DiffAntiSeq: A Controllable Diffusion Model for Efficient Antibody Library Design

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

Abstract

Antibodies comprise the most versatile class of binding molecules. Traditional computational methods for antibody design often rely on evolutionary information but are inadequate for certain applications, particularly when multiple sequence alignments are not robust. Machine learning (ML) approaches have demonstrated impressive success in generating antibody sequences, making them a viable option for effectively representing biological data and rapidly exploring the vast in silico antibody spaces. This work proposes DiffAntiSeq, a controllable diffusion-generative model to construct high-quality virtual antibody libraries. DiffAntiSeq conducts the denoising procedure in the latent residue embedding space and is guided by an additional protein language model (PLM) classifier to steer the generation process toward desired properties, such as improved binding affinity and specificity For verification, we integrate target-specific binding affinities with information from millions of antibody sequences in AlphaSeq into our DiffAntiSeq framework and design thousands of single-chain variable fragments (scFvs) that are then empirically measured. Extensive experiments show that the produced antibodies generally have stronger binding strength against the SARS-CoV-2 target peptide, outperforming existing ML-directed evolution approaches. We expect this controllable diffusion method to be broadly applicable and provide value to other protein engineering-related tasks.

1 Introduction

Antibodies have become critical therapeutics due to their high specificity and lower adverse effects compared to small-molecule drugs. Efficient computational methods to explore and prioritize antibody sequences within the vast sequence space are essential, as exhaustive evaluation is impractical (Li et al., 2022a). Constructing smart antibody libraries—diverse yet stable and specific—is central to accelerating therapeutic discovery (Stokes et al., 2020).

Two primary approaches have emerged for library design. The first leverages natural sequence alignments to understand positional constraints and interdependencies among amino acids. However, alignment-based methods struggle with variable-length and hypermutated complementarity-determining regions (CDRs), crucial for antibody specificity.

Alternatively, deep learning (DL) models, capable of capturing complex patterns, have been successfully applied to protein structure prediction and drug discovery (Jumper et al., 2021; Liu et al., 2020; Wu et al., 2021; 2022a; 2023b;a; 2025b; Tang et al., 2024; Deng et al., 2025). Recent DL methods co-design antibody sequences and structures using geometric graph networks (Jin et al., 2021; 2022; Luo et al., 2022; Shi et al., 2022), but these approaches require known antigen or antibody-antigen complex structures, limiting their applicability and iterative refinement capabilities. Sequence-only DL methods avoid structural constraints but typically adopt auto-regressive models, introducing errors from cumulative inference and restrictive directional assumptions.

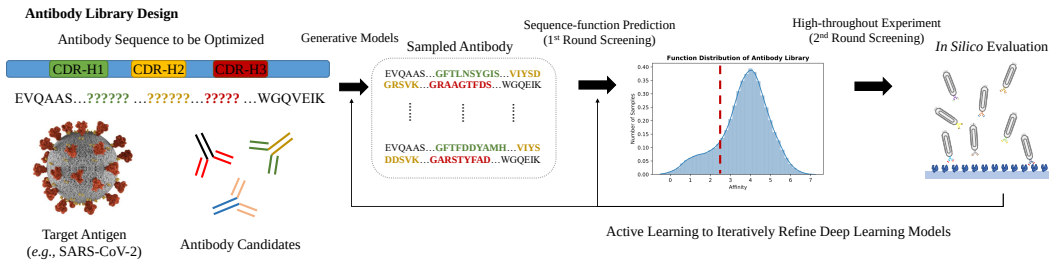

Figure 1: Illustration of the antibody library design task, where an end-to-end diffusion-based algorithm is proposed to design new antibodies. Given a target antigen, the goal is to generate a diverse set of antibody sequences that can bind to the epitope with high affinity. This involves designing the complementarity-determining regions (CDRs) of the antibody while ensuring structural stability and manufacturability. Then, active learning is conducted in the iterative process of refining the antibody library by selecting high-quality candidates for further optimization.

Addressing these limitations, we introduce DiffAntiSeq, an end-to-end denoising diffusion framework combining advanced diffusion techniques with large-scale protein language models (PLMs). DiffAntiSeq transforms initial Gaussian noise vectors into amino acid sequences through progressive denoising steps, generating antibody sequences in a non-autoregressive, full-shot manner. We also propose a gradient-based control algorithm to steer generation toward desired properties such as higher binding strength and specificity, maintaining evolutionary context.

Empirical validation in a virtual library of single-chain variable fragments (scFv) targeting SARS-CoV-2 shows that more than 70% of antibodies generated by DiffAntiSeq outperform the initial candidates. Comparative experiments with state-of-the-art DL-based methods, including BioTransfer (Li et al., 2023) and DiffAb (Luo et al., 2022), confirm that DiffAntiSeq significantly improves antibody quality, underscoring its efficacy in antibody library design.

2 Method

2.1 Task Formulation

We represent the antibody sequence composed of n residues as $\mathbf{a} = [a_1, ..., a_n] \in \mathcal{A}$, where $a_i \in \mathcal{V}$ is an amino acid token. $\mathcal{V}$ is the token vocabulary that consists of twenty amino acid tokens 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 CDR residue b_i in the antibody $\mathbf{a}$. The antigen sequence is consisted 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 via computational tools like AlphaFold (Jumper et al., 2021).

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 full antibody sequence $\mathbf{a}$ is partially known, and the goal shifts to optimizing specific regions, such as CDRs. 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 a given 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

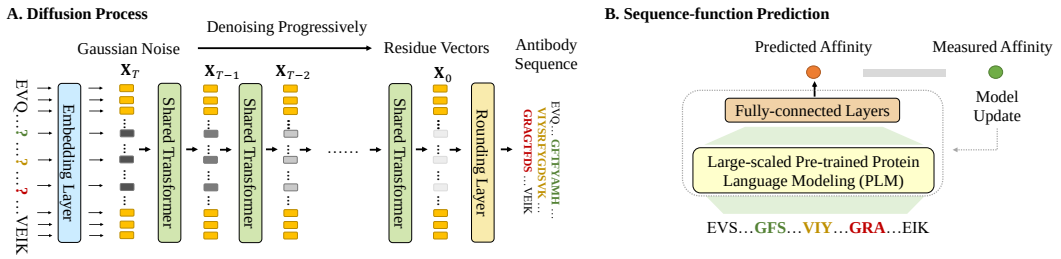

Figure 2: **A.** The diffusion model iteratively denoises a given antibody sequence whose CDRs are filled with Gaussian vectors into residue vectors. It yields an intermediate latent variable of decreasing noise level $\mathbf{x}_T ... \mathbf{x}_0$. A final rounding layer is followed to transfer the residue vectors to discrete antibody sequences. **B.** The generated antibody sequences are forwarded into large-scale pre-trained protein language models (PLM) to obtain the sequence representations, fed into a fully-connected layer to forecast the antibody function further. Quantitative affinity data measured by high-throughput experiments are used to supervise the training of this deep learning model.

K. The goal is to maximize the average binding score across the generated antibodies while promoting diversity. Formally, we solve:

$$\max_{\mathcal{V}_{\mathbf{a}}} \mathbb{E}_{\mathbf{a} \in \mathcal{V}_{\mathbf{a}}}[f(\mathbf{a}, \mathbf{c}, \mathcal{G}_{\text{ag}})], \text{ s.t., } \text{card}(\mathcal{V}_{\mathbf{a}}) \leq K, \tag{1}$$

where $\text{card}(\cdot)$ computes the element number of the set $\mathcal{V}_{\mathbf{a}}$.

2.2 Preliminary of Diffusion Models

Denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020; Nichol & Dhariwal, 2021; Song & Ermon, 2019) 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 modeled as a Gaussian transition $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$, where μ_θ and Σ_θ are predicted by deep networks such as U-Net (Ronneberger et al., 2015) or Transformers (Vaswani et al., 2017).

The forward (noising) process adds Gaussian noise to data over T steps via $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})$, with a predefined variance schedule $\{\beta_t\}_{t=1}^T$. This process produces tractable posteriors $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$, enabling efficient training by minimizing a variational bound on the log-likelihood $\log p_\theta(\mathbf{x}_0)$:

$$\mathbb{E}_{\mathbf{x}_{1:T} \sim q(\cdot|\mathbf{x}_0)} \left[\log \frac{q(\mathbf{x}_T|\mathbf{x}_0)}{p_\theta(\mathbf{x}_T)} + \sum_{t=2}^{T} \log \frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)}{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} - \log p_\theta(\mathbf{x}_0|\mathbf{x}_1) \right]. \tag{2}$$

To improve training stability, Ho et al. (2020) propose simplifying the loss using closed-form KL divergences between Gaussians, yielding a weighted mean-squared error:

$$\mathcal{L}_{\text{ELBO}}(\mathbf{x}_0) = \sum_{t=1}^{T} \gamma_t \, \mathbb{E}_{\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)} \left[\|\mu_\theta(\mathbf{x}_t, t) - \hat{\mu}(\mathbf{x}_t, \mathbf{x}_0)\|^2 \right], \tag{3}$$

where $\hat{\mu}(\mathbf{x}_t, \mathbf{x}_0)$ denotes the mean of the posterior $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$, and γ_t is a weighting schedule. Although no longer a true ELBO, this objective empirically improves sample quality and stabilizes training (Nichol & Dhariwal, 2021).

2.3 Diffusion Models for Antibody Sequence Design

Applying a continuous diffusion model to a discrete antibody sequence is challenging. A recent study (Luo et al., 2022) sets the forward diffusion process in a way that converts

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**
  $\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1} \leftarrow \mu_\theta\left(\mathbf{x}_t, t | \mathbf{c}, \mathcal{G}_{\text{ag}}\right)$
  $\mathbf{x}_{t-1} \leftarrow$ 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 a

the multinomial distribution to the uniform distribution of twenty residue types. This is inevitably suboptimal because it constrains the noise to be a 20-dimensional vector. Here, we borrow the idea from natural language generation (NLG) (Li et al., 2022b; He et al., 2023; Lovelace et al., 2024; Lyu et al., 2023; Liu et al., 2024a;b) and perturb the distribution of residues in a much higher-dimensional vector space, where the noise can be more complex and unconstrained.

To begin with, an embedding function $h_\phi(\cdot)$ is first introduced to map each amino acid to a vector in $\mathbb{R}^d$. Then the embedding of an antibody or antigen sequence is obtained as the two formulas:

$$h_\phi(\mathbf{a}) = [h_\phi(a_1), ..., h_\phi(a_n)] \in \mathbb{R}^{nd}, \qquad h_\phi(\mathbf{c}) = [h_\phi(c_1), ..., h_\phi(c_m)] \in \mathbb{R}^{md}. \tag{4}$$

It is worth noting that we propose to jointly train the diffusion model parameters θ and residue embeddings ϕ. In preliminary experiments, we explored pretrained residue embeddings based on ESM-2 (Lin et al., 2022) but found fixed embeddings inferior to the end-to-end training paradigm. 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 $p_\theta\left(\mathbf{a} \mid \mathbf{x}_0\right) = \prod_{i=1}^{n} p_\theta\left(a_i \mid x_i\right)$, where $p_\theta\left(a_i \mid x_i\right)$ is a *Softmax* distribution. Then, our final training loss is written as follows:

$$\mathcal{L}(\mathbf{a}) = \mathbb{E}_{\mathbf{x}_{0:T} \sim q_\phi(\mathbf{x}_{0:T}|\mathbf{a})}\left[\mathcal{L}_{\text{ELBO}}\left(\mathbf{x}_0\right) + \|h_\phi(\mathbf{a}) - \mu_\theta(\mathbf{x}_1, 1|\mathbf{c}, \mathcal{G}_{\text{ag}})\|^2 - \log p_\theta(\mathbf{a}|\mathbf{x}_0)\right], \tag{5}$$

where $\mathcal{L}_{\text{ELBO}}\left(\mathbf{x}_0\right)$ is derived from Equation 3. Since 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, (Li et al., 2022b) demonstrate that to directly predict the mean of $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ by $\mu_\theta(\mathbf{x}_t, t|\mathbf{c}, \mathcal{G}_{\text{ag}})$ for each denoising step t needs careful tuning, and empirical experiments show that the model usually fails to generate $\mathbf{x}_0$ that commits to a sequence with high probability p_θ. As an alternative choice, we re-parameterize $\mathcal{L}_{\text{ELBO}}$ so that our model is forced to explicitly emphasize $\mathbf{x}_0$ in every term of the loss objective, and it takes the following form:

$$\mathcal{L}'_{\text{ELBO}}\left(\mathbf{x}_0\right) = \sum_{t=1}^{T} \gamma_t \mathbb{E}_{\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)}\left[\left\|\mu_\theta\left(\mathbf{x}_t, t|\mathbf{c}, \mathcal{G}_{\text{ag}}\right) - \mathbf{x}_0\right\|^2\right], \tag{6}$$

where our model $\mu_\theta\left(\mathbf{x}_t, t|\mathbf{c}, \mathcal{G}_{\text{ag}}\right)$ forecasts $\mathbf{x}_0$ immediately. This forces the network to attain $\mathbf{x}_0$ in every step, and (Li et al., 2022b) proved that this objective helps $\mathbf{x}_0$ quickly converge to the token embeddings.

2.4 Target-specific Generation with Desired Antibody Properties

Ideally, the construction of an antibody library ought to satisfy several important requirements. For instance, antibodies need to bind to target molecules with improved binding

strength or specificity. Besides, the library should have rich sequence diversity. To meet these goals, we consider the problem of controllable antibody generation.

To begin with, we describe a plug-and-play procedure that enables the incorporation of antigen information and a generation tendency towards better biological properties. We first leverage a PLM $f_\tau : \mathcal{A} \times \mathcal{C} \to \mathcal{Y}$ to classify the biological property y of any input antibody sequence $\mathbf{a}$, which is bound to the given antigen $\mathbf{c}$. After that, controlling $\mathbf{x}_{0:T}$ is equivalent to decoding from the posterior $p(\mathbf{x}_{0:T} \mid y) = \prod_{t=1}^{T} p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y\right)$. Then this joint inference formula can be decomposed to a sequence of control tasks at each diffusion step, *i.e.*, $p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y\right) \propto p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right) \cdot p\left(y \mid \mathbf{x}_{t-1}, \mathbf{x}_t\right)$. We further simplify $p\left(y \mid \mathbf{x}_{t-1}, \mathbf{x}_t\right) = p(y \mid \mathbf{x}_{t-1})$ via conditional independence assumptions, namely, $y \perp \mathbf{x}_t | \mathbf{x}_{t-1}$. Consequently, the gradient update for the t-th step becomes:

$$\nabla_{\mathbf{x}_{t-1}} \log p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y\right) = \nabla_{\mathbf{x}_{t-1}} \log p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right) + \nabla_{\mathbf{x}_{t-1}} \log p\left(y \mid \mathbf{x}_{t-1}\right), \tag{7}$$

where $\log p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right)$ and $\log p\left(y \mid \mathbf{x}_{t-1}\right)$ are differentiable. The former is parameterized by the diffusion architecture $\mu_\theta(\cdot)$ and $h_\phi(\cdot)$, while the latter is parameterized by the pre-defined classifier $f_\tau(\cdot)$ for binding affinity or specificity. Here, we omit the antigen term and directly forecast y if the target molecule is unique, that is, $f_\tau : \mathcal{A} \to \mathcal{Y}$. Similar to work in the computer vision setting, the classifier $f_\tau(\cdot)$ is trained on the diffusion latent variables, and a gradient update is run on the latent space $\mathbf{x}_{t-1}$ such that it is steered towards fulfilling the control. Notably, these image diffusion studies take one gradient step towards $\nabla_{\mathbf{x}_{t-1}} \log p\left(y \mid \mathbf{x}_{t-1}\right)$ per diffusion step.

To generate biologically reasonable antibodies, we introduce an additional evolutionary regularization as $\lambda \log p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right) + \log p\left(y \mid \mathbf{x}_{t-1}\right)$, where λ is a hyperparameter to trade off homogeneity (the first term) and control (the second term). It is worth mentioning that existing controllable generation methods for diffusion do not include the $\lambda \log p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t\right)$ term in the objective; we found this term to be instrumental for generating biologically reasonable antibody sequences (Li et al., 2022b). The resulting controllable generation process can be viewed as a stochastic decoding method that balances maximizing and sampling $p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, y\right)$. The sampling procedure of our DiffAntiSeq is depicted in Algorithm 1.

2.5 Reprogramming Protein Language Models for Structure-aware Antibody Design

PLMs (Rives et al., 2021; Zheng et al., 2023; Wu et al., 2024b;a) encode rich evolutionary and structural priors, making them powerful engines for structure-conditioned sequence generation. We leverage PLMs as the sequence decoder $\mu_\theta(\mathbf{x}_t, t \mid \mathbf{c}, \mathcal{G}_{\text{ag}})$ in our antibody design framework, enhanced via parameter-efficient fine-tuning (PEFT) to retain modeling strength with minimal overhead.

Our PEFT scheme integrates structural adapters (Zheng et al., 2023) with LoRA (Hu et al., 2022), using a low-rank setup (r=4, α=8). Antigen structural context $\mathcal{G}_{\text{ag}}$ is extracted via a GVP-GNN (Jing et al., 2020). While the optimal PEFT strategy for PLMs remains unsettled (Sledzieski et al., 2024), our hybrid design consistently outperforms individual methods in structure-informed antibody generation.

3 Experiments

Antibody therapies represent a valuable tool to reduce COVID-19 deaths and hospitalizations. To justify the advantages of our DiffAntiSeq, we built a new antibody library against SARS-CoV-2, a strain of coronavirus that causes COVID-19, and then quantitatively investigated the characteristics of this library. Extra experiments are performed to validate the effectiveness of DiffAntiSeq's constituents.

3.1 Dataset

We use AlphaSeq (Engelhart et al., 2022a) as the database, downloaded from `https://github.com/mit-ll/AlphaSeq_Antibody_Dataset`. This dataset contains quantitative binding scores of scFv-format antibodies against a SARS-CoV-2 target peptide collected via an

Table 1: Performance of different PLMs in predicting the measured binding affinity in AlphaSeq.

Model	Fine-tune	**RMSE** ↓	**Spearman** ↑	**Pearson** ↑
Transformer	–	1.2958	0.37	0.41
General PLMs				
ProtTrans	✗	0.8025	0.54	0.61
ESM-1	✗	0.6817	0.70	0.76
ESM-2	✗	0.6221	0.72	0.77
MSA-1b	✗	0.6318	0.71	0.75
Antibody PLMs				
AbLang	✗	0.5622	0.74	0.79
AntiBERTa	✗	0.5297	0.76	0.80
EATLM	✗	0.4966	0.81	0.85
EATLM	✓	**0.2352**	**0.93**	**0.97**

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.

Dataset	mean	std	min	25%	50%	75%	max
AlphaSeq	3.6810	1.2385	-1.4271	3.0367	3.8399	4.5104	7.3483
dyMEAN (Kong et al., 2023)	4.0691	0.9336	1.9605	3.4207	4.0640	4.7341	6.5002
DiffAb (Luo et al., 2022)	3.9879	1.0258	1.5485	3.3315	3.9742	4.6649	6.5322
ProGen2 (Nijkamp, 2023)	0.8658	1.0297	-2.6447	0.1149	0.6385	1.4354	6.0774
BioTransfer (Li et al., 2022a)	0.6655	**1.0103**	-2.6903	-0.0696	0.4282	1.2141	5.5576
DiffAntiSeq	**0.4650**	1.0300	**-2.9571**	**-0.2706**	**0.2409**	**1.0274**	**5.5098**

AlphaSeq assay (Engelhart et al., 2022b). It starts from three seed sequences identified from 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 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, for a total of four sets. Of the 119,600 designs, 104,972 were successfully built into the AlphaSeq library and were subsequently measured with 71,384 designs, resulting in a predicted affinity value for at least one of the triplicate measurements. Data include antibodies with predicted affinity measurements ranging from -1.43 to 7.35. We use Kd as the primary metric for binding affinity, directly provided by the AlphaSeq dataset. Lower Kd values indicate stronger binding. To our knowledge, this dataset is the largest publicly available dataset that contains antibody sequences, antigen sequences, and quantitative measurements. It provides an opportunity to serve as a benchmark to evaluate antibody-specific representation models for DL.

3.2 Results and Analysis

3.2.1 Binding Affinity Prediction

We first train a standalone PLM to predict binding affinity, serving dual purposes: validating the effectiveness of generated libraries and acting as a classifier to propagate gradients and guide the diffusion process. We evaluate various general-purpose PLMs, including ProtTrans (Elnaggar et al., 2021), ESM-1, ESM-2 (Lin et al., 2022), and MSA-1b (Rao et al., 2021), as well as antibody-specific PLMs such as AbLang (Olsen et al., 2022), AntiBERTa (Leem et al., 2022), and EATLM (Wang et al., 2023). These PLMs are tested under linear-probing and fully fine-tuned settings, with results summarized in Tab. 1.

Notably, antibody-specific PLMs generally outperform general-purpose models, with EATLM achieving the best performance. EATLM records the lowest RMSE (0.4966) and the highest Spearman and Pearson correlations. Further fine-tuning EATLM yields even better results, reducing RMSE to 0.2352 and increasing Spearman and Pearson correlations to

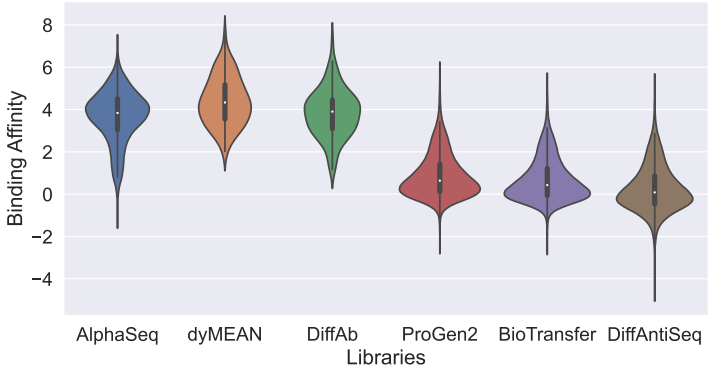

Figure 3: Measured affinity distributions of antibodies in different datasets. A DiffAntiSeq-optimized antibody library outperforms other ML-directed evolution approaches with a high percentage of success.

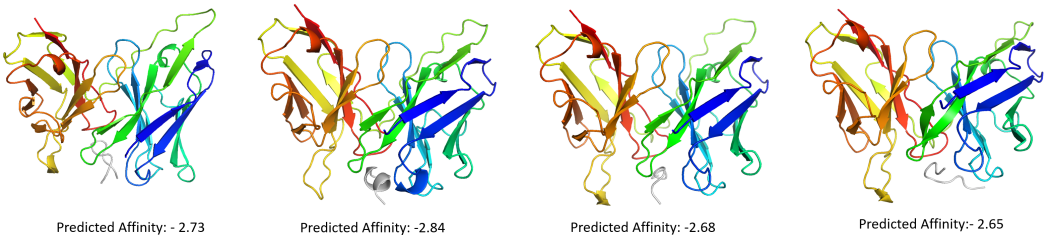

Figure 4: Structural visualization of selected antibody examples in our DiffAntiSeq library against the peptide epitope (the grey segment) from a SARS-CoV-2 Spike protein, which usually elicits strong T cell responses in COVID-19 patients. The complex structures are obtained via Alphafold-3.

0.93 and 0.97, respectively. These results highlight EATLM's capability as a highly accurate adjudicator, effectively validating the efficacy of various antibody design algorithms.

3.2.2 Antibody Library Design

We rigorously compare DiffAntiSeq with BioTransfer, dyMEAN, and DiffAb by reporting the mean, standard deviation, minimum, and maximum binding affinities of AlphaSeq and other generated antibody libraries in Tab. 2. Additionally, we visualize the distribution of measured affinities in Fig. 3 and also draw a couple of complex structures predicted by Alphafold-3 in Fig. 4. Our results show that DiffAntiSeq achieves the highest library success rate. Antibodies generated using DiffAntiSeq-optimized libraries exhibit significantly stronger binding affinities compared to the baselines, with an increased frequency of beneficial mutations. Interestingly, libraries designed using DiffAb perform worse than those designed using AlphaSeq, suggesting that an unconditional diffusion model may not be ideal for target-specific antibody design. This underscores the importance of continuous diffusion models in overcoming the discrete nature of amino acids. These findings demonstrate the strong potential of controllable diffusion models, such as DiffAntiSeq, for antibody library design.

3.3 Evaluation of Antibody Library

To further evaluate the designed libraries, we select the top 10 best antibody sequences from each library and employ a 3-step pipeline to estimate its binding affinity more thoroughly. Specifically, we use ESM-Fold (Lin et al., 2022) to predict antibody structures,

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

Method	Average Aff.	Normalized ED
AlphaSeq	3.68	<u>0.54</u>
dyMEAN (Kong et al., 2023)	4.06	0.22
DiffAb (Luo et al., 2022)	3.98	0.28
ProGen2 (Nijkamp, 2023)	0.86	**0.64**
BioTransfer (Li et al., 2022a)	<u>0.66</u>	0.45
DiffAntiSeq	**0.46**	0.49

HADDOCK (De Vries et al., 2010) to acquire the complexes, and Rosetta (Das & Baker, 2008) to estimate binding affinities against the target antigen. The results indicate that libraries designed by DiffAntiSeq achieved an average $\Delta\Delta G$ of -31.5 kcal/mol, surpassing AlphaSeq (-25.4 kcal/mol) and BioTransfer (-24.7 kcal/mol). Moreover, enhanced binding interfaces with improved hydrophobic and electrostatic interactions were observed in DiffAntiSeq antibodies. This highlights DiffAntiSeq's ability to produce high-affinity antibodies by optimizing binding interfaces through targeted mutations.

3.3.1 Additional Results

In addition to binding affinity measurement, we provide a comprehensive assessment of the sequence diversity of designed antibody libraries. Specifically, we leverage the edit distance (*i.e.*, Levenshtein distance) to calculate the similarity between all pairs of sequences in the library. Lower similarity scores indicate higher diversity. Edit distance measures the minimum number of single-character edits (insertions, deletions, or substitutions) required to change one sequence into another.

Considering that differently designed sequences can have different lengths of CDRs, we utilize the normalized edited distance. That is a scaled version of edit distance and more suitable for comparing sequences of varying lengths, defined as $\text{Normalized Edit Distance} = \frac{\text{Edit Distance}}{\max(\text{ Length of Seq 1, Length of Seq 2})}$. As a consequence, the range of normalized edit distance is between 0 and 1. 0 means the sequences are identical, while 1 means the sequences are completely different (no common characters).

The results in Tab. 3 highlight key trade-offs between sequence diversity and binding affinity across antibody design methods. ProGen2 achieves the highest sequence diversity (normalized edit distance = 0.64), outperforming even the original AlphaSeq dataset (0.54). This high diversity can be attributed to ProGen2's extensive pretraining on a vast corpus of protein sequences, including genomics, metagenomics, and immune repertoire data. However, this comes at the cost of lower binding affinity (0.86). In contrast, structure-based methods like DiffAb and dyMEAN exhibit lower diversity (0.28 and 0.22, respectively) but achieve stronger binding affinities (3.98 and 4.06), as they prioritize structural optimization over sequence exploration. DiffAntiSeq strikes a balance, maintaining moderate diversity (0.49) while achieving the best affinity score (0.46), demonstrating its ability to generate high-quality, diverse antibody libraries. This balance underscores the practical effectiveness of DiffAntiSeq in antibody design.

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 called DiffAntiSeq with two main innovations. Firstly, it performs continuous diffusion on the latent space despite the inherently discrete nature of amino acids. Secondly, we control the diffusion process via gradients to generate antibodies with desired properties. Comprehensive experiments demonstrate the ability of DiffAntiSeq to rapidly design large libraries of potently binding antibodies. Our framework can also be

extended to other domains of protein engineering where large-scale functional mutagenesis screens are applied. We envision our algorithm to solve real-world drug discovery problems.

5 Limitations and Future Works

Despite the progress of DiffAntiSeq in constructing large-scale antibody libraries targeting specific receptors, there are several restrictions in extending our mechanism to real-world applications. Firstly, our model was evaluated using a DL model instead of 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 merely justified the efficacy of DiffAntiSeq on a single antigen (*e.g.*, SARS-CoV-2), which limits the generalizability of this model.

References

Namrata Anand and Tudor Achim. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv:2205.15019*, 2022.

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

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

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.

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.

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, 2022a.

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

Nate Gruver et al. Protein design with guided discrete diffusion. *NeurIPS*, 36, 2024.

Zhengfu He et al. Diffusionbert: Improving generative masked language models with diffusion models. In *Proceedings of the 61st ACL*, pp. 4521–4534, 2023.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv:2207.12598*, 2022.

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

Emiel Hoogeboom et al. Equivariant diffusion for molecule generation in 3d. In *ICML*, pp. 8867–8887. PMLR, 2022.

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

Lei Huang et al. Mdm: Molecular diffusion model for 3d molecule generation. *arXiv:2209.05710*, 2022.

Ilia Igashov et al. Equivariant 3d-conditional diffusion models for molecular linker design. *arXiv:2210.05274*, 2022.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Antibody-antigen docking and design via hierarchical equivariant refinement. *arXiv:2207.06616*, 2022.

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

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *arXiv:2206.01729*, 2022.

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

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

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

Zhifeng Kong et al. Diffwave: A versatile diffusion model for audio synthesis. *arXiv:2009.09761*, 2020.

Gideon D Lapidoth et al. Abdesign: A n algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins: Structure, Function, and Bioinformatics*, 83(8):1385–1406, 2015.

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

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

Lin Li et al. Machine learning optimization of candidate antibody yields highly diverse sub-nanomolar affinity antibody libraries. *Nature Communications*, 14(1):3454, 2023.

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

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

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

Pan Liu et al. Enable fast sampling for seq2seq text diffusion. In *Findings of the ACL: EMNLP 2024*, pp. 8495–8505, 2024a.

Yuhan Liu et al. P3sum: Preserving author's perspective in news summarization with diffusion language models. In *NAACL*, pp. 2154–2173, 2024b.

Justin Lovelace et al. Diffusion guided language modeling. In *Findings of the ACL ACL 2024*, pp. 14936–14952, 2024.

Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *CVPR*, pp. 2837–2845, 2021.

Shitong Luo, Chence Shi, Minkai Xu, and Jian Tang. Predicting molecular conformation via dynamic graph score matching. *NeurIPS*, 34:19784–19795, 2021.

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

Yiwei Lyu et al. Fine-grained text style transfer with diffusion-based language models. In *RepL4NLP 2023*, pp. 65–74, 2023.

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

Erik aand others Nijkamp. Progen2: exploring the boundaries of protein language models. *Cell systems*, 14(11):968–978, 2023.

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

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

Alexander Rives et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*, 118(15):e2016239118, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Arne Schneuing et al. Structure-based drug design with equivariant diffusion models. *arXiv:2210.13695*, 2022.

Chence Shi et al. Protein sequence and structure co-design with equivariant translation. *arXiv:2210.08761*, 2022.

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

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *NeurIPS*, 32, 2019.

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

Xiangru Tang, Howard Dai, Elizabeth Knight, Fang Wu, Yunyang Li, Tianxiao Li, and Mark Gerstein. A survey of generative ai for de novo drug design: new frontiers in molecule and protein generation. *Briefings in Bioinformatics*, 25(4), 2024.

Brian L Trippe et al. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv:2206.04119*, 2022.

Ashish Vaswani et al. Attention is all you need. *NeurIPS*, 30, 2017.

Jordan Venderley. Antibarty diffusion for property guided antibody design. *arXiv:2309.13129*, 2023.

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

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

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

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*, 2022a.

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*, 2023a.

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, 2023b.

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

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*, 2024b.

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, 2025a.

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, 2025b.

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

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

Kevin E Wu et al. Protein structure generation via folding diffusion. *arXiv:2209.15611*, 2022c.

Lemeng Wu et al. Diffusion-based molecule generation with informative prior bridges. *arXiv:2209.00865*, 2022d.

Minkai Xu et al. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv:2203.02923*, 2022.

Ling Yang et al. Diffusion models: A comprehensive survey of methods and applications. *arXiv:2209.00796*, 2022.

Peiyu Yu et al. Latent diffusion energy-based model for interpretable text modeling. *arXiv:2206.05895*, 2022.

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.

A Baseline Methods

Antibody library design is an emerging field, and we select three strong and recent baselines for comparison. To be specific, **BioTransfer** (Li et al., 2022a) is the first DL-driven algorithm for antibody library design. It collected training data via random mutations of the candidate scFv antibody along the entire CDR and high-throughput binding quantification. 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, resulting in 248,921 new scFvs. **DiffAb** (Luo et al., 2022) is one of the earliest diffusion probabilistic models for protein structures targeting specific antigen structures. Here, we discard the structure recovery part and merely keep the sequence diffusion module for our antibody library design, gaining 25k antibodies. Moreover, as SARS-CoV-2 is picked up, for instance, and its structural information is widely accessible, we include another algorithm **dyMEAN** (Kong et al., 2023) for comparison. It is an end-to-end full atom model for E(3)-equivariant antibody design, given the epitope and the incomplete antibody sequence, and does not require complex structures. We feed the epitope information into dyMEAN and randomly select 100 antibody sequences with CDR masked as the model input. ProGen2 (Nijkamp, 2023) 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.

B Experimental Details

DiffAntiSeq is founded on ESM-2, which adopts the Transformer (Vaswani et al., 2017) architecture with 650M parameters. The maximum sequence length is $n = 1024$, the number of diffusion steps is $T = 2000$, and a square-root noise schedule is utilized. That is, $\bar{\alpha}_t = 1 - \sqrt{t/T + \eta}$, where η is a small constant that corresponds to the starting noise level. The embedding dimension is aligned with ESM-2 as $d = 1024$. The classifier takes advantage of the same architecture of the diffusion model, but with a predictive head for attaining binding strength. 25,000 antibodies were sampled by DiffAntiSeq to constitute the library for validation. Following (Wang et al., 2023), we adopted the base Transformer architecture (Vaswani et al., 2017) with 12 layers, 12 heads, and 768 hidden states. The total parameters are 86 M. For the binding affinity prediction task, we conducted a 10-fold validation and reported the average results. For finetuning, we limit the max epochs to 30 and use the Adam optimizer with a max learning rate of 3e-5. We use the mean representation of 12 layers as the sequence representation.

We implement all experiments on 4 A100 GPUs, each with 80G memory. DiffAntiSeq is trained with an AdamW optimizer without weight decay and with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. A ReduceLROnPlateau scheduler is employed to automatically adjust the learning rate with a patience of 10 iterations and a minimum learning rate of $1.e-6$. The regularization weight item λ is set as 1.5. The maximum iterations are 200K, and the validation frequency is 5K iterations. The batch size is set to 64, and the initial learning rate is $1.e-4$ with a dropout rate of 0.1.

BioTransfer was implemented using its official code at `https://github.com/AIforGreatGood/biotransfer`. DiffAb was accessed from its official GitHub at `https://github.com/luost26/diffab`. dyMEAN was examined using its publicly available code at `https://github.com/THUNLP-MT/dyMEAN`. ProGen2 was conducted using its official release at `https://github.com/enijkamp/progen2`. The antibody protocol of HADDOCK was used via the code at `https://github.com/haddocking/HADDOCK-antibody-antigen`.

C Related Work

C.1 Computational Antibody Design

Early methods primarily rely on sampling algorithms applied to hand-crafted and statistical energy functions for antibody optimization, involving iterative modifications to protein sequences and structures (Lapidoth et al., 2015; Wu et al., 2025a). However, energy-based methods suffer from the insufficient expressive power of the statistical energy functions. As a remedy, recent advancements in deep learning have demonstrated substantial enhancements over sampling mechanisms. Specifically, a line of research co-designs the CDR sequences and 3D structures simultaneously, such as Refine-GNN (Jin et al., 2021), HERN (Jin et al., 2022), MEAN, and HTP (Wu & Li, 2023). They all attempt to recover the CDR's sequence and structure while keeping the other parts unchanged. Though this direction seems promising, those methodologies assume the existence of a complex structure, which is usually hard to obtain in real-world circumstances. Moreover, the efficacy of some existing co-design approaches (Jin et al., 2021; 2022) is predominantly limited by the small number of antibody structures.

To overcome these obstacles, another line of research employs language models to generate protein sequences, resulting in increased efficiency. The progress of general PLMs, including ESM, ProGen, and ProTrans, and specific antibody PLMs, such as AntiBerta (Leem et al., 2022), AbLang (Olsen et al., 2022), and EATLM (Wang et al., 2023), provides new prospects for antibody design. It is proven that general PLMs can effectively transfer to antibody tasks and that antibody PLMs improve model performance in antibody paratope predictions. However, how to proficiently unite these PLMs with advanced generative models like diffusion for antibody design remains unexplored.

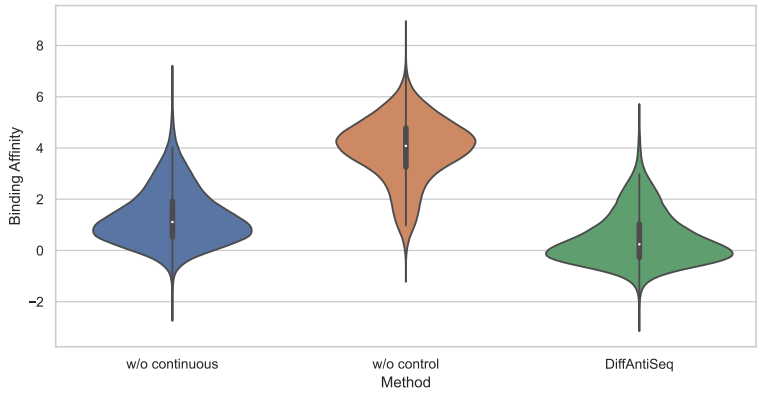

Figure 5: Additional ablation results, where we remove the continuous diffusion and controllable generation, separately.

C.2 Diffusion Models for Proteins

Diffusion models (Yang et al., 2022) have become a new state-of-the-art generative modeling method in the past few years. They are inspired by non-equilibrium thermodynamics and have been invented to learn data distributions by modeling a reverse denoising process. They achieve record-breaking success in various domains including image generation (Wang et al., 2022; Ho & Salimans, 2022), text generation (Li et al., 2022b), interpretable text modeling (Yu et al., 2022), audio synthesis (Kong et al., 2020), and point cloud reconstruction (Luo & Hu, 2021).

Recent efforts employ diffusion models in solving scientific problems, particularly, in drug design with equivariant geometric networks in the 3D space (Hoogeboom et al., 2022; Wu et al., 2022d; Huang et al., 2022; Igashov et al., 2022; Schneuing et al., 2022). They are utilized to generate molecular conformations (Jing et al., 2022; Xu et al., 2022; Luo et al., 2021) or accelerate the simulation of molecular dynamics (MD) (Wu et al., 2022b). In addition to small molecules, they are also applied in the field of larger macromolecules, such as designing new protein backbone structures (Wu et al., 2022c; Anand & Achim, 2022; Shi et al., 2022) or a scaffold structure that supports a desired motif (Trippe et al., 2022). For example, (Luo et al., 2022) presents a diffusion model that targets specific antigen structures with corresponding antibodies. LaMBO (Gruver et al., 2024) proposes guidance over discrete diffusions for antibody design. AntiBARTy (Venderley, 2023) trains a property-conditional diffusion model for guided IgG de novo design. Despite their fruitful progress, none have successfully leveraged diffusion models to generate a smart antibody library to guide the search for potential drugs.

D Ablation Studies

We explore the contributions of different key components of our DiffAntiSeq through two ablation studies, each sampling 25K antibodies. Fig. 5 shows that the removal of either the continuous diffusion or the controllable mechanism induces performance detriment. This is reasonable since the control technique offers specific guidance for diffusion models to sample high-affinity scFvs. In addition, it is observed that the continuous diffusion makes the binding affinity distributions of generated scFvs more condensed. We expect more advanced conditional diffusion techniques to be developed for this essential design problem.