Antibody Design Using a Score-based Diffusion Model Guided by Evolutionary, Physical and Geometric Constraints

Tian Zhu [1 2] **Milong Ren** [1 2] **Haicang Zhang** [# 1 2]

Abstract

Antibodies are central proteins in adaptive immune responses, responsible for protecting against viruses and other pathogens. Rational antibody design has proven effective in the diagnosis and treatment of various diseases like cancers and virus infections. While recent diffusion-based generative models show promise in designing antigen-specific antibodies, the primary challenge lies in the scarcity of labeled antibody-antigen complex data and binding affinity data. We present AbX, a new score-based diffusion generative model guided by evolutionary, physical, and geometric constraints for antibody design. These constraints serve to narrow the search space and provide priors for plausible antibody sequences and structures. Specifically, we leverage a pre-trained protein language model as priors for evolutionary plausible antibodies and introduce additional training objectives for geometric and physical constraints like van der Waals forces. Furthermore, as far as we know, AbX is the first score-based diffusion model with continuous timesteps for antibody design, jointly modeling the discrete sequence space and the $\mathrm{SE}(3)$ structure space. Evaluated on two independent testing sets, we show that AbX outperforms other published methods, achieving higher accuracy in sequence and structure generation and enhanced antibody-antigen binding affinity. Ablation studies highlight the clear contributions of the introduced constraints to antibody design.

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. [2]University of Chinese Academy of Sciences, Beijing, China. Correspondence to: Haicang Zhang <zhanghaicang@ict.ac.cn>.

Proceedings of the 41^{st} International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

1. Introduction

The rational design of antigen-specific antibodies plays a pivotal role in the prevention, diagnosis, and treatment of various diseases (Carter, 2006). Antibodies perform their functions by binding to antigens, triggering a cascade of immune responses. While the framework regions of antibodies exhibit high conservation, their complementarity-determining regions (CDRs) are variable both in sequence and structure, mainly determining the binding affinity and specificity to antigens (Al-Lazikani et al., 1997; North et al., 2011). Therefore, the primary objective of rational antibody design is to optimize the CDRs for effective binding to the targeted antigen.

Experimentally exploring the huge structure and sequence space of CDRs and validating their binding affinity to antigens is inefficient in wet lab settings. Therefore, computational methods have been proposed to optimize CDRs, with antibody-antigen binding affinity as the primary objective. These include traditional classical methods such as RAbD (Pantazes & Maranas, 2010; Lapidoth et al., 2015; Adolf-Bryfogle et al., 2018), as well as more recent deep learning-based methods.

Diffusion models, commonly applied in generating image data, have recently been introduced to antibody design. For example, DiffAb proposes a denoising diffusion probabilistic model on both discrete sequence space and $\mathrm{SE}(3)$ structure space of antibody, targeting specific antigens (Luo et al., 2022).

The scarcity of available antibody-antigen complex data poses a significant challenge for these diffusion-based generative models. In the field of computer vision, models often require significantly larger datasets for effective training, and diffusion models trained on limited datasets tend to exhibit restricted expressiveness, resulting in the generation of less diverse and potentially biased images (Hur et al., 2024). For the case of antibody design, the widely-used SAbDab database (Dunbar et al., 2014) of antibody-antigen complexes comprises only thousands of non-redundant samples, leading to a high risk of overfitting.

We develop AbX, a new score-based diffusion model that is guided by evolutionary, physical, and geometric constraints.

These constraints serve as sequence and structure priors, narrowing the exploration to a more plausible space and mitigating the risk of overfitting. First, we incorporate the language-language model ESM-2 (Lin et al., 2023)as a prior for generating evolutionary plausible antibody sequences. Second, we integrate geometric and physical constraints to generate rational antibody structures and improve the antibody-antigen binding affinity through additional training objectives, including Frame Aligned Point Error (FAPE) and van der Waals loss. Moreover, AbX is the first model to implement a continuous-time score-based diffusion approach in both discrete sequence space and $\mathrm{SE}(3)$ structure space for antibody design.

Our contributions are summarized as follows:

- AbX integrates evolutionary, physical, and geometric constraints in a unified framework for antibody design.
- Unlike previous diffusion-based methods using discrete-time schemes, AbX introduces the first continuous-time score-based diffusion model for joint modeling of discrete sequence space and $\mathrm{SE}(3)$ structure space.
- Evaluation on independent testing sets demonstrates that AbX achieves state-of-the-art performance in antibody design and optimization tasks, highlighting the clear contribution of the introduced constraints.

2. Related Work

2.1. Computational Antibody Design

The primary objective of antibody design is to optimize antibody sequence and structure, particularly the CDRs, to effectively bind to specific target antigens. Traditional classical computational approaches like RAbD (Adolf-Bryfogle et al., 2018) and AbDesign (Lapidoth et al., 2015) explore a library of clustered CDRs and optimize binding energy using Monte Carlo-based methods, which can be time-intensive.

Recent deep learning-based methods have been proposed to improve antibody design. These methods can be broadly categorized into discriminative models and generative models. Discriminative models typically utilize graph neural networks to extract features from antigen structure and to predict the most likely antibody structure and sequence (Jin et al., 2022; Kong et al., 2023a; Gao et al., 2023; Kong et al., 2023b). Generative models (Rezende & Mohamed, 2015; Song et al., 2020; Goodfellow et al., 2020) explicitly describe the generative process of antibody structure and sequence and can sample a group of diverse antibodies (Eguchi et al., 2022). Among these, diffusion-based models stand out for their high capacity to model the complex distributions of structure and sequence (Luo et al., 2022; Peng et al., 2023). All these methods are limited to the scarcity of training samples.

Introducing additional constraints into a generative model serves as a strategy to address the limitation of scarce training datasets. For instance, additional priors have been introduced to improve video generation (Ge et al., 2023; Jin et al., 2023). In the case of antibody design, AbDiffuser and DiffProperty are the most related to our approach. AbX shows significant differences with these methods. AbDiffuser estimates a position-specific profile as sequence priors from the limited available structure data. In contrast, our approach utilizes a large pre-trained language model as sequence priors, which has been proven as an effective indicator of evolutionary plausibility in wet-lab experiments (Hie et al., 2023); While DiffProperty employs existing methods to generate candidate antibodies and use property predictors for further screening (Villegas-Morcillo et al., 2023), AbX directly integrates constraints into the model.

2.2. Diffusion-based Generative Model

Diffusion-based generative models have recently shown great potential in various fields like image, text, and video (Ho et al., 2020; Austin et al., 2021; Mei & Patel, 2023). It has also been applied in protein-related tasks such as protein design (Watson et al., 2023; Ingraham et al., 2023; Alamdari et al., 2023; Yim et al., 2023), protein structure prediction (Jing et al., 2023), and protein side-chain packing (Liu et al., 2023; Zhang et al., 2023).

DDPMs and score-based models are mainly two kinds of diffusion-based models. Score-based models have demonstrated superior performance over DDPMs in various domains (Campbell et al., 2022; Vahdat et al., 2021). While previous diffusion-based models for antibody design such as DiffAb (Luo et al., 2022) and AbDiffuser (Martinkus et al., 2023) leverage DDPMs to co-generate CDRs sequences and structures, our model utilizes a score-based diffusion model.

3. Methods

This section is structured as follows: In section 3.1, we introduce the notations used throughout the paper and formally define the problem. Section 3.2 provides an overview of our model. Section 3.3 describes the diffusion process applied to modeling antibodies. Section 3.4 details our score network for the diffusion process and our conditional encoder within the model. Section 3.5 outlines the training objectives of our model. Section 3.6 explains our strategy for sampling protein sequences and structures.

3.1. Preliminaries and Notations

An antibody-antigen complex comprises four chains: two identical heavy chains and two identical light chains, as illustrated in Figure 1 (Dunbar et al., 2014). These chains

are divided into variable (V) and constant (C) domains. The variable domains, VH and VL, each contain three CDRs each: H1, H2, H3 in VH, and L1, L2, L3 in VL (Al-Lazikani et al., 1997; North et al., 2011).

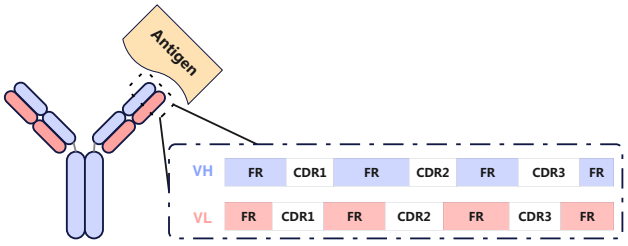


Figure 1. Illustration of antibody-antigen complex structure and its CDRs.

Each antibody residue, denoted as $\{a_1, a_2, \ldots, a_n\}$, is characterized by its amino acid type $a_i \in \{\text{ACDEFGHIKLMNPQRSTVWY}\}$, the rotation $\mathbf{R}_i \in \text{SO}(3)$, the 3D coordinate $\mathbf{x}_i \in \mathbb{R}^3$, we can combine the rotation and 3D coordinate together denoted as frame $T_i = (\mathbf{R}_i, \mathbf{x}_i) \in \text{SE}(3)$. Then the antibody can be denoted as $\mathcal{P}_{\text{Ab}} = \{(a_i, T_i) \mid i = 1, \cdots n\}$. Antigen also includes one or more chains, and the set of its residues is $\{n+1, n+2, \ldots, m\}$. So we can denote the antigen as $\mathcal{P}_{\text{Ag}} = \{(a_i, T_i) \mid i = n+1, \ldots, m\}$. In this work, our goal is to design CDRs given the antigen and antibody framework regions. We assume the CDRs to be generated are indexed by the set $\mathcal{C} \subset \{1, 2, \ldots, n\}$, where $|\mathcal{C}| = N$. Therefore, the CDRs can be represented as $\mathcal{P}_C = \{(a_i, T_i) \mid i \in \mathcal{C}\}$, with $\mathbf{A} = \{a_i \mid i \in \mathcal{C}\}$ denoting the sequence and $\mathbf{T} = \{T_i \mid i \in \mathcal{C}\}$ representing the structure. Formally, our goal is to model the distribution of $\mathcal{P}_C$ given $\mathcal{P}_{\text{FC}} = \{(a_i, T_i) \mid i \in \{1, \ldots, m\} \backslash \mathcal{C}\}$.

3.2. Overview of AbX

Figure 2 provides a comprehensive overview of the AbX model. The model integrates physical and geometric constraints as part of the training objectives, and the evolutionary constraint is informed by the large language model ESM-2 within the model architectures.

During inference, the process begins by sampling the structure and sequence of the CDRs from prior distributions. At each step, AbX receives the antigen and antibody from the previous timestep as input. The antigen and framework regions of the antibody are encoded by the conditional encoder, and the large language model, ESM-2, within the score network, is then utilized to generate antibody sequence embeddings. Subsequently, the conditional embedding and the sequence embedding are processed by the score network. Finally, the score network generates the initial structure $\mathbf{T}_\theta^0$ and sequence distribution $p_{0|t}^\theta(a_0|a, \mathcal{P}_{\text{FC}})$ for the CDRs. Based on the initial structure, we compute the scores for both sequence and structure. Utilizing these scores, we then perform a reverse step to derive the sequence and structure at the subsequent timestep. In summary, through the implementation of a reverse diffusion process, AbX refines these generated CDRs to yield the finalized structure and sequence of the CDRs.

3.3. Score-based Diffusion Processes for Structure and Sequence

Score-based diffusion models utilize stochastic differential equations (SDE) to characterize the diffusion process. In this section, we briefly outline the forward and reverse processes for structures and sequences. Let $\mathcal{P}_C^t = \{(a_j^t, T_j^t) \mid j \in \mathcal{C}\}, t \in \mathcal{U}(0,1)$ represent the intermediate state of the CDRs at time t. The forward diffusion process progresses from $t = 0$ to $t = 1$, and the generative diffusion process proceeds in reverse. This diffusion process applies to both the sequence $\{a_j^t \mid j \in \mathcal{C}\}$ and the structure $\{T_j^t \mid j \in \mathcal{C}\}$.

Diffusion process for sequence Given the discrete nature of sequences, we follow τLDR to establish a Continuous Time Markov Chain (CTMC) for sequence space, analogous to a score-based diffusion process (Campbell et al., 2022). Contrasting with τLDR, our method alters the original diffusion model into a conditional version, incorporating conditional embeddings of the target antigens and the framework regions of antibodies. We define a prior distribution $p_{\text{ref}}^{\text{Seq}}$ and a transition rate matrix $\mathbf{S}^t \in \mathbb{R}^{20\times 20}$. For a current state $\tilde{a}$, the matrix entry $\mathbf{S}^t(\tilde{a}, a)$ indicates the instantaneous transition rate to state a. Detailed information about the Continuous Time Markov Chain is available in the Appendix A.1.

The CDRs sequence is denoted as $(\mathbf{A}^t)_{t\sim\mathcal{U}(0,1)} = (\{a_i^t | i \in \mathcal{C}\})_{t\sim\mathcal{U}(0,1)}$, initiated at $p_0(\mathbf{A}|\mathcal{P}_{\text{FC}}) = p_{\text{data}}(\mathbf{A}|\mathcal{P}_{\text{FC}})$ at $t = 0$ and transitioning to $p_1(\mathbf{A}|\mathcal{P}_{\text{FC}})$ at $t = 1$, approximating the prior distribution $p_{\text{ref}}^{\text{Seq}}(\mathbf{A}) = (\mathcal{U}(\{0, \ldots, 19\}))^{\otimes N}$, where $\mathcal{U}$ denotes the uniform distribution. The infinitesimal transition probabilities for the forward and reverse processes at each residue a are defined as:

$$p_{t|t-\Delta t}(a|\tilde{a}) = \delta_{a,\tilde{a}} + \mathbf{S}^t(\tilde{a}, a)\Delta t + \text{o}(\Delta t), \quad (1)$$
$$p_{t|t+\Delta t}(\tilde{a}|a, \mathcal{P}_{\text{FC}}) = \delta_{\tilde{a},a} + \hat{\mathbf{S}}^t(a, \tilde{a}, \mathcal{P}_{\text{FC}})\Delta t + \text{o}(\Delta t).$$

Here, $\text{o}(\Delta t)$ means terms that tend to zero at a faster rate than Δt, and $\hat{\mathbf{S}}^t \in \mathbb{R}^{20\times 20}$ represents a reverse transition rate matrix which is calculated from $\mathbf{S}_t$ as:

$$\hat{\mathbf{S}}^t(a, \tilde{a}, \mathcal{P}_{\text{FC}}) = \mathbf{S}^t(\tilde{a}, a) \sum_{a_0} \frac{p_{t|0}(\tilde{a}|a_0)}{p_{t|0}(a|a_0)} p_{0|t}(a_0|a, \mathcal{P}_{\text{FC}}),\ a \neq \tilde{a}, \quad (2)$$

where $p_{t|0}$ are the marginals of the forward process, and $p_{0|t}(a_0|a, \mathcal{P}_{\text{FC}}) = p_{t|0}(a|a_0)\frac{p_{\text{data}}(a_0|\mathcal{P}_{\text{FC}})}{p_t(a|\mathcal{P}_{\text{FC}})}$ with $p_t(a|\mathcal{P}_{\text{FC}})$

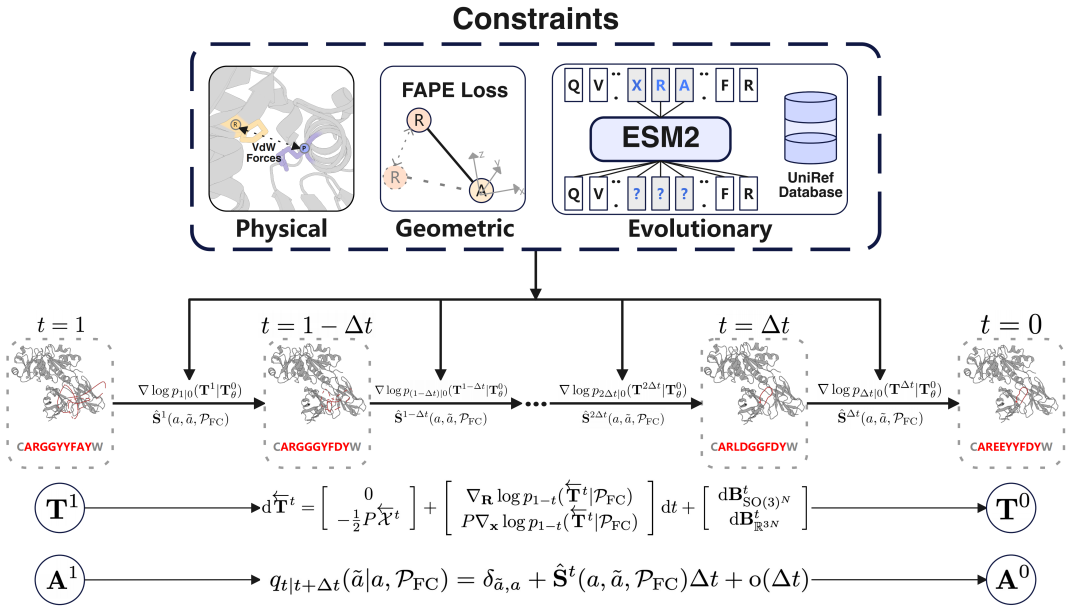


Figure 2. Overview of the AbX.AbX leverages a Continuous Time Markov Chain (CTMC) and a score-based diffusion model to generate CDRs sequences and structures. It incorporates physical and geometric constraints during training and applies evolutionary constraints within its model architecture. The grey part of the protein structure and sequence represents the target antigens and the antibody framework, while the red part denotes the CDRs in the antibody.

being the forward process marginal at time t. Considering the intractability of directly computing $p_{0|t}(a_0|a)$, we approximate it as $p^{\theta}_{0|t}(a_0|a,\mathcal{P}_{\mathrm{FC}})$ with the score network, and we can get approximation reverse time transition rate matrix $\hat{\mathbf{S}}^t_\theta$. More details about the choice of transition rate matrix can be found in the Appendix A.1.

Diffusion process for structure We employ a score-based diffusion model in the $\mathrm{SE}(3)^N$ space as proposed in FrameDiff to describe the generative process for CDRs structures (Yim et al., 2023). Our primary extension from FrameDiff for structure modeling is that our approach is a conditional diffusion model, taking into account the target antigens and the framework regions of antibodies. The intermediate states of CDRs structures can be denoted as $(\mathbf{T}^t)_{t\in\mathcal{U}(0,1)}=(\{\mathbf{R}^t,\mathbf{X}^t\})_{t\in\mathcal{U}(0,1)}$, where $\mathbf{R}^t=\{\mathbf{R}^t_i\}_{i\in\mathcal{C}},\mathbf{X}^t=\{\mathbf{x}^t_i\}_{i\in\mathcal{C}}$. Let $(\overleftarrow{\mathbf{T}}^t)=(\mathbf{T}^{1-t})_{t\in\mathcal{U}(0,1)}$, with $\overleftarrow{\mathbf{T}}^t=(\overleftarrow{\mathbf{R}}^t,\overleftarrow{\mathbf{X}}^t)$.

We select prior distribution as $p^{\mathrm{SE}(3)^N}_{\mathrm{ref}}(\mathbf{T})=\mathrm{P}_{\#}(\mathcal{N}(0,\mathrm{Id})^{\otimes N})\otimes\mathcal{U}(\mathrm{SO}(3))^{\otimes N}$ and the initial distribution $p_0(\mathbf{T}|\mathcal{P}_{\mathrm{FC}})=p_{\mathrm{data}}(\mathbf{T}|\mathcal{P}_{\mathrm{FC}})$. The forward SDE and reverse SDE on $\mathrm{SE}(3)^N$ can be defined as follows:

$$\mathrm{d}\mathbf{T}^t=\begin{bmatrix}0\\ P\frac{1}{2}\mathbf{X}^t\end{bmatrix}\mathrm{d}t+\begin{bmatrix}\mathrm{d}\mathbf{B}^t_{\mathrm{SO}(3)^N}\\ P\mathrm{d}\mathbf{B}^t_{\mathbb{R}^{3N}}\end{bmatrix}, \tag{3}$$

$$\mathrm{d}\overleftarrow{\mathbf{T}}^t=\begin{bmatrix}0\\ -\frac{1}{2}P\overleftarrow{\mathbf{X}}^t\end{bmatrix}+\begin{bmatrix}\nabla_{\mathbf{R}}\log p_{1-t}(\overleftarrow{\mathbf{T}}^t|\mathcal{P}_{\mathrm{FC}})\\ P\nabla_{\mathbf{x}}\log p_{1-t}(\overleftarrow{\mathbf{T}}^t|\mathcal{P}_{\mathrm{FC}})\end{bmatrix}\mathrm{d}t+\begin{bmatrix}\mathrm{d}\mathbf{B}^t_{\mathrm{SO}(3)^N}\\ \mathrm{d}\mathbf{B}^t_{\mathbb{R}^{3N}}\end{bmatrix}. \tag{4}$$

Here, $\mathbf{B}^t_{\mathrm{SO}(3)^N}$ and $\mathbf{B}^t_{\mathbb{R}^{3N}}$ represent the Brownian motion on $\mathrm{SO}(3)^N$ and $\mathbb{R}^{3N}$, respectively, and $P\in\mathbb{R}^{3N\times 3N}$ can be defined as the projection matrix that removes the center of mass $\frac{1}{N}\sum_{i\in\mathcal{C}}\mathbf{x}_i$. The score functions $\nabla_{\mathbf{R}}\log p_{1-t}(\overleftarrow{\mathbf{T}}^t|\mathcal{P}_{\mathrm{FC}})$ and $\nabla_{\mathbf{x}}\log p_{1-t}(\overleftarrow{\mathbf{T}}^t|\mathcal{P}_{\mathrm{FC}})$ can be approximated by $s^{\mathbf{R}}_\theta$ and $s^{\mathbf{x}}_\theta$ using $\mathbf{T}^t$ and $\mathbf{T}^0_\theta$ as per the methods described in FrameDiff (Yim et al., 2023). The computation of these scores is detailed in the Appendix A.2. The generated initial structure $\mathbf{T}^0_\theta=(\mathbf{R}^0_\theta,\mathbf{X}^0_\theta)$ can be estimated with the score network, with further details provided in Section 3.4.

3.4. Model Architectures

Conditional encoder We leverage a Multiple Layer Perceptrons (MLP), denoted as $G_\theta(\mathcal{P}_{\mathrm{FC}})$, to learn single and pair representations from the given antibody framework

region $\mathcal{P}_{\mathrm{FC}}$ and the antigen $\mathcal{P}_{\mathrm{Ag}}$

For the single representation $\mathbf{s}$, the inputs include amino acid types, the torsion angles ϕ and ψ, and 3D coordinates of all atoms. For the pair representations $\mathbf{z}$, the inputs include pair amino acid identities, the inter-residue distances of all-atom pairs, and the relative positional encoding.

Score network The score network employs triangular multiplicative update and Invariant Point Attention (IPA) layers (Jumper et al., 2021) to generate the initial structure $\mathbf{T}_\theta^0$ and sequence distribution $p_{0|t}^\theta(a_0|a, \mathcal{P}_{\mathrm{FC}})$.

The input to our score network $F_\theta(\mathbf{T}^t, \mathbf{A}^t, \mathbf{s}, \mathbf{z})$ includes the noised sequence and structure, supplemented by single embeddings $\mathbf{s}$ and pair embeddings $\mathbf{z}$ from the conditional encoder. Recognizing the superiority of general protein language models in enhancing the evolutionary plausibility of antibody designs, as established by Hie et al. (2023), we have integrated ESM-2 (3B) as our evolutionary constraint in the score network. ESM-2 is utilized for encoding the noised sequence $\mathbf{A}^t \cup (\mathbf{A}_{\mathrm{Ab}} \backslash \mathbf{A})$, feeding into our score network.

Moreover, our score network incorporates a recycling mechanism, acting as self-conditioning within the diffusion model framework (Watson et al., 2023). During this recycling process, the initially generated sequence is encoded using ESM-2 and combined with the single and pair embeddings from IPA, augmenting the conditional embeddings. More details about model architectures can be found in Appendix A.3

3.5. Training Objectives

In our work, the process of learning the initial structure $\mathbf{T}_\theta^0$ and the sequence distribution $p_{0|t}^\theta(a_0|a)$ involves minimizing not only the denoising score matching losses but also incorporating auxiliary losses designed to the specificity of antibody design tasks. These auxiliary objectives include physical and geometric constraints, improving target antigen recognition and binding.

Physical constraint In our model, we embed physical constraint to guide the generation of antibodies with high binding affinity to target antigens. This includes a structural violation loss $\mathcal{L}_{\mathrm{violations}}$ (Jumper et al., 2021) to prevent violations in covalent peptide bond angles and lengths among neighboring residues (detailed in Appendix A.4), and a van der Waals loss $\mathcal{L}_{\mathrm{VdW}}$ to approximate the van der Waals forces within neighboring non-bonded backbone atoms. The van der Waals loss is defined as:

$$\mathcal{L}_{\mathrm{VdW}} = \sum_{i=1}^{N_{\mathrm{nbpairs}}} \max(d_{\mathrm{lit}}^i - \tau - d_{\mathrm{pred}}^i, 0), \tag{5}$$

where d_{pred}^i is the distance between two non-bonded backbone atoms in the generated initial structure and d_{lit}^i is the literature-supported plausible distance based on van der Waals radii and force. N_{nbpairs} represents the number of non-bonded backbone atom pairs between CDRs and other regions including antigens. The tolerance τ is set to 1.5Å. The physical constraint is expressed as:

$$\mathcal{L}_{\mathrm{Physical}} = 0.03\mathcal{L}_{\mathrm{VdW}} + 0.03\mathcal{L}_{\mathrm{violation}}. \tag{6}$$

Geometric constraint The geometric constraint is specifically formulated to accurately depict the rigidity and flexibility inherent in antibody structures. For the CDRs structures $\mathbf{T} = (\mathbf{R}, \mathbf{X})$, we incorporate the FAPE loss $\mathcal{L}_{\mathrm{FAPE}}$, distogram loss $\mathcal{L}_{\mathrm{distogram}}$ and lDDT-Cα loss $\mathcal{L}_{\mathrm{lDDT}}$ (Jumper et al., 2021), aiming to generate more rational structure. The geometric constraint is thus defined as:

$$\mathcal{L}_{\mathrm{Geometric}} = \mathcal{L}_{\mathrm{FAPE}} + 0.5\mathcal{L}_{\mathrm{distogram}} + 0.1\mathcal{L}_{\mathrm{lDDT}}. \tag{7}$$

Denoising score matching losses For the CDRs structure, we minimize the DSM loss $\mathcal{L}_{\mathrm{DSM}}$ as per previous work (Yim et al., 2023). In the sequence space, the continuous time negative Evidence Lower Bound (ELBO) $\mathcal{L}_{\mathrm{ELBO}}$ and the cross-entropy loss $\mathcal{L}_{\mathrm{CE}}$ are minimized. The continuous time negative ELBO for the reverse time CTMC is defined as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{ELBO}} =& \mathbb{E}_{t\sim\mathcal{U}(0,1)p_t(a)r_t(\tilde{a}|a)}[\sum_{a'\neq a} \hat{\mathbf{S}}_\theta^t(a,a') \\ &- \mathcal{Z}^t(a)\log(\hat{\mathbf{S}}_\theta^t(\tilde{a},a))] + C,\end{aligned} \tag{8}$$

where C is a constant independent of θ, and

$$\mathcal{Z}^t(a) = \sum_{a'\neq a} \mathbf{S}_t(a,a'),\ r_t(\tilde{a}|a) = (1-\delta_{\tilde{a},a})\frac{\mathbf{S}_t(a,\tilde{a})}{\mathcal{Z}^t(a)}.$$

The denoising score matching losses are denoted as:

$$\mathcal{L}_{\mathrm{Diff}} = \mathcal{L}_{\mathrm{DSM}} + 0.2\mathcal{L}_{\mathrm{ELBO}} + 0.2\mathcal{L}_{\mathrm{CE}}. \tag{9}$$

The overall training objective We aggregate all the aforementioned losses to form the final training objective function. The geometric and physical constraints are applied only when t is sampled near 0, where fine-grained characteristics emerge. The complete training loss is expressed as:

$$\mathcal{L}_{\mathrm{All}} = \mathcal{L}_{\mathrm{Diff}} + \mathbb{I}_{t<\frac{1}{4}}(\mathcal{L}_{\mathrm{Geometric}} + \mathcal{L}_{\mathrm{Physical}}), \tag{10}$$

where $\mathbb{I}_{t<\frac{1}{4}}$ is an indicator variable that equals 1 only when $t < \frac{1}{4}$. More details about model training can be found in Appendix C.

3.6. Sampling Algorithms

Our algorithm integrates sequence sampling and structure sampling for antibody design. We employ the tau-leaping method (Gillespie, 2001) for sequence generation and geodesic random walk approach (Jørgensen, 1975) for structure sampling, as detailed in Algorithm 1.

Sampling algorithm for sequence We utilize a tau-leaping approach, assuming constant $\hat{\mathbf{S}}_\theta^t$ and $\mathbf{A}^t$ within the interval $[t-\tau, t]$, where τ is a small time step. In the beginning, we sample the sequence in CDRs from the uniform distribution i.e. $\mathbf{A}^1 \sim (\mathcal{U}(\{0,\dots,19\}))^{\otimes N}$. The number of transitions from $\mathbf{A}^t$ to $\mathbf{A}^{t-\tau}$ in this interval follows a Poisson distribution with a mean of $\tau\hat{\mathbf{S}}_\theta^t(a^t, a^{t-\tau})$. At time $t-\tau$, we aggregate all transitions:

$$\mathbf{A}^{t-\tau} = \mathbf{A}^t + \sum_{d=1}^{N} \sum_{s=0\setminus a_d^t}^{19} P_s(s - a_d^t)\mathbf{e}_d. \qquad (11)$$

Here, $\mathbf{e}_d$ is a one-hot vector with a 1 at the dth dimension, and P_s is a Poisson variable, $P_s \sim \text{Poisson}(\tau\hat{\mathbf{S}}_\theta^t(a^t, s))$.

Sampling algorithm for structure For sampling in the $\text{SE}(3)^N$ space of structures, we apply Euler-Maruyama discretization of Eq. (4), following established methods (Yim et al., 2023). In the initial phase of our sampling procedure, we adopt distinct sampling approaches for translations and rotations. For translations, we use a standard Gaussian distribution in $\mathbb{R}^3$: $\mathbf{Z}_n^{\mathbf{x}} \sim \mathcal{N}(0, \text{Id}_3)$. For rotations, we employ a uniform distribution on $\text{SO}(3)$: $\mathbf{Z}_n^{\mathbf{R}} \sim \text{IGSO}(3)$. Our score network is also designed to generate the {Cα, C, N, O} atoms.

Upon generating the backbone and sequence, we construct the remaining side-chain atoms using the Rosetta side-chain packing methods (Misura et al., 2004). Finally, the full atom structure in CDRs is refined using Rosetta's *FastRelax* (Alford et al., 2017) component using the *ref2015* score function.

4. Experiments

In section 4.1, we assess the performance of AbX on the RAbD test dataset. In section 4.2, we evaluate AbX in antibody optimization using the DiffAb test dataset. In section 4.3, we present the results of the ablation studies on the key components of AbX. Finally, we illustrate case studies to validate the efficacy of introduced contraints by showing the trajectories of designed CDRs structures in section 4.4..

4.1. Sequence and Structure Co-design

We evaluate AbX in CDRs structure and sequence co-design on the RAbD test dataset, consisting of 60 diverse antibody-antigen complexes (Adolf-Bryfogle et al., 2018). In this task, we design all six CDRs, conditioned on the antibody framework regions and the target antigens. For model training, we utilized data collected from the Structural Antibody Database (SAbDab) (Dunbar et al., 2014) up to July 2023. (Dunbar et al., 2014). We strictly remove the overlap between the training and testing sets using a CDR-H3 sequence identity threshold of 40%. Further details on model training are available in the Appendix C.

Baseline models As discussed in Section 2.1, computational antibody design methods can be categorized into two groups: discriminative models and generative models. To benchmark the performance, we incorporate state-of-the-art methods from both categories. These include DiffAb (Luo et al., 2022), a diffusion-based generative model, and dyMEAN (Kong et al., 2023b), which employs a graph neural network. Further details on these baseline models can be found in Appendix D.

Evaluation metrics The models are evaluated using various metrics: (1) Amino Acid Recovery (AAR, %): measures the sequence recovery accuracy of the generated sequences compared with the native sequences; (2) Root Mean Square Deviation (RMSD, Å): calculates the structure deviation between Cα coordinates of generated and native CDRs; (3) Plausibility: is assessed using likelihood under another independent antibody language model, AntiBERTy (Ruffolo et al., 2021); (4) Improvement Percentage (IMP, %): indicates the proportion of designed antibodies with improved binding energy (ΔG), computed using *InterfaceAnalyzer* in Rosetta.

While the first four and last two residues in the CDR-H3 are highly conserved, the middle loop residues predominantly contribute to antigen binding (Kong et al., 2023b; Huang et al., 2023). Therefore, we introduce additional metrics that evaluate performance specifically on the middle loop residues. These metrics include Loop RMSD and Loop AAR, assessing RMSD and AAR in the middle loop region within CDR-H3, respectively. For each antibody-antigen complex in the test dataset, we design 100 candidates and calculate the average of these metrics across the candidates. Further details on the metrics are available in Appendix D.2.

Experimental results As shown in Table 1, AbX outperforms other methods in each metric. Particularly, AbX exhibits a significant improvement in IMP and Plausibility, indicating the efficacy of the introduced constraints in generating more plausible antibodies capable of binding to target antigens. Moreover, Table 2 provides detailed metrics for each CDR region, showing AbX's superior performance in both AAR and RMSD compared to both the discriminative and generative methods across all CDR regions.

Table 1. Evaluation of *de novo* designed CDRs in RAbD test dataset.

Metrics	DiffAb	dyMEAN	AbX
IMP(%)↑	12.07	0.00	**18.64**
Plausibility ↑	-1.38	-1.21	**-1.01**
Loop AAR↑	21.25	22.25	**30.80**
Loop RMSD↓	3.45	5.14	**3.24**

In addition, we have developed a full-atom version of AbX, named AbX-FullAtom, which generates both the side-chain and backbone atoms. As shown in Table S1 and Table S2, AbX-FullAtom demonstrates performance comparable to AbX. The advantage of AbX-FullAtom lies in its ability to efficiently generate side-chain atoms without the need for separate side-chain packing using Rosetta, which can be time-consuming. More details on AbX-FullAtom are available in Appendix B and Appendix D.

4.2. Antibody Optimization

We further evaluate AbX on the DiffAb testing set (Luo et al., 2022), comprising 19 well-known antigens including SARS-CoV-2 and MERS. All methods followed the dataset-splitting approach established by DiffAb and retrained for a fair comparison. We first evaluate the performance on DiffAb testing set using the same metrics previously employed for RAbD test set. As shown in Table 3, AbX maintains superior performance over dyMean and DiffAb across all metrics.

Table 3. Evaluation of *de novo* designed CDR-H3 in DiffAb test dataset.

Metrics	DiffAb	dyMEAN	AbX
IMP(%)↑	37.13	5.26	**42.93**
Plausibility ↑	-0.70	-0.77	**-0.67**
Loop AAR↑	21.76	29.31	**30.25**
Loop RMSD↓	3.60	4.80	**3.41**

We next evaluate AbX in antibody optimization. Optimizing existing antibodies to enhance binding affinity has broad applications in therapeutics. Following the optimization outlined in DiffAb, we utilize the CDRs of existing antibodies as the starting point for the inference process.

In this task, we specifically compare AbX with the generative model DiffAb. Given that AbX and DiffAb use different noise schedules, we compare the performance at each noise scale level instead of the optimization step. Further details on this experiment are provided in Appendix E.

As illustrated in Figure 3, the antibodies optimized by AbX consistently exhibit higher binding affinity than those optimized by DiffAb across different noise scales. This highlights the superior efficacy of AbX in antibody optimization.

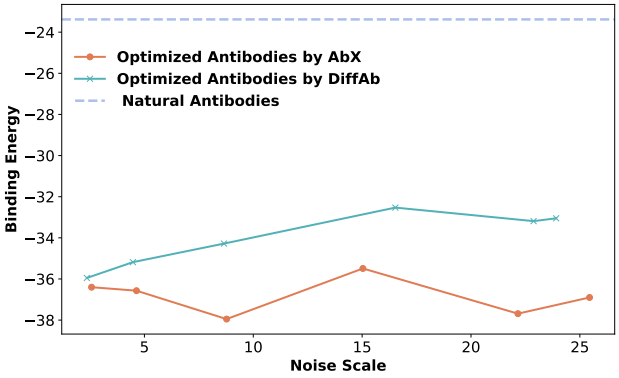


Figure 3. Binding energy of designed antibody-antigen complex in different noise scales. Binding energy is used as an approximation for binding affinity in antibody-antigen interactions.

4.3. Ablation Studies

To validate our motivation and investigate the relative contributions of key components of AbX, we conducted a series of ablation studies, specifically: (1) Without both geometric and physical constraints. (2) Without the physical constraint. (3) Without the evolutionary constraint. For these experiments, we employed all models trained on the DiffAb dataset as detailed in Section 4.2. In our ablation studies, we generated all six CDRs using these models.

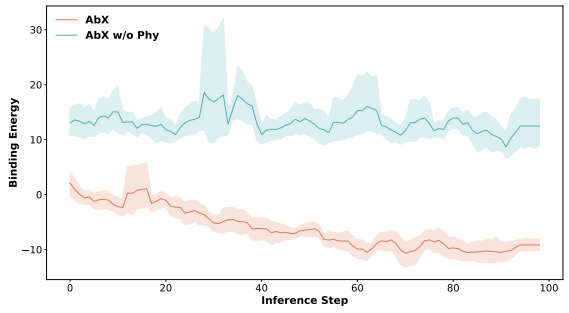
Table 4 illustrates that the inclusion of geometric, physical, and evolutionary constraints significantly contributes to the enhanced performance of AbX. The removal of the evolutionary constraint led to reduced Plausibility and Loop AAR, which underscores the critical role of evolutionary constraint in achieving evolutionary plausible antibodies. Furthermore, the exclusion of evolutionary constraint affected both IMP and RMSD, indicating its importance in generating structurally accurate and plausible antibodies. The absence of geometric and physical constraints resulted in a marked decrease in RMSD and IMP demonstrating that these constraints are crucial for generating antibodies with tight binding to target antigens. The decrease in IMP observed without the physical constraint confirms its significance in the generation of antibodies with low binding affinity. Collectively, these findings underscore the vital contribution of each constraint to the superior performance of our model.

Table 2. Evaluation of *de novo* designed CDRs across each CDR in RAbD test dataset.

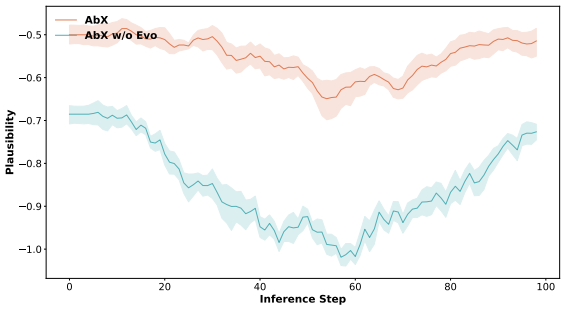
CDR	Method	AAR(%) ↑	RMSD(Å)↓	CDR	Method	AAR(%) ↑	RMSD(Å) ↓
	DiffAb	70.01	0.88		DiffAb	61.07	0.85
H1	dyMEAN	75.71	1.09	L1	dyMEAN	75.55	1.03
	AbX	**80.72**	**0.85**		AbX	**79.37**	**0.78**
	DiffAb	38.52	0.78		DiffAb	58.58	0.55
H2	dyMEAN	68.48	1.11	L2	dyMEAN	83.09	0.66
	AbX	**70.73**	**0.76**		AbX	**84.53**	**0.45**
	DiffAb	28.05	2.86		DiffAb	47.57	1.39
H3	dyMEAN	37.50	3.88	L3	dyMEAN	52.11	1.44
	AbX	**45.18**	**2.50**		AbX	**65.92**	**1.18**

Table 4. Ablation studies for AbX in DiffAb test dataset.

Geometric Constraint	Physical Constraint	Evolutionary Constraint	IMP (%)	Plausibility	H3 AAR(%)	H3 RMSD
✓	✓	✓	**54.82**	**-0.67**	49.17	**2.68**
✓	✓	✗	46.50	-0.77	45.32	3.19
✗	✗	✓	19.36	-0.70	53.21	3.62
✓	✗	✓	52.02	-0.69	**53.84**	2.86



(a)



(b)

Figure 4. (a) Binding energy of designed antibody-antigen complexes during the diffusion process. (b) Plausibility of designed antibody-antigen complexes during the diffusion process.

4.4. Case Studies on Trajectories of Antibody Design

To illustrate the contribution of the introduced constraints in AbX, we compare the trajectories of antibodies designed by AbX with the ablation models that exclude these constraints. We selected a specific antibody-antigen complex (PID: 5TLJ) and designed all six CDRs for the antigen. We generated 10 antibodies, and at each diffusion step, we assessed both the binding energy to the target antigen and evolutionary plausibility.

First, as illustrated in Figure 4, the binding energy decreases progressively during the inference process for all models. Second, when compared with the ablation model lacking the physical and geometric constraints (referred to as AbX w/o Phy), AbX consistently produces antibodies with lower binding energy at each inference step. Third, in comparison with the ablation model that omits the evolutionary constraint (denoted as AbX w/o Evo), AbX consistently yields antibodies with higher evolutionary plausibility. Fourth, the number of atom clashes in generated antibody structures decreases as the inference progresses (Figure S2 in Appendix F).

5. Conclusions

To address the challenge of limited training data in computational antibody design, we present AbX, a score-based diffusion model guided by physical, geometric, and evolutionary constraints. Our results demonstrate that the intro-

duced constraints in AbX lead to significant improvements in antibody design and optimization.

While we have employed various *in silico* metrics to evaluate the performance, one main limitation of our work is the absence of wet-lab experimental validation, which remains our future work. Our model holds potential for several extensions. For instance, we can enhance it by incorporating additional classical physical constraints, such as modeling full-atom energy using established force fields like AMBER (Maier et al., 2015).

6. Code Avilibility

The AbX software is available on Github (https://github.com/zhanghaicang/carbonmatrix_public)

Impact Statement

This paper presents work whose goal is to advance the field of AI for Science. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgements

We acknowledge the financial support from the National Natural Science Foundation of China (grant no. 32370657) and the Project of Youth Innovation Promotion Association CAS to H.Z. We also acknowledge the financial support from the Development Program of China (grant no. 2020YFA0907000) and the National Natural Science Foundation of China (grant nos. 32271297 and 62072435). We thank the ICT Computing-X Center, Chinese Academy of Sciences, for providing computational resources.

We thank Xiaoyang Hou and Zaikai He for the useful discussions.

References

Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B. D., Hu, X., Adachi, Y., Schief, W. R., and Dunbrack Jr, R. L. RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLoS computational biology*, 14(4):e1006112, 2018.

Al-Lazikani, B., Lesk, A. M., and Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *Journal of molecular biology*, 273(4):927–948, 1997.

Alamdari, S., Thakkar, N., van den Berg, R., Lu, A. X., Fusi, N., Amini, A. P., and Yang, K. K. Protein generation with evolutionary diffusion: sequence is all you need. *bioRxiv*, pp. 2023–09, 2023.

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., et al. The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation*, 13(6):3031–3048, 2017.

Austin, J., Johnson, D. D., Ho, J., Tarlow, D., and Van Den Berg, R. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993, 2021.

Campbell, A., Benton, J., De Bortoli, V., Rainforth, T., Deligiannidis, G., and Doucet, A. A continuous time framework for discrete denoising models. *Advances in Neural Information Processing Systems*, 35:28266–28279, 2022.

Carter, P. J. Potent antibody therapeutics by design. *Nature reviews immunology*, 6(5):343–357, 2006.

Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J., and Deane, C. M. SAbDab: the structural antibody database. *Nucleic acids research*, 42 (D1):D1140–D1146, 2014.

Eguchi, R. R., Choe, C. A., and Huang, P.-S. Ig-VAE: Generative modeling of protein structure by direct 3D coordinate generation. *PLOS Computational Biology*, 18 (6):1–18, 06 2022. doi: 10.1371/journal.pcbi.1010271.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. Protein complex prediction with AlphaFold-Multimer. *biorxiv*, pp. 2021–10, 2021.

Gao, K., Wu, L., Zhu, J., Peng, T., Xia, Y., He, L., Xie, S., Qin, T., Liu, H., He, K., et al. Pre-training Antibody Language Models for Antigen-Specific Computational Antibody Design. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 506–517, 2023.

Ge, S., Nah, S., Liu, G., Poon, T., Tao, A., Catanzaro, B., Jacobs, D., Huang, J.-B., Liu, M.-Y., and Balaji, Y. Preserve Your Own Correlation: A Noise Prior for Video Diffusion Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22930–22941, October 2023.

Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 115(4):1716–1733, 2001.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Hie, B. L., Shanker, V. R., Xu, D., Bruun, T. U., Weidenbacher, P. A., Tang, S., Wu, W., Pak, J. E., and Kim, P. S. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 2023.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Huang, Y., Paul, S., and Marks, D. An Energy Based Model for Incorporating Sequence Priors for Target-Specific Antibody Design. In *NeurIPS 2023 Generative AI and Biology (GenBio) Workshop*, 2023.

Hur, J., Choi, J., Han, G., Lee, D.-J., and Kim, J. Expanding expressiveness of diffusion models with limited data via self-distillation based fine-tuning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5028–5037, 2024.

Ingraham, J. B., Baranov, M., Costello, Z., Barber, K. W., Wang, W., Ismail, A., Frappier, V., Lord, D. M., Ng-Thow-Hing, C., Van Vlack, E. R., et al. Illuminating protein space with a programmable generative model. *Nature*, pp. 1–9, 2023.

Jin, P., Li, H., Cheng, Z., Li, K., Ji, X., Liu, C., Yuan, L., and Chen, J. DiffusionRet: Generative Text-Video Retrieval with Diffusion Model. In *ICCV*, pp. 2470–2481, 2023.

Jin, W., Barzilay, R., and Jaakkola, T. Antibody-antigen docking and design via hierarchical structure refinement. In *International Conference on Machine Learning*, pp. 10217–10227. PMLR, 2022.

Jing, B., Erives, E., Pao-Huang, P., Corso, G., Berger, B., and Jaakkola, T. S. EigenFold: Generative Protein Structure Prediction with Diffusion Models. In *ICLR 2023-Machine Learning for Drug Discovery workshop*, 2023.

Jørgensen, E. The central limit problem for geodesic random walks. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1-2):1–64, 1975.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

Kong, X., Huang, W., and Liu, Y. Conditional Antibody Design as 3D Equivariant Graph Translation. In *The Eleventh International Conference on Learning Representations*, 2023a.

Kong, X., Huang, W., and Liu, Y. End-to-End Full-Atom Antibody Design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17409–17429. PMLR, 23–29 Jul 2023b.

Lapidoth, G. D., Baran, D., Pszolla, G. M., Norn, C., Alon, A., Tyka, M. D., and Fleishman, S. J. AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins: Structure, Function, and Bioinformatics*, 83(8):1385–1406, 2015.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574.

Liu, S., Zhu, T., Ren, M., Chungong, Y., Bu, D., and Zhang, H. Predicting mutational effects on protein-protein binding via a side-chain diffusion probabilistic model. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Luo, S., Su, Y., Peng, X., Wang, S., Peng, J., and Ma, J. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances in Neural Information Processing Systems*, 35: 9754–9767, 2022.

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

Mariani, V., Biasini, M., Barbato, A., and Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013.

Martinkus, K., Ludwiczak, J., LIANG, W.-C., Lafrance-Vanasse, J., Hotzel, I., Rajpal, A., Wu, Y., Cho, K., Bonneau, R., Gligorijevic, V., and Loukas, A. AbDiffuser:

full-atom generation of in-vitro functioning antibodies. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Mei, K. and Patel, V. Vidm: Video implicit diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 9117–9125, 2023.

Misura, K. M., Morozov, A. V., and Baker, D. Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. *Journal of molecular biology*, 342(2):651–664, 2004.

North, B., Lehmann, A., and Dunbrack Jr, R. L. A new clustering of antibody CDR loop conformations. *Journal of molecular biology*, 406(2):228–256, 2011.

Pantazes, R. and Maranas, C. D. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Engineering, Design & Selection*, 23(11):849–858, 2010.

Peng, Z., Han, C., Wang, X., Li, D., and Yuan, F. Generative diffusion models for antibody design, docking, and optimization. *bioRxiv*, pp. 2023–09, 2023.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

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

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2020.

Vahdat, A., Kreis, K., and Kautz, J. Score-based generative modeling in latent space. *Advances in Neural Information Processing Systems*, 34:11287–11302, 2021.

Villegas-Morcillo, A., Weber, J., and Reinders, M. Guiding diffusion models for antibody sequence and structure codesign with developability properties. In *NeurIPS 2023 Generative AI and Biology (GenBio) Workshop*, 2023.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620 (7976):1089–1100, 2023.

Yim, J., Trippe, B. L., De Bortoli, V., Mathieu, E., Doucet, A., Barzilay, R., and Jaakkola, T. SE(3) Diffusion Model with Application to Protein Backbone Generation. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Zhang, Y., Zhang, Z., Zhong, B., Misra, S., and Tang, J. DiffPack: A Torsional Diffusion Model for Autoregressive Protein Side-Chain Packing. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

A. Details of Diffusion Processes

A.1. Diffusion Process for Sequence

For the completeness of our study, we provide a brief introduction to the Continuous Time Markov Chain (CTMC). A CTMC is a continuous stochastic process $\{a^t\}_{t\in\mathcal{U}(0,1)}$ satisfying the Markov property. In this process, a^t takes values in a discrete state space, and a CTMC is characterized by its transition probabilities. The subsequent state in the chain is selected based on a transition probability $p_{t|t-\Delta t}(a|\tilde{a})$. We define this transition probability with a transition rate matrix $\mathbf{S}\in\mathbb{R}^{20\times 20}$, formulated as:

$$\mathbf{S}(\tilde{a},a)=\lim_{\Delta t\to 0}\frac{p_{t|t-\Delta t}(a|\tilde{a})-\delta_{a,\tilde{a}}}{\Delta t}, \tag{12}$$

$$p_{t|t-\Delta t}(a|\tilde{a})=\delta_{a,\tilde{a}}+\mathbf{S}(\tilde{a},a)\Delta t+o(\Delta t), \tag{13}$$

where $\mathbf{S}(\tilde{a},a)$ represents the $(\tilde{a},a)$ element of the transition rate matrix, $p_{t|t-\Delta t}(a,\tilde{a})$ denotes the infinitesimal transition probability from state $\tilde{a}$ at time $t-\Delta t$ to state a at time t. The transition rate matrix has the following properties: $\mathbf{S}(\tilde{a},a)\geq 0$ for $\tilde{a}\neq a,\ \mathbf{S}(a,a)\leq 0,\ \mathbf{S}(a,a)=-\sum_{\tilde{a}\neq a}\mathbf{S}(\tilde{a},a)$.

Building on the framework established by τLDR (Campbell et al., 2022), we set the transition rate matrix as $\mathbf{S}^t=\beta(t)\mathbf{S}$, where $\mathbf{S}=\mathbb{1}\mathbb{1}^T-20\mathrm{Id}$, with $\mathbb{1}\mathbb{1}^T$ being a matrix of ones and Id the identity. This leads to the analytic expression of the marginal distribution:

$$p_{t|0}(a|a_0)=\exp(\int_0^t\mathbf{S}^s\mathrm{d}s)=Q\exp[\Lambda\int_0^t\beta(s)ds]Q^{-1}, \tag{14}$$

where $\mathbf{S}=Q\Lambda Q^{-1}$ is the eigendecomposition of matrix $\mathbf{S}$, and exp is applied element-wise. For a more detailed exploration of continuous-time score-based diffusion models for discrete space, we refer to the work of τLDR (Campbell et al., 2022).

A.2. Diffusion Process for Structure

Following the approach outlined by FrameDiff (Yim et al., 2023), we calculate the score on $\mathrm{SE}(3)$ space, utilizing the generated initial structure $\mathbf{T}_\theta^0=(\mathbf{R}_\theta^0,\mathbf{X}_\theta^0)$. This computation enables us to derive the score on both $\mathrm{SO}(3)$ and $\mathbb{R}^3$ as follows:

$$\nabla_{\mathbf{R}}\log p_{t|0}(\mathbf{R}^t|\mathbf{R}_\theta^0)=\frac{\mathbf{R}^t}{\mathbf{W}^t}\log\{\mathbf{R}_\theta^{(0,t)}\}\frac{\partial_{\mathbf{W}}f(\mathbf{W}^t,t)}{f(\mathbf{W}^t,t)}, \tag{15}$$

$$\nabla_{\mathbf{x}}\log p_{t|0}(\mathbf{X}^t|\mathbf{X}_\theta^0)=(1-e^{-t})^{-1}(e^{-\frac{t}{2}}\mathbf{X}_\theta^0-\mathbf{X}^t), \tag{16}$$

where f represents the Brownian motion on $\mathrm{SO}(3)$, $\mathbf{W}(\mathbf{R})$ denotes the rotation angle in radians for any $\mathbf{R}\in\mathrm{SO}(3)$, $\mathbf{R}^{(0,t)}$is defined as $(\mathbf{R}^0)^{\mathrm{T}}\mathbf{R}^t$, and $\mathbf{W}^t=\mathbf{W}(\mathbf{R}^{(0,t)})$. Here, log is the inverse of the exponential map on $\mathrm{SO}(3)$. These equations are fundamental to our diffusion process in $\mathrm{SE}(3)$ space. For a more comprehensive understanding of this process, we refer to the detailed explanation provided in the work of FrameDiff (Yim et al., 2023).

A.3. Details of Model Architectures

Figure S1 illustrates our score network, which draws inspiration from AlphaFold2 (Jumper et al., 2021). Our model utilizes the triangular multiplicative update method to effectively capture complex interactions within antibody-antigen complexes. Additionally, we employ an $\mathrm{SE}(3)$-invariant Invariant Point Attention (IPA) architecture for generating the initial structure $\mathbf{T}_\theta^0$ and sequence marginals $p_{0|t}^\theta(a_0|a,\mathcal{P}_{\mathrm{FC}})$.

The inputs to our score network include the noised sequence and structure, alongside the output from the conditional encoder. The noised sequence $\mathbf{A}^t\cup(\mathbf{A}_{\mathrm{Ab}}\backslash\mathbf{A})$ is processed through the language model ESM-2, and the noised structure serves as the initial frame for the IPA network. Embeddings from both the language model and the conditional encoder undergo processing via the triangular multiplicative update. Subsequently, the IPA generates the final structure and marginal distribution of the sequence.

In the recycling phase, the sequence is obtained by taking the value with the highest probability at each position according to the marginal distribution $p_{0|t}^\theta(a_0|a,\mathcal{P}_{\mathrm{FC}})$. This sampled sequence replaces the noised sequence in the language model.

The final single, pair embeddings and distance map of the structure from the IPA are then incorporated into the conditional encoder output. The entire process described above is then repeated. This recycling process is repeated three times to refine the predictions.

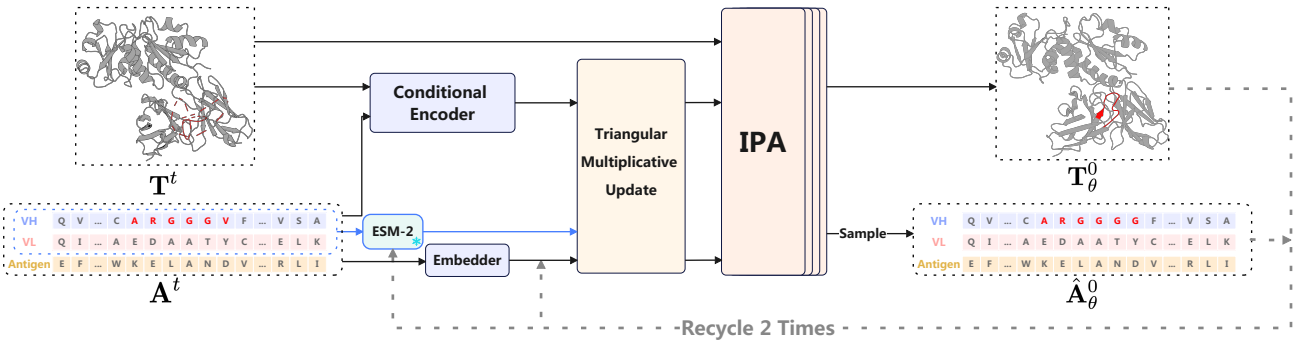


Figure S1. Illustration of AbX's model architectures.

A.4. Details of Training Objectives

Details of physical constraint To learn the geometry of inter-residue covalent bonds and prevent structural violations, particularly in regions lacking ground truth atom coordinates, we incorporate the structural violation loss from AlphaFold2 (Jumper et al., 2021). This loss is comprised of bond length and angle violation losses.

The bond length violation loss is formulated as follows:

$$\mathcal{L}_{\text{bond length}} = \frac{1}{N_{\text{bonds}}} \sum_{i=1}^{N_{\text{bonds}}} \max(|\ell^i_{\text{pred}} - \ell^i_{\text{lit}}| - \tau, 0), \tag{17}$$

where ℓ^i_{pred} denotes the bond length in the generated structure between the i-th N atom and the $(i+1)$-th Cα atom, forming a peptide bond. ℓ^i_{lit} is the literature value of this bond length. N_{bonds} represents the total number of peptide bonds in the generated CDRs structure, and the tolerance τ is set to $12\sigma_{\text{lit}}$, with σ_{lit} being the standard deviation of this bond length from the literature.

For bond angle violations, the loss is calculated using the cosine of the angle derived from the dot product of unit vectors of the bonds:

$$\mathcal{L}_{\text{bond angle}} = \frac{1}{N_{\text{angles}}} \sum_{i=1}^{N_{\text{angles}}} \max(|\cos\alpha^i_{\text{pred}} - \cos\alpha^i_{\text{lit}}| - \tau, 0), \tag{18}$$

where α^i_{pred} is the predicted bond angle, and α^i_{lit} is its corresponding literature value. N_{angles} is the total count of peptide bond angles in the generated CDRs structure. The tolerance τ ranges from -12 to 12 times the standard deviation of this bond angle as per literature.

The overall structure violation loss is computed as the sum of bond angle and bond length violation losses:

$$\mathcal{L}_{\text{violation}} = \mathcal{L}_{\text{bond angle}} + \mathcal{L}_{\text{bond length}}. \tag{19}$$

Details of geometric constraint Drawing inspiration from the breakthroughs achieved by AlphaFold2 (Jumper et al., 2021), our model integrates the Frame Aligned Point Error (FAPE) loss $\mathcal{L}_{\text{FAPE}}$, distogram loss $\mathcal{L}_{\text{distogram}}$, and lDDT-C$\alpha$ loss $\mathcal{L}_{\text{lDDT}}$. The FAPE loss, which is SE(3)-invariant, measures the structural similarity between the generated and ground truth antibodies. It is calculated as:

$$\mathcal{L}_{\text{FAPE}} = \frac{1}{Z}\frac{1}{{N_{\text{res}}}^2} \sum_{i,j=1}^{N_{\text{res}}} \min\left(\sqrt{\left\|T_i^{-1} \circ \mathbf{x}_j - T_i^{\text{true}^{-1}} \circ \mathbf{x}_j^{\text{true}}\right\|}, d_{\text{clamp}}\right), \tag{20}$$

where Z, set to 10Å, is the length scale, and d_{clamp} is a clamping constant and also set to 10Å. N_{res} denotes the number of residues in the antibody-antigen complex. Given the complex interactions within antibody-antigen complexes, our model also considers the interface FAPE loss (Evans et al., 2021).

To enhance the efficiency of pair representations, we incorporate the distogram loss. This involves projecting the symmetrized pair representations ($\mathbf{z}_{ij} + \mathbf{z}_{ji}$) into 64 distance bins to obtain bin probabilities p_{ij}^b using softmax. The bins span distances from 2Å to 22Å. The label y_{ij}^b is encoded as one-hot binned residue distances from ground-truth beta carbon positions. The distogram loss is defined as:

$$\mathcal{L}_{\text{distogram}} = -\frac{1}{N_{\text{res}}^2} \sum_{i,j} \sum_{b=1}^{64} y_{ij}^b \log p_{ij}^b. \tag{21}$$

To improve the stereochemical plausibility of the generated structure, we include the per-residue lDDT-Cα loss (Mariani et al., 2013). This metric is computed against the ground truth structure and discretized into 50 bins, denoted as $p_i^{\text{true LDDT}}$. The final single representation from the IPA is projected into 50 bins, referred to as p_i^{pLDDT}, for predicting the score. The per-residue lDDT-Cα loss is described as:

$$\mathcal{L}_{\text{lDDT}} = \frac{1}{N_{\text{res}}} \sum_{i=1}^{N_{\text{res}}} (p_i^{\text{true LDDT}^T} \log p_i^{\text{pLDDT}}). \tag{22}$$

Details of denoising score matching losses Following the approach described by FrameDiff (Yim et al., 2023), the Denoising Score Matching (DSM) loss is computed as follows:

$$\mathcal{L}_{\text{DSM}}^{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^{N} \lambda_t \|\nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}^t|\mathbf{R}_\theta^0) - \nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}^t|\mathbf{R}^0)\|^2, \tag{23}$$

$$\mathcal{L}_{\text{DSM}}^{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_\theta^0 - \mathbf{x}^0\|^2, \tag{24}$$

where λ_t is calculated as $\lambda_t = 1/(\frac{1}{N}\sum_{i=1}^{N} \nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}^t|\mathbf{R}^0))$. The SE(3) DSM loss is then defined as the sum of these individual losses:

$$\mathcal{L}_{\text{DSM}} = \mathcal{L}_{\text{DSM}}^{\mathbf{R}} + \mathcal{L}_{\text{DSM}}^{\mathbf{x}}. \tag{25}$$

A.5. Details of Sampling Algorithm

Details of the overall sampling algorithm are provided in Algorithm 1.

B. Full-atom version of AbX

We have developed AbX-FullAtom, a full-atom extension of the AbX model, capable of generating antibodies with complete backbone and side-chain atoms. Following AlphaFold2 (Jumper et al., 2021), side-chain atoms in AbX-FullAtom are represented as torsion angles. Due to the variability in the number of side-chain atoms among different residues, the count of side-chain torsion angles varies for each residue. To account for the varied distribution of torsion angles across different residues, AbX-FullAtom predicts side-chain torsion angles for all possible residues simultaneously. Subsequently, based on the sequence sampled by our model, we select the appropriate torsion angles to construct the final structural output.

The training objectives for the full-atom version are enhanced by incorporating a side-chain FAPE loss (Jumper et al., 2021), and expanding the van der Waals loss to encompass all atoms, not just those in the backbone. This is specifically implemented to ensure the accuracy and quality of the side-chain conformations in the generated structures. Finally, we utilize the *fastrelax* in Rosetta to refine these structures.

C. Implementation Details of AbX

In configuring the diffusion process for structures in both AbX and AbX-FullAtom, we followed a hyperparameter setting consistent with that used in the score-based generative model by FrameDiff (Yim et al., 2023). For translations, we set

Algorithm 1 Sampling Procedure of AbX

Input: $\theta, N, N_{\text{steps}}, \epsilon, \zeta, \mathcal{P}_{\text{FC}}$
$\Delta t = \frac{1-\epsilon}{N_{\text{steps}}}$
$\mathbf{s}, \mathbf{z} = G_\theta(\mathcal{P}_{\text{FC}})$ # Convert $\mathcal{P}_{\text{FC}}$ prior to conditional embeddings
$\mathbf{A}^1 \sim p_{\text{ref}}^{\text{Seq}}, \mathbf{T}^1 \sim P \cdot p_{\text{ref}}^{\text{SE}(3)^N}$ # Initialize sampling from prior distributions
for $t = 1, 1-\Delta t, 1-2\Delta t, \ldots, \epsilon$ **do**
  $\mathbf{T}_\theta^0, p_{0|t}^\theta(\mathbf{A}^0|\mathbf{A}^t, \mathcal{P}_{\text{FC}}) = F_\theta(\mathbf{T}^t.\mathbf{A}^t, \mathbf{s}, \mathbf{z})$
  $\{(s_{\theta,n}^{\mathbf{R}}, s_{\theta,n}^{\mathbf{x}})\}_{n=1}^N = \nabla_{\mathbf{T}^t} \log p_{t|0}(\mathbf{T}^t|\mathbf{T}_\theta^0)$ # Calculate the score function of structure
  $\hat{\mathbf{S}}_\theta^t(a, \tilde{a}) = \mathbf{S}^t(\tilde{a}, a) \sum_{a^*} \frac{p_{t|0}(\tilde{a}|a^*)}{p_{t|0}(a|a^*)} p_{0|t}^\theta(a^*|a, \mathcal{P}_{\text{FC}})$ # Compute transition rate matrix of sequence
  for $(\mathbf{R}_n^t, \mathbf{x}_n^t, a_n^t) = (T_1^t, a_1^t), \ldots, (T_N^t, a_N^t)$ **do**
    $\mathbf{Z}_n^{\mathbf{x}} \sim \mathcal{N}(0, \text{Id}_3), \mathbf{Z}_n^{\mathbf{R}} \sim \text{IGSO}(3)$ # Simulate reverse step for structure
    $W_n^{\mathbf{x}} = P(P\Delta t[\frac{1}{2}\mathbf{x}_n^t + s_{\theta,n}^{\mathbf{x}}] + \zeta\sqrt{\Delta t}\mathbf{Z}_n^{\mathbf{x}})$ # Remove center of mass
    $W_n^{\mathbf{R}} = \Delta t s_{\theta,n}^{\mathbf{R}} + \zeta\sqrt{\Delta t}\mathbf{Z}_n^{\mathbf{R}}$
    $T_n^{t-\Delta t} = \exp_{T_n^t}\{(W_n^{\mathbf{R}}, W_n^{\mathbf{x}})\}$
    $P_s \sim \text{Possion}(\Delta t \hat{\mathbf{S}}_\theta^t(a^t, s))$ # Simulate reverse step for sequence
    $a_n^{t-\Delta t} = a_n^t + \sum_{s=0}^{19} P_s \times (s - a_n^t)$
    $a_n^{t-\Delta t} = \text{Clamp}(a_n^{t-\gamma}, \min = 1, \max = 19)$
  end for
end for
$\mathbf{T}_\theta^0, p_{0|t}^\theta(\mathbf{A}^0|\mathbf{A}^\epsilon, \mathcal{P}_{\text{FC}}) = F_\theta(\mathbf{T}^\epsilon.\mathbf{A}^\epsilon, \mathbf{s}, \mathbf{z})$
$\hat{\mathbf{A}}_\theta^0 = \text{Sample}(p_{0|t}^\theta(\mathbf{A}^0|\mathbf{A}^\epsilon, \mathcal{P}_{\text{FC}}))$
Output: $\mathbf{T}_\theta^0, \hat{\mathbf{A}}_\theta^0$

$f_x(s) = \frac{1}{2}\beta(s)$ and $g_x(s) = \sqrt{\beta(s)}$, with the schedule $\beta(s)$ defined as $\beta(s) = \beta_{\min} + t(\beta_{\max} - \beta_{\min})$. Here, $\beta_{\min}$ and $\beta_{\max}$ are set at 0.1 and 20, respectively. For rotations, the diffusion schedule is given by $\sigma_r(s) = \log(s \cdot \exp(\sigma_{\max}) + (1-s)\exp(\sigma_{\min}))$, where $\sigma_{\min}^2 = 0.01$ and $\sigma_{\max}^2 = 2.25$. For the diffusion process for sequences in both AbX and AbX-FullAtom, we selected $\beta(t)$ as 0.3.

The dimensions of the single and pair representations, as well as the hidden representation in the IPA layer, are set to 512, 128, and 128, respectively. We utilize eight IPA layers and one triangular multiplicative update layer, bringing the total count of trainable parameters to 11,859,323. In our model, the ESM-2 (3B) language model is utilized, which is accessible via their GitHub repository (`https://github.com/facebookresearch/esm`). During training, the Adam optimizer was employed with a batch size of 128. The learning rate followed a polynomial ramp from 0 to 100. We set the total number of training iterations at 25,000, and the reverse diffusion model operates with 100 steps. The training involved random sampling of combinations of CDRs for masking and generation, consistent with the training strategy outlined in the *codesign_multicdrs.yml* configuration file used in DiffAb.

D. Details of Structure and Sequence Co-design Tasks

D.1. Baseline Models

DiffAb (Luo et al., 2022) We utilized DiffAb from the GitHub repository (`https://github.com/luost26/diffab`). To ensure a fair comparison, we retrained DiffAb on the same dataset as our model, using the configuration file *codesign_multicdrs.yml*. For the designed antibodies, side-chain atoms were added using Rosetta Packing, and the structures were further refined using the *fastrelax* feature in Rosetta.

dyMEAN (Kong et al., 2023b) We employed dyMEAN from its GitHub repository (`https://github.com/THUNLP-MT/dyMEAN`). For comparability, dyMEAN was retrained on the same dataset as ours, adhering to the *multi_cdr_design.json* configuration. Similar to DiffAb, the antibodies designed by dyMEAN were refined using *fastrelax* in Rosetta.

RAbD (Adolf-Bryfogle et al., 2018) We intended to include Rosetta Antibody Design (RAbD) in our comparative analysis.

However, the preprocessing service PyIgClassify (http://dunbrack2.fccc.edu/PyIgClassify) for antibodies, was not operational, preventing a direct comparison with our model.

D.2. Metrics

Plausibility For evaluating the plausibility of our designed sequences, we utilize the pseudo log-likelihood scores calculated by antibody language model AntiBerty (Ruffolo et al., 2021), in line with the approach described in AbDiffuser (Martinkus et al., 2023). AntiBerty, an antibody-specific transformer language model, has been pre-trained on a dataset comprising 558M natural antibody sequences. We implemented AntiBerty from their GitHub repository (https://github.com/jeffreyruffolo/AntiBERTy).

Improvement Percentage (IMP, %) To assess IMP, we calculate the binding energy of antibody-antigen complexes using the *InterfaceAnalyzer* and the *ref2015* score function in pyRosetta. IMP is then determined by computing the proportion of designed antibody-antigen complexes whose binding energy are lower (indicating improvement) than those in natural complexes. For the relaxation and scoring of these complexes, we utilized the 2023.33 version of pyRosetta.

Root Mean Square Deviation (RMSD) Initially, we align the generated antibody-antigen complexes with their corresponding natural complexes using Kabsch alignment. Subsequently, we calculate the RMSD for each region of these aligned complexes.

Loop AAR and **Loop RMSD** DiffAb utilizes the Chothia scheme for numbering CDRs, whereas dyMEAN and AbX employ the IMGT scheme. Given the difference in CDR numbering between these schemes, we standardize the comparison by adjusting for CDR lengths. Specifically, for models using the IMGT scheme, we exclude the first four and last two residues, while for the Chothia scheme, we omit the first two and last two residues. This approach ensures we focus on the middle loop residues of equivalent lengths across different schemes, thereby enabling a fair and accurate comparison.

D.3. Results of AbX-FullAtom

In our experimental setup, AbX-FullAtom was trained using the identical dataset and configuration parameters as those employed for AbX. We employed AbX and AbX-FullAtom to generate 100 candidate structures for each antibody-antigen complex in the RAbD test dataset. Subsequently, we calculated the average values for each metric across all these candidates. In Table S1, we observe that AbX-FullAtom exhibits marginally lower performance compared to AbX. Despite this, AbX-FullAtom still outperforms other methods, underscoring the effectiveness of our models, AbX and AbX-FullAtom.

Table S1. Evaluation of *de novo* designed CDRs in RAbD test dataset.

Metrics	AbX-FullAtom	AbX
IMP(%)↑	16.29	**18.64**
Plausibility ↑	-1.07	**-1.01**
Loop AAR↑	30.09	**30.80**
Loop RMSD↓	3.55	**3.24**

Table S2 presents the RMSD and AAR metrics for each CDR. Here, AbX-FullAtom demonstrates comparable performance to AbX. These findings confirm that our full-atom antibody generation method, AbX-FullAtom, is also reliable.

E. Details of CDR-H3 Generation and Optimization Task

We adopted a data-splitting approach similar to that of DiffAb for the CDR-H3 generation task. Our training dataset, derived from the SAbDab database accessed in July 2023, was clustered based on 50% sequence identity in CDR-H3 sequences. We excluded clusters containing test data which encompasses 19 antibody-antigen complexes. To ensure a fair comparison, DiffAb was retrained using this same dataset, employing the *codesign_multicdrs.yml* configuration file. In this task, the original CDR-H3 is removed from the antibody-antigen complexes, and both the sequence and structure of the missing CDR-H3 are sampled.

In the optimization task, we followed the methodology proposed by DiffAb to optimize antibodies (Luo et al., 2022). This

Table S2. Evaluation of *de novo* designed CDRs across each CDR in RAbD test dataset.

CDR	Method	AAR(%) ↑	RMSD(Å)↓	CDR	Method	AAR(%) ↑	RMSD(Å) ↓
H1	AbX-FullAtom	**81.16**	0.87	L1	AbX-FullAtom	74.24	0.88
	AbX	80.72	**0.85**		AbX	**79.37**	**0.78**
H2	AbX-FullAtom	67.32	0.83	L2	AbX-FullAtom	78.79	0.53
	AbX	**70.73**	**0.76**		AbX	**84.53**	**0.45**
H3	AbX-FullAtom	**45.32**	2.72	L3	AbX-FullAtom	57.12	**1.13**
	AbX	45.18	**2.50**		AbX	**65.92**	1.18

process involves perturbing the CDR sequence and structure at time t using forward diffusion, then denoising from time t to time 0 in reverse diffusion to generate a set of 100 optimized antibody candidates. Due to differences in the diffusion process between AbX and DiffAb, we replace optimization steps in DiffAb with noise scale levels in the diffusion process for structural comparison. The noise scale is determined by calculating the RMSD of Cα atoms between the noised structures and the corresponding ground truth structures. The objective of the task is to identify the best candidate with optimal binding energy to the target antigen across various noise scales. For this purpose, the average binding energy of the top candidates from the test dataset is calculated using Rosetta. This average binding energy is then selected as the key metric for assessment.

F. Visualization of Generated Antibody-antigen Complex During the Generative Process

Figure S2 presents antibody-antigen complexes generated by our score network at each timestep. Our analysis reveals a decrease in the number of atom clashes and binding energy as the inference proceeds. Notably, the final generated complex exhibits improved binding affinity compared to the natural complex. These findings underscore the effectiveness of our biophysical constraints in enhancing antibody-antigen complex generation at each timestep.

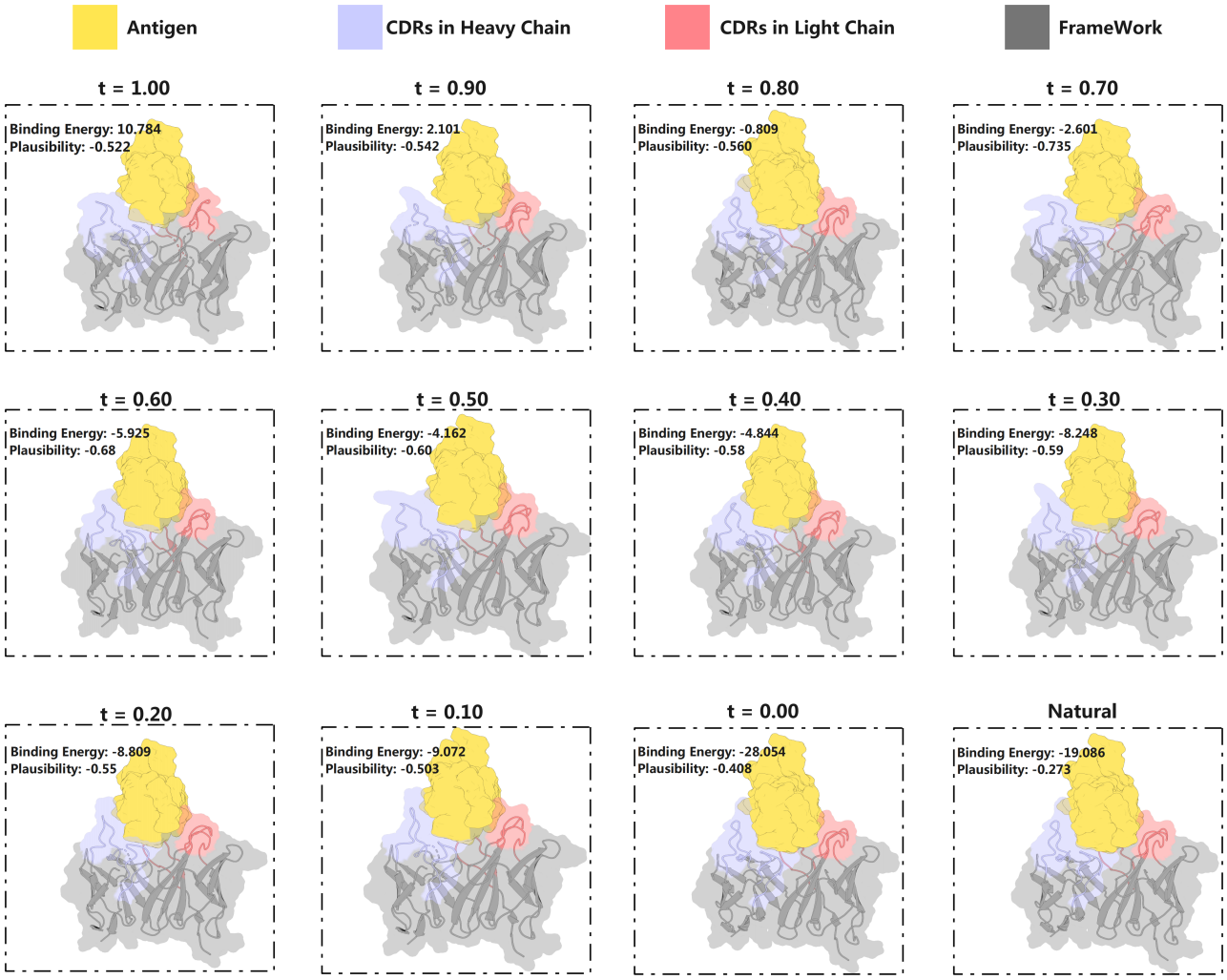


Figure S2. Visualization of generated antibody-antigen complexes during the generative process. The heavy, light, and antigen chains of the antibody-antigen complex (PID:5TLJ) are denoted as D, C, and X.