [46, 2, 32] avoid explicit structural inputs and scale well with data, but they are predominantly autoregressive. This design choice imposes a fixed generation order, introduces exposure bias through cumulative decoding errors, and makes it difficult to globally revise sequences during optimization.

Diffusion-based generative models offer a compelling alternative by enabling non-autoregressive, full-shot generation through iterative denoising. Recent work has successfully applied diffusion to protein structures and sequences, demonstrating improved global coherence and controllability [10, 25, 36]. However, existing diffusion approaches for proteins either operate in discrete token spaces with limited expressiveness or lack effective mechanisms for target-specific functional control. As a result, unconditional or weakly conditioned diffusion models often fail to generate antibodies with strong, specific binding properties.

This work introduces **DiffAntiSeq**, a controllable diffusion framework for antibody library design that operates in a continuous latent sequence space. DiffAntiSeq performs denoising diffusion over learned residue embeddings rather than discrete amino acid tokens, enabling richer noise modeling and more flexible sequence refinement. To steer generation toward desired functional outcomes, we integrate gradient-based classifier guidance derived from protein language models trained to predict antibody-antigen binding affinity and specificity. This plug-and-play control mechanism allows DiffAntiSeq to balance functional optimization with evolutionary plausibility during sampling.

We evaluate DiffAntiSeq on large-scale antibody sequence and binding data from AlphaSeq, focusing on the design of single-chain variable fragment (scFv) libraries targeting a SARS-CoV-2 peptide. Across extensive in silico and structure-based evaluations, DiffAntiSeq consistently outperforms state-of-the-art machine-learning-driven evolution methods, producing antibody libraries with stronger binding affinity while maintaining substantial sequence diversity. In summary, our contributions are threefold:

- We propose a continuous latent diffusion model for antibody sequence generation that avoids the limitations of autoregressive decoding and discrete diffusion.
- We introduce a target-specific, gradient-guided control strategy using protein language models to steer antibody generation toward improved binding properties.
- We demonstrate, through large-scale experiments and structural validation, that DiffAntiSeq enables efficient construction of high-quality, diverse antibody libraries.

2 METHOD

2.1 PRELIMINARY

Task Formulation. An antibody sequence is composed of n residues as $\mathbf{a} = [a_1, ..., a_n] \in \mathcal{A}$, where $a_i \in \mathcal{V}$ is an amino acid. $\mathcal{V}$ is the token vocabulary that consists of twenty amino acids and four auxiliary tokens (*i.e.,* 'PAD', 'END', 'START', 'UNKNOWN'). The CDRs of this antibody are a m-length subsequence of $\mathbf{a}$ denoted as $\mathbf{b} = [b_1, ..., b_m]$, where $b_i = a_{e_i}$ and e_i is the index of

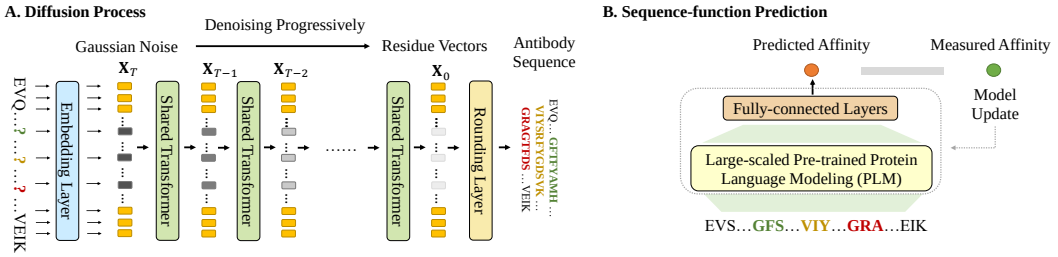

Figure 2: **A.** The diffusion iteratively denoises the antibody's CDR sequence in the latent space, yielding an intermediate latent variable of decreasing noise level $\mathbf{x}_T ... \mathbf{x}_0$. A final rounding layer then transfers the residue vectors to discrete antibody sequences. **B.** Designed antibody sequences are forwarded into protein language models with a fully-connected layer to forecast the functionality. Quantitative affinity data obtained from high-throughput experiments are used to guide training.

CDR residue b_i in the antibody $\mathbf{a}$. The antigen sequence consists of n' amino acids, represented as $\mathbf{c} = [c_1, ..., c_{n'}] \in \mathcal{C}$, and its corresponding structure is denoted as $\mathcal{G}_{\text{ag}}$, which can be obtained by X-ray crystallography or computational tools like AlphaFold [14].

Controllable antibody generation refers to the task of sampling antibody sequences $\mathbf{b}$ from a PLM, represented as a conditional distribution $p_{\text{PLM}}(\mathbf{a} \mid \mathbf{c})$. In some settings, the antibody sequence $\mathbf{a}$ is partially known, and the goal shifts to optimizing specific regions. This leads to conditional generation of CDRs $\mathbf{b}$ via $p_{\text{PLM}}(\mathbf{b} \mid \mathbf{a}, \mathbf{c})$ or $p_{\text{PLM}}(\mathbf{b} \mid \mathbf{a} - \mathbf{b}, \mathbf{c})$, where $\mathbf{a} - \mathbf{b}$ denotes the framework region. We extend this formulation by introducing a ground-truth mapping function $f : \mathcal{A} \times \mathcal{C} \to \mathcal{Y} \subset \mathbb{R}$, where $\mathcal{A}$ and $\mathcal{C}$ denote the spaces of antibody and antigen sequences, respectively, and $f(\cdot)$ evaluates properties such as binding affinity or specificity for an antibody-antigen pair $(\mathbf{a}, \mathbf{c})$.

Our objective is to train a generative model $\mu_\theta(\cdot \mid \mathbf{c}, \mathcal{G}_{\text{ag}})$ that, conditioned on the antigen sequence $\mathbf{c}$ and its structure $\mathcal{G}_{\text{ag}}$, can construct a virtual antibody library $\mathcal{V}_\mathbf{a}$ of size at most K. The goal is to maximize the average binding score across the generated antibodies while promoting diversity. Formally, we solve $\max_{\mathcal{V}_\mathbf{a}} \mathbb{E}_{\mathbf{a} \in \mathcal{V}_\mathbf{a}}[f(\mathbf{a}, \mathbf{c}, \mathcal{G}_{\text{ag}})]$, s.t., $\text{card}(\mathcal{V}_\mathbf{a}) \leq K$, where $\text{card}(\cdot)$ computes the element number of the set $\mathcal{V}_\mathbf{a}$.

Diffusion Models. DDPMs [10, 25, 53] are latent variable models that generate data $\mathbf{x}_0 \in \mathbb{R}^d$ by reversing a Markovian diffusion process. Starting from Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$, DDPMs iteratively denoise a latent trajectory $\mathbf{x}_T \to \cdots \to \mathbf{x}_0$ to recover samples from the data distribution. Each reverse step is a transition $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$, where μ_θ and Σ_θ are predicted by networks. The forward (noising) process adds Gaussian noise to data over T steps via $q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})$, with a predefined variance schedule $\{\beta_t\}_{t=1}^T$. This process produces tractable posteriors $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0)$, enabling efficient training by minimizing a variational bound on the log-likelihood. To improve training stability, [10] proposes simplifying the loss using closed-form KL divergences between Gaussians, yielding a weighted mean-squared error:

$$\mathcal{L}_{\text{ELBO}}(\mathbf{x}_0) = \sum_{t=1}^{T} \gamma_t \, \mathbb{E}_{\mathbf{x}_t \sim q(\mathbf{x}_t \mid \mathbf{x}_0)} \left[\left\| \mu_\theta(\mathbf{x}_t, t) - \hat{\mu}(\mathbf{x}_t, \mathbf{x}_0) \right\|^2 \right], \tag{1}$$

where $\hat{\mu}(\mathbf{x}_t, \mathbf{x}_0)$ is the mean of the posterior $q(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0)$, and γ_t is a weighting schedule. Although no longer a true ELBO, this objective empirically improves sample quality and stabilizes training [25].

2.2 Diffusion Models for Antibody Sequence Design

Applying a continuous diffusion model to a discrete antibody sequence is challenging. Prior studies [23] converts the multinomial distribution over 20 residue types into a uniform distribution during the forward process. This is inevitably suboptimal because it constrains the noise to lie in a 20-dimensional vector space. Here, we perturb the distribution of residues in a much higher-dimensional vector space, where the noise can be more complex and unconstrained.

Algorithm 1 DiffAntiSeq Sampling Process

Input: diffusion model $\mu_\theta(\cdot)$, classifier $f_\tau(\cdot)$, initial noise level σ_0, gradient scale s, antigen sequence $\mathbf{c}$ and structure $\mathcal{G}_{\text{ag}}$
$\mathbf{x}_T \leftarrow$ sample from $\mathcal{N}(\mathbf{0}, \sigma_0 \mathbf{I})$
for t from T to 1 **do**
  $\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1} \leftarrow \mu_\theta\left(\mathbf{x}_t, t \mid \mathbf{c}, \mathcal{G}_{\text{ag}}\right)$
  $\mathbf{x}_{t-1} \leftarrow$ sample from $\mathcal{N}\left(\boldsymbol{\mu}_{t-1} + s\boldsymbol{\Sigma}_{t-1}\nabla_{\mathbf{x}_t} \log p_\tau(y \mid \mathbf{x}_t), \boldsymbol{\Sigma}_{t-1}\right)$ ▷ Gradients from an extra binding affinity classifier $f_\tau(\cdot)$ is used as guidance
end for
$\mathbf{a} \leftarrow p_\theta(\mathbf{x}_0)$ ▷ Rounding function maps $\mathbf{x}_0$ from latent space $\mathcal{X}$ to discrete token space $\mathcal{A}$
Return $\mathbf{a}$

To begin with, an embedding function $h_\phi(\cdot)$ maps each amino acid to a vector in $\mathbb{R}^d$. Then sequence embeddings are obtained as $h_\phi(\mathbf{a}) = [h_\phi(a_1), ..., h_\phi(a_n)] \in \mathbb{R}^{nd}$ and $h_\phi(\mathbf{c}) = [h_\phi(c_1), ..., h_\phi(c_m)] \in \mathbb{R}^{md}$. It is worth noting that we propose jointly training the diffusion model parameters θ and the residue embeddings ϕ. In preliminary experiments, we explored pretrained residue embeddings based on ESM-2 [21] but found that fixed embeddings were inferior to end-to-end training. After that, a Markov transition is implemented to transfer from discrete amino acids $\mathbf{a}$ to $\mathbf{x}_0$ in the forward process, as $q_\phi\left(\mathbf{x}_0 \mid \mathbf{a}\right) = \mathcal{N}\left(h_\phi(\mathbf{a}), \sigma_0 \mathbf{I}\right)$.

In the reverse process, we add a trainable rounding step, parameterized by $p_\theta\left(\mathbf{a} \mid \mathbf{x}_0\right) = \prod_{i=1}^{n} p_\theta\left(a_i \mid \mathbf{x}_i\right)$, where $p_\theta\left(a_i \mid \mathbf{x}_i\right)$ is a *Softmax* distribution. The training loss is therefore:

$$\mathcal{L}(\mathbf{a}) = \mathbb{E}_{\mathbf{x}_{0:T} \sim q_\phi(\mathbf{x}_{0:T} \mid \mathbf{a})}\left[\mathcal{L}_{\text{ELBO}}\left(\mathbf{x}_0\right) + \left\|h_\phi(\mathbf{a}) - \mu_\theta(\mathbf{x}_1, 1 \mid \mathbf{c}, \mathcal{G}_{\text{ag}})\right\|^2 - \log p_\theta(\mathbf{a} \mid \mathbf{x}_0)\right], \quad (2)$$

where $\mathcal{L}_{\text{ELBO}}\left(\mathbf{x}_0\right)$ is from Equ. 1. As the learned embeddings $h_\phi(\cdot)$ define a mapping from discrete residue types $\mathbf{a}$ to continuous latent space $\mathcal{X}$, the inverse process requires a similar operation to round a predicted $\hat{\mathbf{x}}_0$ back to a discrete antibody sequence. In particular, [20] demonstrate that to directly predict the mean of $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ by $\mu_\theta(\mathbf{x}_t, t \mid \mathbf{c}, \mathcal{G}_{\text{ag}})$ for each denoising step t needs careful tuning, and empirical results show that the model often fails to generate $\mathbf{x}_0$ that commits to a sequence with high probability p_θ. Alternatively, we re-parameterize $\mathcal{L}_{\text{ELBO}}$ to force our model to explicitly emphasize $\mathbf{x}_0$ in every term of the loss objective, and it takes the following form:

$$\mathcal{L}'_{\text{ELBO}}\left(\mathbf{x}_0\right) = \sum_{t=1}^{T} \gamma_t \mathbb{E}_{\mathbf{x}_t \sim q(\mathbf{x}_t \mid \mathbf{x}_0)}\left[\left\|\mu_\theta\left(\mathbf{x}_t, t \mid \mathbf{c}, \mathcal{G}_{\text{ag}}\right) - \mathbf{x}_0\right\|^2\right], \quad (3)$$

where $\mu_\theta\left(\mathbf{x}_t, t \mid \mathbf{c}, \mathcal{G}_{\text{ag}}\right)$ forecasts $\mathbf{x}_0$ immediately. This forces the network to quickly converge to the token embeddings $\mathbf{x}_0$ [20].

2.3 Target-Specific Generation with Desired Properties

The construction of an antibody library must satisfy several essential requirements. For instance, generated antibodies should exhibit improved binding affinity or specificity toward a target antigen, while simultaneously maintaining sufficient sequence diversity for downstream screening. Thus, we formulate antibody design as a controllable generation problem within the diffusion framework.

Predictor-based guidance. We introduce a plug-and-play guidance mechanism that incorporates antigen-specific functional information into the reverse diffusion process. Specifically, we first train a PLM predictor $f_\tau : \mathcal{A} \times \mathcal{C} \to \mathcal{Y}$, which estimates a biological property y (e.g., binding affinity or specificity) for an antibody–antigen pair $(\mathbf{a}, \mathbf{c})$. Because diffusion operates in the continuous latent space, we extend this predictor to latent variables by evaluating it on the representation $\mathbf{x}_t$. We interpret f_τ as parameterizing a conditional likelihood $p_\tau(y \mid \mathbf{x}_t)$. Concretely, we adopt an energy-based interpretation in which $\log p_\tau(y \mid \mathbf{x}_t) \propto f_\tau(\mathbf{x}_t, \mathbf{c})$, so that higher predicted affinity corresponds to higher log-likelihood under the guidance model.

Posterior-guided diffusion. Our goal is to guide the reverse diffusion trajectory $\mathbf{x}_T \to \mathbf{x}_0$ toward sequences with desirable biological properties. From a probabilistic perspective, this corresponds to approximately sampling from the posterior $p(\mathbf{x}_{0:T} \mid y) = \prod_{t=1}^{T} p(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y)$. Thus, controllable

generation can be viewed as modifying each reverse diffusion step to incorporate information about y. Using Bayes' rule, each controlled transition can be written as $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y) \propto p(\mathbf{x}_{t-1} \mid \mathbf{x}_t) \cdot p_\tau(y \mid \mathbf{x}_{t-1}, \mathbf{x}_t)$. To simplify this expression, we make a conditional independence assumption: the biological property y depends only on the underlying antibody representation at step $t-1$ and is conditionally independent of the previous noisy state $\mathbf{x}_t$. Formally, we assume $y \perp \mathbf{x}_t \mid \mathbf{x}_{t-1}$, which yields $p_\tau(y \mid \mathbf{x}_{t-1}, \mathbf{x}_t) = p_\tau(y \mid \mathbf{x}_{t-1})$.

Under this assumption, the gradient of the controlled reverse transition becomes $\nabla_{\mathbf{x}_{t-1}} \log p(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y) = \nabla_{\mathbf{x}_{t-1}} \log p(\mathbf{x}_{t-1} \mid \mathbf{x}_t) + \nabla_{\mathbf{x}_{t-1}} \log p_\tau(y \mid \mathbf{x}_{t-1})$. The first term corresponds to the standard diffusion score, parameterized by the denoising network $\mu_\theta(\cdot)$ together with the embedding function $h_\phi(\cdot)$. The second term is the guidance signal derived from the biological property predictor $f_\tau(\cdot)$ through the likelihood $p_\tau(y \mid \mathbf{x}_{t-1})$. In practice, at each reverse diffusion step, we perform a single gradient update in the latent space along the direction $\nabla_{\mathbf{x}_{t-1}} \log p_\tau(y \mid \mathbf{x}_{t-1})$, thereby steering the denoising trajectory toward antibody sequences with improved predicted properties.

When the target antigen is fixed and unique, we simplify notation by omitting the antigen argument and directly predicting y from the antibody representation, i.e., $f_\tau : \mathcal{A} \to \mathcal{Y}$. In this case, antigen information is implicitly encoded during training. The sampling procedure is depicted in Alg. 1.

Evolutionary regularization. While strong guidance can improve predicted affinity, excessive steering may distort the underlying sequence distribution and lead to biologically implausible antibodies. To mitigate this issue, we introduce an additional evolutionary regularization term that balances fidelity to the pretrained diffusion model and property optimization. Specifically, we consider the combined objective $\lambda \log p(\mathbf{x}_{t-1} \mid \mathbf{x}_t) + \log p_\tau(y \mid \mathbf{x}_{t-1})$, where $\lambda > 0$ is a hyperparameter controlling the trade-off between maintaining evolutionary plausibility (modeled by the unconditional diffusion transition) and optimizing the desired biological property.

Unlike standard classifier-guided diffusion methods, which typically rely solely on the property gradient, our formulation explicitly preserves the contribution of the generative prior. Empirically, we find that retaining the $\lambda \log p(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ term is crucial for generating structurally and biologically reasonable antibody sequences [20]. Overall, the resulting controllable generation procedure can be interpreted as a stochastic decoding process that balances sampling from the learned antibody distribution with targeted optimization toward improved binding properties.

2.4 Reprogramming Protein Language Models

PLMs encode rich evolutionary and structural priors, making them effective backbones for structure-conditioned sequence generation. Recent work, including LM-Design [54], InstructPLM [28], KW-Design [9], and VFN-IF-ESM [24], demonstrates huge gains in structure-aware sequence modeling, with improvements of 10.8% (recovery $50.22\% \rightarrow 55.65\%$), 73.9% (perplexity $10.28 \rightarrow 2.68$), 14.4% (recovery $54.74\% \rightarrow 62.67\%$), and 17.6% (recovery $51.66\% \rightarrow 60.77\%$), respectively, on the CATH 4.2 benchmark.

In DiffAntiSeq, we repurpose PLMs as the sequence decoder $\mu_\theta(\mathbf{x}_t, t \mid \mathbf{c}, \mathcal{G}_{\text{ag}})$, conditioned jointly on the diffusion state and antigen structure. To adapt PLMs efficiently while preserving their pretrained capacity, we employ parameter-efficient fine-tuning (PEFT). Specifically, we combine structural adapters [54] with LoRA [11] using a low-rank configuration ($r = 4$, $\alpha = 8$). Antigen structural context $\mathcal{G}_{\text{ag}}$ is encoded using a GVP-GNN [13]. Although the optimal PEFT strategy for PLMs remains an open question [30], we empirically find that this hybrid adapter-plus-LoRA design consistently outperforms either approach in isolation for structure-conditioned antibody generation.

3 Experiments

3.1 Experimental Setups

Data and Split. We use AlphaSeq [7] as the database, which contains quantitative binding scores of scFv-format antibodies against a SARS-CoV-2 target peptide collected via an AlphaSeq assay [8]. It begins with three seed sequences identified in a phage display campaign using a human naive library. Sets of 29,900 antibodies were designed *in silico* by creating all $k = 1$ mutations and random $k = 2$ and $k = 3$ mutations throughout CDRs. Diversity was introduced in the heavy-chain

Table 1: Performance of different PLMs in predicting the measured binding affinity in AlphaSeq. Results shown as mean ± std over 3 runs.

Model	Fine-tune	**MAE** ↓	**RMSE** ↓	$\mathbf{R}^2$ ↑	**Spearman** ↑	**Pearson** ↑
Transformer	–	1.031 ± 0.043	1.295 ± 0.051	0.170 ± 0.015	0.375 ± 0.018	0.412 ± 0.021
General PLMs						
ProtTrans	✗	0.648 ± 0.028	0.803 ± 0.034	0.382 ± 0.019	0.546 ± 0.022	0.612 ± 0.024
ESM-1	✗	0.540 ± 0.023	0.682 ± 0.029	0.581 ± 0.018	0.701 ± 0.015	0.767 ± 0.013
ESM-2	✗	0.499 ± 0.021	0.622 ± 0.025	0.594 ± 0.016	0.720 ± 0.012	0.774 ± 0.011
MSA-1b	✗	0.508 ± 0.022	0.632 ± 0.027	0.562 ± 0.020	0.712 ± 0.014	0.753 ± 0.013
MINT	✗	0.437 ± 0.019	0.543 ± 0.023	0.627 ± 0.014	0.758 ± 0.011	0.803 ± 0.009
Antibody PLMs						
AbLang	✗	0.456 ± 0.020	0.562 ± 0.024	0.638 ± 0.013	0.749 ± 0.012	0.797 ± 0.010
AntiBERTa	✗	0.421 ± 0.018	0.530 ± 0.022	0.662 ± 0.012	0.768 ± 0.010	0.805 ± 0.009
EATLM	✗	$\underline{0.395 \pm 0.017}$	$\underline{0.497 \pm 0.020}$	$\underline{0.708 \pm 0.011}$	$\underline{0.818 \pm 0.008}$	$\underline{0.850 \pm 0.007}$
EATLM	✓	0.182 ± 0.008	0.235 ± 0.010	0.951 ± 0.004	0.939 ± 0.005	0.975 ± 0.003

Table 2: The affinity statistics of different designed antibody datasets in Kd (the lower the better), including the original AlphaSeq and other DL-generated libraries.

Method	mean	std	min	25%	50%	75%	max
AlphaSeq	3.6810	1.2385	-1.4271	3.0367	3.8399	4.5104	7.3483
dyMEAN [15]	4.0691	0.9336	1.9605	3.4207	4.0640	4.7341	6.5002
DiffAb [23]	3.9879	1.0258	1.5485	3.3315	3.9742	4.6649	6.5322
ProGen2 [26]	0.8658	1.0297	-2.6447	0.1149	0.6385	1.4354	6.0774
BioTransfer [18]	0.6655	**1.0103**	-2.6903	-0.0696	0.4282	1.2141	<u>5.5576</u>
DPLM [35]	0.6201	1.0215	-2.7542	-0.0124	0.3855	1.1687	5.7315
DPLM-2 [36]	<u>0.5743</u>	<u>1.0189</u>	<u>-2.8108</u>	<u>-0.1285</u>	<u>0.3264</u>	<u>1.0952</u>	5.6221
DiffAntiSeq	**0.4650**	1.0300	**-2.9571**	**-0.2706**	**0.2409**	**1.0274**	**5.5098**

CDRs for seed sequence one, in the light-chain CDRs for seed sequence two, and independently in the heavy- and light-chain CDRs of seed sequence three, yielding four sets. Of the 119,600 designs, 104,972 were successfully incorporated into the AlphaSeq library and subsequently measured, yielding 71,384 designs with predicted affinity values for at least one of the triplicate measurements. The data include antibodies with predicted affinity measurements ranging from -1.43 to 7.35. We use Kd in AlphaSeq as the metric for binding affinity. Lower Kd values indicate stronger binding.

Baselines. Antibody library design is an emerging field, and we select several strong baselines for comparison. To be specific, **BioTransfer** [19] is the first DL-driven algorithm for antibody library design. It collected training data by randomly mutating the candidate scFv antibody across the entire CDR and by quantifying high-throughput binding. Then, it performs supervised fine-tuning of pretrained PLMs to predict binding affinities with uncertainty assessment. In silicon, scFv antibody design is conducted via Bayesian optimization over an ML-extrapolated fitness landscape, yielding 248,921 new scFvs. **DiffAb** [23] is among the earliest diffusion-based probabilistic models for protein structures that target specific antigen structures. Here, we discard the structure-recovery component and retain only the sequence diffusion module for our antibody library design, yielding 25k antibodies. Moreover, as SARS-CoV-2 is identified and its structural information is widely accessible, we include another algorithm, **dyMEAN** [15], for comparison. It is an end-to-end, full-atom model for E(3)-equivariant antibody design, given the epitope and an incomplete antibody sequence, and does not require complex structures. We feed the epitope information into dyMEAN and randomly select 100 antibody sequences, with the CDRs masked, as the model input. **ProGen2** [26] is a decoder-only PLM trained on datasets collectively totaling 1B protein sequences from genomic, metagenomic, and immune repertoire databases. A *ProteinGen2-small* with 151M parameters is used. Two diffusion PLM, **DPLM** [35] and **DPLM-2** [36], are also included.

3.2 Results and Analysis

Binding Affinity Prediction. We evaluate the accuracy of PLMs for predicting antibody-antigen binding affinity on AlphaSeq [7], as these predictors serve a dual role: (i) validating the quality of the generated antibody libraries, and (ii) acting as gradient-based classifiers to guide the diffu-

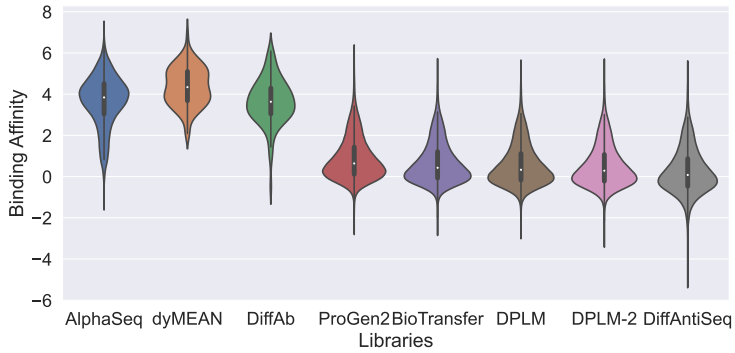

Figure 3: Measured affinity distributions of antibodies in different datasets. A DiffAntiSeq-optimized antibody library outperforms other ML-directed evolution approaches.

sion process. We benchmark both general-purpose PLMs, including ProtTrans [6], ESM-1, ESM-2, MSA-1b [29], and MINT [33], as well as antibody-specialized models such as AbLang [27], AntiBERTa [16], and EATLM [34]. All models are evaluated under linear-probing and fully fine-tuned settings, with results summarized in Tab. 1.

Across all metrics, antibody-specific PLMs consistently outperform general-purpose counterparts, highlighting the importance of immunoglobulin-aware pretraining. Among the models evaluated, EATLM achieves the lowest RMSE (0.4966) and the highest rank and linear correlations. When fully fine-tuned, EATLM further reduces RMSE to 0.2352 and improves Spearman's and Pearson's correlations to 0.93 and 0.97, respectively. These establish EATLM as a highly reliable affinity estimator and justify its use as the guidance predictor.

Antibody Library Design. We next compare DiffAntiSeq with diverse antibody design methods by analyzing the distribution of binding affinities across generated libraries. Tab. 2 reports the mean, standard deviation, and percentile statistics of predicted dissociation constant (K_d), where lower values indicate stronger binding. Baselines include BioTransfer [18], ProGen2 [26], DiffAb [23], dyMEAN [15], and diffusion-based PLMs such as DPLM and DPLM-2 [35, 36].

DiffAntiSeq achieves the lowest mean K_d (0.4650) among all evaluated methods, outperforming BioTransfer, ProGen2, DPLM, and DPLM-2 by a clear margin. In addition, DiffAntiSeq produces a heavier left tail in the affinity distribution, as reflected by more favorable minimum and lower-quartile values. These results indicate that DiffAntiSeq not only improves average binding strength but also increases the frequency of highly potent antibody candidates.

Interestingly, DiffAb [23] yields worse affinity statistics than the original AlphaSeq library. This suggests that unconditional or weakly conditioned diffusion models are insufficient for target-specific antibody optimization. In contrast, DiffAntiSeq explicitly incorporates target-conditioned guidance in a continuous latent space, enabling more effective exploration of functionally relevant sequence regions. Overall, these findings demonstrate the advantage of controllable diffusion over both autoregressive PLMs and structure-only diffusion baselines for antibody library design.

Structure-Based Affinity Evaluation. To validate the functional quality of designed antibodies beyond sequence-level predictors, we perform structure-based affinity estimation on the top-10 candidates from each library. Specifically, we predict antibody structures using ESMFold [21], construct antibody-antigen complexes with HADDOCK [4], and estimate binding free energies using Rosetta [3]. DiffAntiSeq-designed antibodies achieve an average $\Delta\Delta G$ of -31.5 kcal/mol, substantially outperforming AlphaSeq (-25.4 kcal/mol) and BioTransfer [18] (-24.7 kcal/mol). Visual inspection of binding interfaces reveals improved hydrophobic packing and more favorable electrostatic interactions in DiffAntiSeq-generated complexes. These results provide orthogonal structural evidence that DiffAntiSeq produces antibodies with stronger and more stable antigen binding.

Summary. Results show that DiffAntiSeq consistently outperforms existing antibody design methods across predictor-based, distributional, and structure-based evaluations. By combining con-

tinuous latent diffusion with target-specific gradient guidance, DiffAntiSeq effectively balances optimization and exploration, yielding antibody libraries with superior affinity and practical diversity.

3.3 Additional Results

Library Diversity. Beyond binding affinity, we conduct a comprehensive evaluation of the sequence diversity of the designed antibody libraries. We quantify diversity using the edit distance (Levenshtein distance), which measures the minimum number of single-residue edits (insertions, deletions, or substitutions) required to transform one sequence into another. Lower pairwise similarity corresponds to higher library diversity.

Table 3: Quantitative evaluation of sequence diversities among different library design algorithms.

Method	Average Aff.	Normalized ED
AlphaSeq	3.682	0.545
dyMEAN [15]	4.060	0.223
DiffAb [23]	3.984	0.287
ProGen2 [26]	0.861	**0.649**
BioTransfer [19]	<u>0.661</u>	0.452
DPLM [38]	0.812	0.571
DPLM-2 [36]	0.855	<u>0.583</u>
DiffAntiSeq	**0.463**	0.496

Because designed antibodies may exhibit varying CDR lengths, we adopt the normalized edit distance, which accounts for length differences and enables fair comparison across libraries. Specifically, the normalized edit distance is defined as $\text{Normalized Edit Distance} = \frac{\text{Edit Distance}}{\max(\text{Length of Seq 1}, \text{Length of Seq 2})}$, yielding values in the range $[0, 1]$, where 0 indicates identical sequences and 1 indicates maximally dissimilar sequences.

Tab. 3 reveals clear trade-offs between sequence diversity and binding affinity across antibody design methods. ProGen2 achieves the highest diversity (normalized edit distance $= 0.64$), exceeding even the original AlphaSeq dataset (0.54), which can be attributed to its large-scale pretraining on diverse protein corpora spanning genomic, metagenomic, and immune repertoire data. However, this high diversity comes at the expense of reduced binding affinity (mean $K_d = 0.86$).

In contrast, structure-driven methods such as DiffAb and dyMEAN exhibit substantially lower diversity (0.28 and 0.22, respectively) but achieve higher binding affinities (3.98 and 4.06), reflecting their emphasis on structural optimization over broad sequence exploration. DiffAntiSeq strikes a favorable balance between these extremes, maintaining moderate diversity (0.49) while achieving the strongest binding affinity among all methods (mean $K_d = 0.46$). This balance highlights DiffAntiSeq's ability to generate antibody libraries that are both functionally optimized and sufficiently diverse for downstream discovery.

Ablation Studies. We examine the contributions of key components of DiffAntiSeq through two ablation studies, each sampling 25K antibodies. Fig. 4 shows that the removal of either the continuous diffusion or the controllable mechanism induces performance detriment. This is reasonable, as the control technique provides specific guidance to diffusion models for sampling high-affinity scFvs. In addition, continuous diffusion makes the binding affinity distributions of generated scFvs more compact. We expect more advanced conditional diffusion techniques to be developed for this essential design problem.

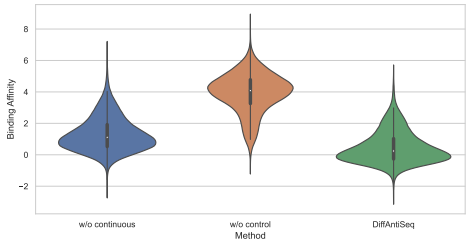

Figure 4: Ablation results, where we remove the continuous diffusion and controllable generation, separately.

4 Conclusion

Despite the importance of therapeutic antibodies, designing early-stage antibody therapeutics remains a time and cost-intensive endeavor. In this paper, we propose a controllable denoising diffusion algorithm, DiffAntiSeq, with two main innovations. Comprehensive experiments demonstrate the ability of DiffAntiSeq to rapidly design large libraries of potently binding antibodies.

LIMITATIONS AND FUTURE WORKS

Despite the progress of DiffAntiSeq in constructing large-scale antibody libraries targeting specific receptors, several limitations impede the extension of our mechanism to real-world applications. First, our model was evaluated using a DL model rather than wet experiments. Those binding affinity data may not be available for the pair of antigen and antibody where the antibody needs to be redesigned. Secondly, we only demonstrated the efficacy of DiffAntiSeq for a single antigen (*e.g.*, SARS-CoV-2), which limits the model's generalizability.

REFERENCES

[1] Hanqun Cao, Marcelo DT Torres, Jingjie Zhang, Zijun Gao, Fang Wu, Chunbin Gu, Jure Leskovec, Yejin Choi, Cesar de la Fuente-Nunez, Guangyong Chen, et al. A deep reinforcement learning platform for antibiotic discovery. *bioRxiv*, 2025.

[2] Hanqun Cao, Hongrui Zhang, Junde Xu, Zhou Zhang, Lingdong Shen, Minghao Sun, Ge Liu, Jinbo Xu, Wu-Jun Li, Jinren Ni, et al. From supervision to exploration: What does protein language model learn during reinforcement learning? *arXiv preprint arXiv:2510.01571*, 2025.

[3] Rhiju Das and David Baker. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.*, 77 (1):363–382, 2008.

[4] Sjoerd J De Vries et al. The haddock web server for data-driven biomolecular docking. *Nature protocols*, 5(5):883–897, 2010.

[5] Arthur Deng, Karsten Householder, Fang Wu, Sebastian Thrun, K Christopher Garcia, and Brian Trippe. Predicting mutational effects on protein binding from folding energy. *arXiv preprint arXiv:2507.05502*, 2025.

[6] Ahmed Elnaggar et al. Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10): 7112–7127, 2021.

[7] Emily Engelhart et al. A dataset comprised of binding interactions for 104,972 antibodies against a sars-cov-2 peptide. *Scientific Data*, 9(1):653, 2022.

[8] Emily Engelhart et al. Massively multiplexed affinity characterization of therapeutic antibodies against sars-cov-2 variants. *Antibody therapeutics*, 5(2):130–137, 2022.

[9] Zhangyang Gao, Cheng Tan, Xingran Chen, Yijie Zhang, Jun Xia, Siyuan Li, and Stan Z Li. Kw-design: Pushing the limit of protein design via knowledge refinement. In *The Twelfth International Conference on Learning Representations*, 2023.

[10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020.

[11] Edward J Hu et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

[12] Wengong Jin et al. Iterative refinement graph neural network for antibody sequence-structure co-design. *arXiv:2110.04624*, 2021.

[13] Bowen Jing et al. Learning from protein structure with geometric vector perceptrons. *arXiv:2009.01411*, 2020.

[14] John Jumper et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596 (7873):583–589, 2021.

[15] Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. *arXiv:2302.00203*, 2023.

[16] Jinwoo Leem et al. Deciphering the language of antibodies using self-supervised learning. *Patterns*, 3(7), 2022.

[17] Guanlue Li, Xufeng Zhao, Fang Wu, and Sören Laue. Joint design of protein surface and backbone using a diffusion bridge model. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[18] Lin Li, Esther Gupta, John Spaeth, Leslie Shing, Rafael Jaimes, Emily Engelhart, Randolph Lopez, Rajmonda S Caceres, Tristan Bepler, and Matthew E Walsh. Machine learning optimization of candidate antibody yields highly diverse sub-nanomolar affinity antibody libraries. *Nature communications*, 14(1):3454, 2023.

[19] Lin Li et al. Machine learning optimization of candidate antibodies yields highly diverse sub-nanomolar affinity antibody libraries. *bioRxiv*, 2022.

[20] Xiang Lisa Li et al. Diffusion-lm improves controllable text generation. *arXiv:2205.14217*, 2022.

[21] Zeming Lin et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv*, 2022:500902, 2022.

[22] Ge Liu et al. Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics*, 36(7):2126–2133, 2020.

[23] Shitong Luo et al. Antigen-specific antibody design and optimization with diffusion-based generative models. *bioRxiv*, 2022.

[24] Weian Mao, Muzhi Zhu, Hao Chen, and Chunhua Shen. Modeling protein structure using geometric vector field networks. *bioRxiv*, pp. 2023–05, 2023.

[25] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, pp. 8162–8171. PMLR, 2021.

[26] Erik Nijkamp et al. Progen2: exploring the boundaries of protein language models. *Cell systems*, 14(11):968–978, 2023.

[27] Tobias H Olsen et al. Ablang: an antibody language model for completing antibody sequences. *Bioinformatics Advances*, 2(1):vbac046, 2022.

[28] Jiezhong Qiu, Junde Xu, Jie Hu, Hanqun Cao, Liya Hou, Zijun Gao, Xinyi Zhou, Anni Li, Xiujuan Li, Bin Cui, et al. Instructplm: Aligning protein language models to follow protein structure instructions. *bioRxiv*, pp. 2024–04, 2024.

[29] Roshan M Rao et al. Msa transformer. In *ICML*, pp. 8844–8856. PMLR, 2021.

[30] Samuel Sledzieski et al. Democratizing protein language models with parameter-efficient fine-tuning. *PNAS*, 121(26):e2405840121, 2024.

[31] Jonathan M Stokes et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4): 688–702, 2020.

[32] Xiangru Tang, Xinwu Ye, Fang Wu, Daniel Shao, Dong Xu, and Mark Gerstein. Bc-design: A biochemistry-aware framework for highly accurate inverse protein folding. In *ICML 2025 Generative AI and Biology (GenBio) Workshop*, 2025.

[33] Varun Ullanat, Bowen Jing, Samuel Sledzieski, and Bonnie Berger. Learning the language of protein-protein interactions. *Nature Communications*, 2026.

[34] Danqing Wang et al. On pre-trained language models for antibody. *bioRxiv*, pp. 2023–01, 2023.

[35] Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Diffusion language models are versatile protein learners. *arXiv preprint arXiv:2402.18567*, 2024.

[36] Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Dplm-2: A multimodal diffusion protein language model. *arXiv preprint arXiv:2410.13782*, 2024.

[37] Zehong Wang, Xiaolong Han, Qi Yang, Xiangru Tang, Fang Wu, Xiaoguang Guo, Weixiang Sun, Tianyi Ma, Pietro Lio, Le Cong, et al. Molecular representations in implicit functional space via hyper-networks. *arXiv preprint arXiv:2601.22327*, 2026.

[38] Zhendong Wang et al. Diffusion-gan: Training gans with diffusion. *arXiv:2206.02262*, 2022.

[39] Fang Wu. A semi-supervised molecular learning framework for activity cliff estimation. *arXiv preprint arXiv:2601.04507*, 2026.

[40] Fang Wu and Stan Z Li. A hierarchical training paradigm for antibody structure-sequence co-design. *arXiv:2311.16126*, 2023.

[41] Fang Wu and Stan Z Li. Dynamics-inspired structure hallucination for protein-protein interaction modeling. *Transactions on Machine Learning Research*, 2025.

[42] Fang Wu, Qiang Zhang, Dragomir Radev, Jiyu Cui, Wen Zhang, Huabin Xing, Ningyu Zhang, and Huajun Chen. 3d-transformer: Molecular representation with transformer in 3d space. 2021.

[43] Fang Wu, Siyuan Li, Lirong Wu, Stan Z Li, Dragomir Radev, and Qiang Zhang. Discovering the representation bottleneck of graph neural networks from multi-order interactions. *arXiv preprint arXiv:2205.07266*, 2022.

[44] Fang Wu, Nicolas Courty, Shuting Jin, and Stan Z Li. Improving molecular representation learning with metric learning-enhanced optimal transport. *Patterns*, 4(4), 2023.

[45] Fang Wu, Huiling Qin, Wenhao Gao, Siyuan Li, Connor W Coley, Stan Z Li, Xianyuan Zhan, and Jinbo Xu. Instructbio: A large-scale semi-supervised learning paradigm for biochemical problems. *arXiv preprint arXiv:2304.03906*, 2023.

[46] Fang Wu, Lirong Wu, Dragomir Radev, Jinbo Xu, and Stan Z Li. Integration of pre-trained protein language models into geometric deep learning networks. *Communications Biology*, 6 (1):876, 2023.

[47] Fang Wu, Shuting Jin, Jianmin Wang, Zerui Xu, Jinbo Xu, Brian Hie, et al. Surfdesign: Effective protein design on molecular surfaces. 2024.

[48] Fang Wu, Tinson Xu, Shuting Jin, Xiangru Tang, Zerui Xu, James Zou, and Brian Hie. D-flow: Multi-modality flow matching for d-peptide design. *arXiv:2411.10618*, 2024.

[49] Fang Wu, Bozhen Hu, and Stan Z Li. Generalized implicit neural representations for dynamic molecular surface modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 877–885, 2025.

[50] Fang Wu, Zhengyuan Zhou, Shuting Jin, Xiangxiang Zeng, Jure Leskovec, and Jinbo Xu. Surface-based molecular design with multi-modal flow matching. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 3192–3203, 2025.

[51] Fang Wu et al. Molformer: Motif-based transformer on 3d heterogeneous molecular graphs. *arXiv:2110.01191*, 2021.

[52] Fang Wu et al. Pre-training of equivariant graph matching networks with conformation flexibility for drug binding. *Advanced Science*, pp. 2203796, 2022.

[53] Fang Wu et al. A score-based geometric model for molecular dynamics simulations. *arXiv:2204.08672*, 2022.

[54] Zaixiang Zheng, Yifan Deng, Dongyu Xue, Yi Zhou, Fei Ye, and Quanquan Gu. Structure-informed language models are protein designers. In *ICML*, pp. 42317–42338. PMLR, 2023.