# Structure-Guided Reinforcement Learning for High-Affinity Antibody Design

**Hanqun Cao** [*1] **Shuaike Shen** [*2] **Weihao Xuan** [3] **Jian Ma** [2] **Pheng Ann Heng** [1] **Fang Wu** [4]

## Abstract

Antibodies enable precise targeted therapies for immune-related diseases through specific recognition and binding to pathogens or inflammatory factors. While recent advances in deep learning have enabled structure-based antibody design with promising accuracy, existing generative models primarily mimic natural antibody distributions without explicit optimization toward binding affinity. To address this gap, we propose EvoAb, a reinforcement learning framework that combines a pretrained antibody co-design model with thermodynamics-grounded reward modeling for high-affinity antibody design. We introduce Multi-round Mutation Policy Optimization (M2PO), an iterative algorithm that integrates affinity-weighted sequence learning with distribution anchoring, progressively enhancing binding affinity while preserving structural plausibility. By leveraging structure-aware reward signals, EvoAb enables efficient *in silico* directed evolution without expensive physics-based calculations or wet-lab experiments. Extensive experiments demonstrate that EvoAb achieves state-of-the-art binding affinity optimization, reducing mean $\Delta\Delta G$ from 2.89 to 2.45 kcal/mol and increasing stabilizing mutation rates from 13% to 19%. Cross-validation with FoldX confirms that our optimization yields physically meaningful improvements, highlighting EvoAb's potential for accelerating therapeutic antibody development.

## 1. Introduction

Antibodies are Y-shaped proteins synthesized by B cells to identify and neutralize foreign pathogens (Chiu et al., 2019). These molecules achieve remarkable specificity by recognizing small regions on target antigens known as epitopes, typically spanning 5–8 amino acids (Chiu et al., 2019). This specificity arises from six hypervariable loops called complementarity-determining regions (CDRs)—three on the heavy chain variable domain (H1, H2, H3) and three on the light chain (L1, L2, L3) (Rabia et al., 2018; Jeon & Kim, 2024), shown in Figure 1. Together, these loops form the paratope, the molecular surface that complementarily engages the antigen's epitope (Chiu et al., 2019; Haakenson et al., 2018). Among the six CDRs, CDR-H3 exhibits the greatest sequence and structural diversity, serving as the primary determinant of antigen recognition and binding affinity. Consequently, engineering CDR-H3 to enhance binding affinity has emerged as a central strategy for developing therapeutic antibodies with improved efficacy, specificity, and clinical outcomes (Paul et al., 2024; Scott et al., 2012).

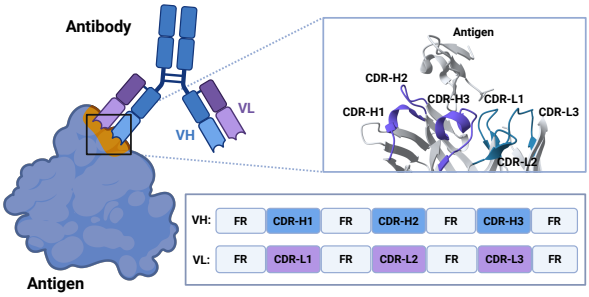


*Figure 1.* **Overview of antibody structure and antigen-binding interface. (Left)** Antibody-antigen complex with Variable Heavy ($V_H$) and Variable Light ($V_L$) domains engaging the target epitope. **(Top-right)** Six hypervariable CDR loops (H1–H3, L1–L3) form the paratope that determines binding specificity. **(Bottom-right)** Domain architecture showing CDR loops interspersed within conserved Framework Regions (FRs).

Computational approaches for CDR design have undergone rapid evolution over the past decade. Early efforts relied on physics-based energy functions and combinatorial sampling strategies (Weitzner et al., 2017; Pantazes & Maranas, 2010; Ravn et al., 2010), which, while interpretable, often struggled to capture the complex sequence-structure-function relationships underlying antibody binding. The advent of deep learning has transformed this landscape: geometric neural networks and generative models now enable

[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong [2]Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University [3]RIKEN AIP [4]Department of Computer Science, Stanford University. Correspondence to: Fang Wu <ffangwu97@stanford.edu>.

*Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)*

direct learning from structural data, achieving substantial improvements in CDR sequence recovery and structural accuracy (Jin et al., 2021; Luo et al., 2022; Abanades et al., 2023; 2022; Kong et al., 2022; Chen et al., 2023; Wu & Li, 2024; Wu et al., 2026). More recently, structure prediction foundation models—including AlphaFold 3 (Abramson et al., 2024), Boltz-2 (Passaro et al., 2025), and Chai-2 (Team et al., 2025a)—have demonstrated remarkable capabilities in modeling antibody-antigen interactions with atomic-level precision. Building on these advances, emerging design frameworks (Team et al., 2025b) have pushed the boundaries further, enabling zero-shot design of binders with favorable developability profiles.

Despite these advances, a critical gap remains: existing generative models primarily learn to mimic natural antibody distributions without explicit optimization toward binding affinity, which is essential for therapeutic development. Recent efforts have explored learning-based approaches for affinity enhancement, including reinforcement learning with sequence-based $\Delta\Delta G$ predictors (Cao et al., 2025) and multi-objective optimization with physics-based energy functions (Wen et al., 2025). However, two fundamental challenges persist. First, reward signals are either unreliable or computationally prohibitive: sequence-based predictors often lack accuracy due to their neglect of structural context, while physics-based methods such as FoldX (Schymkowitz et al., 2005) and Rosetta (Rohl et al., 2004) are too slow for iterative training. Second, naive optimization tends to drift toward implausible sequences that exploit reward model artifacts, necessitating careful regularization to maintain structural validity.

To address these challenges, we propose **EvoAb**, a reinforcement learning framework that bridges structure-aware antibody generation with efficient affinity optimization. EvoAb integrates a pretrained antibody co-design foundation model for atomic-level interface modeling with a thermodynamics-grounded binding affinity predictor that leverages both sequence and structural information, providing reliable reward signals without prohibitive computational cost. Our contributions are three-fold:

- We present EvoAb, a unified framework for affinity-guided CDR-H3 optimization that combines antibody foundation models with structure-aware reward modeling, enabling efficient *in silico* directed evolution without physics-based energy calculations or wet-lab experiments.
- We introduce Multi-round Mutation Policy Optimization (M2PO), an iterative algorithm that integrates affinity-weighted sequence learning with distribution anchoring, progressively enhancing binding affinity while preserving the structural plausibility of generated antibodies.
- Extensive experiments demonstrate that EvoAb consistently generates CDR-H3 variants with significantly improved binding profiles. We further validate that EvoAb-designed antibodies achieve strong performance under independent physics-based evaluation, confirming the generalization of our optimization beyond the training reward model.

## 2. Related Work

### 2.1. Antibody Design Models

Early CDR-H3 design relied on physics-based methods including RosettaAntibody (Weitzner et al., 2017), OptCDR (Pantazes & Maranas, 2010), Next-Generation Sequencing (NGS) (Ravn et al., 2010)), while initial learning-based approaches focused on predicting CDR structures DeepH3 (Ruffolo et al., 2020), Abacus (Powles et al., 2019). With advances in protein language models and structure prediction AlphaFold2 (Jumper et al., 2021), AI-driven methods emerged to design CDRs from both sequence and structural perspectives IgDesign (Shanehsazzadeh et al., 2023), AntiBERTa (Leem et al., 2022). Then, GNN-based generative models have enabled faster, CDR-specific design with parallel design of sequence and structure modalities RefineGNN (Jin et al., 2021), DiffAb (Luo et al., 2022), ImmuneBuilder (Abanades et al., 2023), ABLooper (Abanades et al., 2022), MEAN (Kong et al., 2022), and H3-OPT (Chen et al., 2023).

Beyond unconditional generation, epitope-specific methods such as RFDiffusion Antibody (Bennett et al., 2025), dyMEAN (Kong et al., 2023), UniMoMo (Kong et al., 2025) achieve end-to-end co-design with atomic precision. AlphaFold-series models, including AlphaFold 3 (Abramson et al., 2024), Boltz-2 (Passaro et al., 2025), SeedFold (Yi et al., 2025), and Proteo-R1 (Wu et al., 2026), further account for all-atom environments and dynamic conformational transitions, enabling more accurate complex structure prediction.

### 2.2. Binding Affinity Prediction

Traditional physics-based tools like Rosetta (Rohl et al., 2004), and FoldX (Schymkowitz et al., 2005) estimate binding affinity through atomic energy calculations. With the growth of deep learning and large-scale antibody databases, three categories of learning-based methods have emerged: (1) *Sequence-based models* fine-tune pretrained protein language models on antibody-antigen binding data or mutation outcomes, such as AlphaBind (Agarwal et al., 2025), AttABSeq (Jin et al., 2024), ProtAttBA (Liu et al., 2025). (2) *Structure-based models* represent complexes as graphs to capture precise antigen-CDR interactions. Methods such as Graphinity (Hummer et al., 2025), GearBind (Cai et al., 2024), and light-DDG (Wu et al., 2025) identify physical

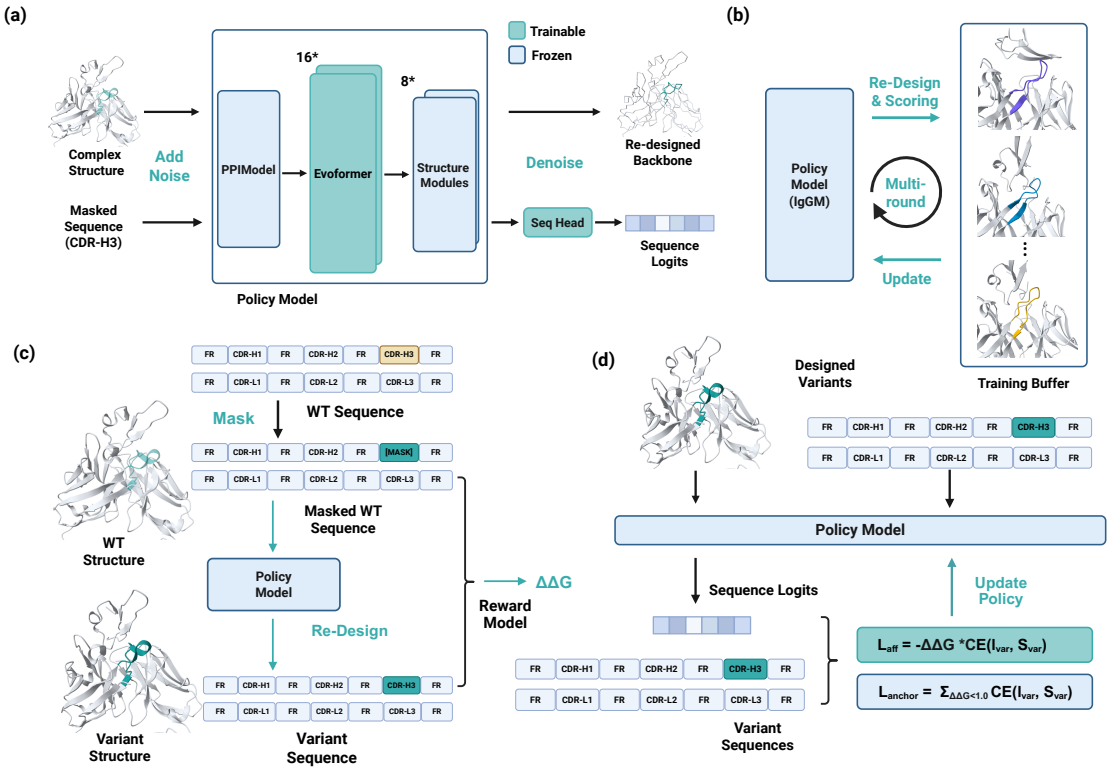


*Figure 2.* **Overview of the EvoAb framework. (a)** Policy model architecture. The model takes antibody-antigen structure and masked CDR sequence as input, jointly redesigning both structure and sequence through iterative denoising. The PPI module encodes complex-level information, which is then processed by the Evoformer and Structure Module to generate the output. During fine-tuning, selected Evoformer layers and the Sequence Head are unfrozen. **(b)** M2PO multi-round offline fine-tuning. The policy is iteratively updated using samples from the training buffer at each round. **(c)** CDR-H3 redesign. Conditioned on wild-type (WT) structure and masked sequence, the policy generates new CDR-H3 variants, which are scored by the reward model to obtain affinity change predictions ($\Delta\Delta G$). **(d)** Policy update. The training loss is computed from output logits and variant sequences, directly updating the policy parameters.

interactions between spatially proximal residues and locate key interfacial hotspots affected by mutations. (3) *Physics-informed models* like StaB-ddG (Deng et al., 2025) leverage thermodynamic parameterization to reformulate binding energy prediction as folding energy differences, enabling transfer learning from abundant protein stability data while ensuring physical antisymmetry and mutational path independence. These learning-based approaches offer faster inference and improved zero-shot scoring, facilitating more efficient antibody design iterations.

## 3. Method

We propose a unified framework for antibody directed evolution that optimizes CDR sequences for enhanced binding affinity. Our approach integrates an antibody co-design foundation model, a thermodynamics-grounded binding affinity predictor, and an offline reinforcement learning method for discrete diffusion models. An overview of EvoAb is illustrated in Figure 2.

### 3.1. Problem Formulation

Given an antibody-antigen complex with backbone structure $X_{\text{complex}}$ and a wild-type antibody sequence $S_{\text{WT}}$, our goal is to redesign the CDR-H3 region to enhance binding affinity. Let $S_{\text{WT}} = [S_{\text{FR}}; S^{\text{WT}}_{\text{CDR-H3}}; S_{\text{other}}]$, where $S_{\text{FR}}$ denotes the framework regions, $S^{\text{WT}}_{\text{CDR-H3}}$ is the wild-type CDR-H3 sequence, and $S_{\text{other}}$ represents the remaining CDR regions (H1, H2, L1, L2, L3). All regions except CDR-H3 remain fixed during optimization.

We formulate this as a conditional sequence optimization problem. The objective is to generate a mutant CDR-H3 sequence $S^{\text{Mut}}_{\text{CDR-H3}}$ that minimizes the predicted binding free energy change:

$$\min_{S^{\text{Mut}}_{\text{CDR-H3}}} \Delta\Delta G = g_\phi(S_{\text{WT}}, S_{\text{Mut}}, X_{\text{complex}}), \qquad (1)$$

where $g_\phi$ is a learned $\Delta\Delta G$ predictor (in kcal/mol), $S_{\text{Mut}} = [S_{\text{FR}}; S^{\text{Mut}}_{\text{CDR-H3}}; S_{\text{other}}]$ denotes the mutant antibody sequence, and negative $\Delta\Delta G$ indicates improved binding affinity relative to the wild-type.

### 3.2. Sequence-Structure Co-Design via IgGM

We employ IgGM (Wang et al., 2025) as our base generative model for CDR-H3 redesign. IgGM performs joint sequence-structure co-design using an SE(3)-equivariant graph neural network, where each residue $i$ is represented by its amino acid type $s_i$, $C_\alpha$ coordinate $x_i \in \mathbb{R}^3$, and orientation $O_i \in SO(3)$.

Given the complex structure and a masked CDR-H3 region, we define the conditioning context as $c \equiv (X_{\text{complex}}, S_{\text{FR}}, S_{\text{other}})$. IgGM generates the CDR-H3 sequence and structure conditioned on this context:

$$S_{\text{CDR-H3}}, X_{\text{CDR-H3}} = \text{IgGM}_\theta(c). \tag{2}$$

**Discrete Diffusion Process.** Let $s_0 \in \{1, \dots, 20\}^L$ denote the clean CDR-H3 sequence of length $L$, with each position represented as a one-hot row vector. IgGM defines a forward noising process via time-indexed transition matrices $Q_t \in \mathbb{R}^{20\times 20}$:

$$q(s_t \mid s_{t-1}) = \text{Cat}(s_t;\, p = s_{t-1}Q_t), \tag{3}$$

applied independently across positions. The marginal distribution at timestep $t$ is:

$$q(s_t \mid s_0) = \text{Cat}(s_t;\, p = s_0\bar{Q}_t), \quad \bar{Q}_t = \prod_{k=1}^{t} Q_k. \tag{4}$$

The reverse denoising process is parameterized by logits $l_\theta(s_t, c) \in \mathbb{R}^{L\times 20}$:

$$\pi_\theta(s_0 \mid s_t, c) = \text{Cat}(s_0;\, \text{softmax}(l_\theta(s_t, c)))\,. \tag{5}$$

Sampling proceeds via ancestral denoising: starting from $s_T$ drawn from the terminal distribution, for $t = T, \dots, 1$, the model predicts $\hat{s}_0 \sim \pi_\theta(s_0 \mid s_t, c)$ and samples $s_{t-1} \sim q(s_{t-1} \mid s_t, \hat{s}_0)$. We overload $\pi_\theta$ to denote both the per-step denoising distribution and the induced generative policy over complete sequences.

**Trainable Parameters.** We initialize from pretrained IgGM weights and fine-tune only a subset of parameters to balance adaptation capacity with computational efficiency and prevent overfitting. In IgGM's architecture, the Evoformer captures complex sequence-structure dependencies through iterative attention mechanisms, while the Sequence Head (LM head) directly produces amino acid logits for sequence generation. In contrast, the Structure Module primarily refines geometric coordinates through SE(3)-equivariant transformations, which are less directly coupled to sequence prediction (Wang et al., 2025). Based on this architectural insight, we selectively unfreeze partial Evoformer layers and the Sequence Head (the components most directly responsible for sequence generation), while keeping the Structure Module frozen to preserve learned geometric priors.

### 3.3. Thermodynamic Reward Modeling

To provide reliable reward signals for reinforcement learning, we employ StaB-ddG (Deng et al., 2025) as the reward model. StaB-ddG is grounded in a fundamental thermodynamic identity that decomposes binding energy into folding energies:

$$\Delta G_{\text{bind}} = \Delta G_{\text{fold}}^{\text{complex}} - \Delta G_{\text{fold}}^{\text{Ab}} - \Delta G_{\text{fold}}^{\text{Ag}}. \tag{6}$$

The model leverages a pretrained inverse folding network (ProteinMPNN (Dauparas et al., 2022)) to estimate folding energies via sequence log-likelihoods, providing strong generalization to novel interfaces. The reward model $g_\phi$ takes as input the wild-type sequence, the mutant sequence, and the complex backbone structure to compute $\Delta\Delta G$ as defined in Eq. (1). The parameters $\phi$ remain frozen throughout training.

### 3.4. Multi-round Mutation Policy Optimization (M2PO)

We propose Multi-round Mutation Policy Optimization (M2PO), an iterative policy refinement algorithm for discrete sequence generation. M2PO progressively improves the generative policy $\pi_\theta$ to produce CDR-H3 sequences with enhanced binding affinity while maintaining structural plausibility.

#### 3.4.1. Affinity-Weighted Sequence Learning

To steer the policy toward generating high-affinity sequences, we introduce an affinity-weighted learning objective. The key insight is that sequences with stronger predicted binding affinity should contribute more to the policy update, analogous to how advantage-weighted methods prioritize high-reward samples in reinforcement learning (Peng et al., 2019). Unlike standard AWR which uses exponentiated advantages $\exp(A/\beta)$ as weights, we directly use the raw affinity signal $w(s_0) = -\Delta\Delta G(s_0)$ for numerical stability, as exponentiated $\Delta\Delta G$ values can lead to extremely large weights that destabilize training.

Given a generated sequence $s_0$ with predicted $\Delta\Delta G$, we define the affinity-weighted loss as:

$$\mathcal{L}_{\text{aff}}(\theta) = \mathbb{E}_{s_0\sim\pi_\theta,\, t\sim\mathcal{U}[1,T],\, s_t\sim q(\cdot|s_0)}\Big[w(s_0) \cdot \mathcal{L}_{\text{CE}}(\pi_\theta, s_0, s_t, c)\Big], \tag{7}$$

where $\mathcal{L}_{\text{CE}}$ denotes the cross-entropy loss over CDR-H3 positions:

$$\mathcal{L}_{\text{CE}}(\pi_\theta, s_0, s_t, c) = -\sum_{i=1}^{L} \log \pi_\theta(s_0^{(i)} \mid s_t, c). \tag{8}$$

Since negative $\Delta\Delta G$ indicates improved binding, sequences with superior affinity receive positive weights and are rein-

forced, effectively guiding the policy toward high-affinity regions of sequence space. Conversely, sequences with positive $\Delta\Delta G$ (reduced affinity) receive negative weights, which actively push the policy away from these unfavorable regions. This bidirectional weighting scheme provides a stronger learning signal than approaches that only reinforce good samples, as it explicitly discourages generation of low-affinity sequences.

#### 3.4.2. Distribution Anchoring

Standard policy optimization methods often employ KL divergence regularization against a reference policy to prevent distribution drift (Schulman et al., 2017). However, this requires maintaining a separate reference model during training, incurring substantial memory and computational overhead. To address this, we introduce a lightweight alternative: *distribution anchoring*, which regularizes the policy by reinforcing generation of high-quality samples without requiring explicit distributional constraints.

Specifically, we apply an unweighted reconstruction loss to sequences with $\Delta\Delta G < \tau$ (i.e., improved or near-wild-type affinity):

$$\mathcal{L}_{\text{anchor}}(\theta) = \mathbb{E}_{s_0\sim\pi_\theta,\, t\sim\mathcal{U}[1,T],\, s_t\sim q(\cdot|s_0)} \Big[\mathbf{1}[\Delta\Delta G(s_0) < \tau] \cdot \mathcal{L}_{\text{CE}}(\pi_\theta, s_0, s_t, c)\Big], \tag{9}$$

where $\tau$ is a threshold set to 1.0 kcal/mol in our experiments. This term serves three purposes: (1) it prevents mode collapse by ensuring the policy retains the capability to generate structurally valid sequences; (2) it anchors learning to the distribution of improved binders, effectively filtering out low-quality samples from the regularization signal; and (3) by using the same cross-entropy loss form as $\mathcal{L}_{\text{aff}}$, the two terms are naturally aligned in scale, avoiding the need to carefully tune a separate coefficient for heterogeneous regularization terms such as KL divergence (Wu et al., 2019; Peters et al., 2010).

#### 3.4.3. Training Objective

The overall training objective combines affinity-weighted learning with distribution anchoring:

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{aff}}(\theta) + \lambda\mathcal{L}_{\text{anchor}}(\theta), \tag{10}$$

where $\lambda$ balances the two components. The affinity-weighted term drives the policy toward high-affinity sequences by maximizing expected binding improvement, while the anchoring term acts as implicit regularization by preserving generative capacity on validated high-quality samples. Note that samples satisfying $\Delta\Delta G < \tau$ contribute to both terms, as such samples should both be prioritized (via $\mathcal{L}_{\text{aff}}$) and serve as anchors for stable learning (via $\mathcal{L}_{\text{anchor}}$).

**Algorithm 1** Multi-round Mutation Policy Optimization

**Require:** Policy $\pi_\theta$ (IgGM), reward model $g_\phi$ (StaB-ddG), training set $\mathcal{D}$
**Require:** Rounds $M$, samples per complex $N$, threshold $\tau$, weight $\lambda$, learning rate $\eta$
**for** $m = 1$ to $M$ **do**
  Initialize batch buffer $\mathcal{B} \leftarrow \emptyset$
  **for all** $(c, S_{\text{WT}}) \in \mathcal{D}$ ▷ $c = (X_{\text{complex}}, S_{\text{FR}}, S_{\text{other}})$ **do**
    **for** $j = 1$ to $N$ **do**
      Sample $S^{(j)}_{\text{CDR-H3}}, X^{(j)}_{\text{CDR-H3}} \sim \pi_\theta(\cdot \mid c)$
      Compute $\Delta\Delta G^{(j)} = g_\phi(S_{\text{WT}}, S^{(j)}_{\text{Mut}}, X_{\text{complex}})$
      $\mathcal{B} \leftarrow \mathcal{B} \cup \{(c, S^{(j)}_{\text{CDR-H3}}, \Delta\Delta G^{(j)})\}$
    **end for**
  **end for**
  **for all** $(c, s_0, \Delta\Delta G) \in \mathcal{B}$ **do**
    Sample $t \sim \mathcal{U}[1, T]$, $s_t \sim q(s_t \mid s_0)$
  **end for**
  Compute $\mathcal{L}_{\text{aff}}$ over $\mathcal{B}$ with weights $w = -\Delta\Delta G$
  Compute $\mathcal{L}_{\text{anchor}}$ over $\{(c, s_0, s_t) \in \mathcal{B} : \Delta\Delta G < \tau\}$
  Update $\theta \leftarrow \theta - \eta\nabla_\theta(\mathcal{L}_{\text{aff}} + \lambda\mathcal{L}_{\text{anchor}})$
**end for**
**return** Optimized policy $\pi_\theta$

#### 3.4.4. Multi-round Optimization

A core challenge of offline reinforcement learning is that training with a fixed dataset can be affected by distribution shifts and out-of-distribution (OOD) actions (Levine et al., 2020), which can severely degrade policy performance. To mitigate this issue, we employ an iterative data collection and policy optimization scheme, interleaving offline training with limited on-policy data acquisition. Specifically, M2PO operates through iterative rounds of generation and policy refinement, as summarized in Algorithm 1. In each round, CDR-H3 sequences are sampled from the current policy $\pi_\theta$, evaluated by the frozen reward model $g_\phi$, and used to update the policy parameters via Eq. (10). Through successive rounds, the policy distribution progressively shifts toward generating high-affinity CDR-H3 sequences while maintaining structural validity.

M2PO can be interpreted as a conservative offline-to-online fine-tuning approach (Nair et al., 2020). By progressively integrating on-policy data generated from the evolving policy, M2PO reduces the distribution mismatch between the behavior policy and the learned policy. This hybrid paradigm has been shown to effectively alleviate OOD issues in offline RL and improve policy robustness (Lee et al., 2022; Mark et al.).

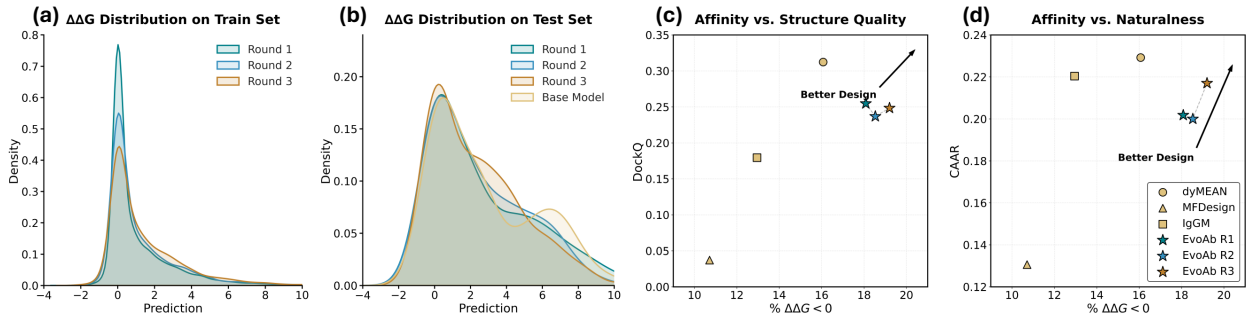


*Figure 3.* **Detailed analysis of multi-round optimization. (a-b)** $\Delta\Delta G$ distribution across training and test sets over multiple optimization rounds. The distribution progressively shifts toward lower (more favorable) values. **(c-d)** Pareto front trade-offs between binding affinity ($\%\Delta\Delta G < 0$) and design quality metrics (DockQ, CAAR) for all baselines and EvoAb across optimization rounds. EvoAb progressively advances toward the Pareto front.

*Table 1.* **Quantitative evaluation of antibody design methods.** Metrics span three dimensions: sequence recovery (AAR, CAAR), structural accuracy (RMSD, DockQ, %DockQ $> 0.23$), and binding affinity ($\Delta\Delta G$, $\%\Delta\Delta G < 0$). **Bold**: best; underline: second best.

| | Sequence | | Structure | | Binding | |
|---|---|---|---|---|---|---|
| Method | AAR ↑ | CAAR ↑ | RMSD ↓ | DockQ ↑ | $\Delta\Delta G$ ↓ | $\%\Delta\Delta G < 0$ |
| dyMEAN | 0.390 | **0.229** | 2.45 | **0.312** | 2.56 | 16.07 |
| MFDesign | 0.305 | 0.131 | 2.67 | 0.037 | 5.57 | 10.71 |
| IgGM | **0.419** | 0.220 | **2.42** | 0.179 | 2.89 | 12.95 |
| **EvoAb** | 0.401 | 0.217 | 2.65 | 0.247 | **2.45** | **19.20** |

## 4. Experiments

In this section, we systematically evaluate the effectiveness of M2PO for antibody redesign. Section 4.1 describes the experimental setup and baseline methods. Section 4.2 presents the main benchmark comparison and demonstrates EvoAb's improvements in binding affinity optimization. Section 4.3 provides a detailed analysis of the multi-round optimization process, examining the trade-offs and correlations between $\Delta\Delta G$ and sequence-structure metrics. Section 4.4 showcases representative cases that are progressively optimized across rounds. Finally, Section 4.5 investigates the impact of different loss configurations on antibody optimization.

### 4.1. Experimental Settings

**Dataset.** We curate our dataset from experimentally resolved antibody-antigen complexes in the SAbDab database (Dunbar et al., 2014). Following the temporal split protocol of IgGM (Wang et al., 2025), structures deposited before December 31, 2022 are used for training, those from January 1 to June 30, 2023 for validation, and those from July 1 to December 31, 2023 for testing. To eliminate redundancy and prevent data leakage, we apply CD-HIT (Fu et al., 2012) clustering at 95% sequence identity on concatenated heavy and light chain sequences, retaining only cluster representatives. For benchmark consistency, we adopt SAb-23H2-Ab from IgGM as our test set, ensuring no temporal overlap with comparative baselines. This yields 2,366 / 267 / 60 complexes for training / validation / testing, respectively.

During multi-round optimization, we generate 8 CDR-H3 variants per complex at each round using the corresponding model checkpoint. Each variant is annotated with $\Delta\Delta G$ predictions from StaB-ddG, yielding 18,928 / 2,136 / 480 samples per round for training / validation / testing.

**Baselines and Metrics.** We compare EvoAb against state-of-the-art antibody design models with CDR redesign capabilities: dyMEAN (Kong et al., 2023), MF-Design (Yang et al., 2025), and IgGM (Wang et al., 2025). Note that IgGM serves as both a baseline and the foundation model for EvoAb, allowing us to directly assess the benefits of our affinity-guided fine-tuning. Design quality is evaluated across three dimensions: (1) *Sequence*: amino acid recovery (AAR) and contact amino acid recovery (CAAR); (2) *Structure*: $C_\alpha$ RMSD, interface quality (DockQ), and acceptable rate (%DockQ $> 0.23$); and (3) *Binding*: mean predicted $\Delta\Delta G$ from StaB-ddG and stabilizing mutation rate ($\%\Delta\Delta G < 0$).

### 4.2. EvoAb Explores Low-Energy Landscapes

We first compare EvoAb against baseline methods on the CDR-H3 redesign task, evaluating sequence recovery, structural accuracy, and binding affinity on the test set. As shown in Table 1, all baselines achieve comparable performance on sequence recovery (AAR) and backbone RMSD, suggesting that current methods can reliably reconstruct CDR-H3 sequences and structures. For interface quality, dyMEAN achieves the highest DockQ score (0.31), substantially outperforming IgGM (0.18) and MF-Design (0.04). In terms of binding affinity, dyMEAN also exhibits a more favorable $\Delta\Delta G$ distribution with a higher proportion of stabilizing mutations (16.07%) among existing methods.

EvoAb demonstrates substantial improvements in binding

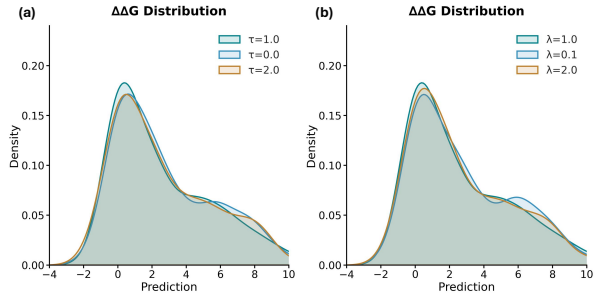


*Figure 4.* **Distribution of $\Delta\Delta G$ under different loss configurations.** **(a)** Effect of varying anchor loss cutoff threshold $\tau \in \{0.0, 1.0, 2.0\}$ kcal/mol. **(b)** Effect of varying anchor loss weight $\lambda \in \{0.1, 1.0, 2.0\}$.

affinity through multi-round offline fine-tuning. As shown in Figure 3, over successive optimization rounds, the proportion of stabilizing mutations increases from 13% to 19%. Notably, EvoAb surpasses all baselines including dyMEAN in binding affinity metrics, achieving state-of-the-art performance and directly validating the effectiveness of iterative optimization. Moreover, although sequence recovery and RMSD show slight degradation compared to baselines, interface quality improves considerably (DockQ: 0.18 $\rightarrow$ 0.25), approaching the design quality of dyMEAN. This observation suggests that by explicitly optimizing for binding affinity, the model implicitly discovers energetically favorable complex conformations rather than simply minimizing spatial deviation from the reference structure.

*Table 2.* **Ablation study on loss configurations.** Design performance of different anchor loss weight $\lambda$ and cutoff threshold $\tau$. **Bold**: best; underline: second best.

| | Sequence | | Structure | | Binding | |
|---|---|---|---|---|---|---|
| Settings | AAR ↑ | CAAR ↑ | RMSD ↓ | DockQ ↑ | $\Delta\Delta G$ ↓ | $\%\Delta\Delta G < 0$ |
| $\lambda = 0.1$ | 0.386 | 0.193 | 2.50 | 0.252 | 2.91 | 13.39 |
| $\lambda = 2.0$ | **0.409** | **0.217** | **2.46** | **0.268** | 2.76 | 15.18 |
| $\tau = 0.0$ | 0.395 | 0.200 | 2.49 | 0.251 | 2.78 | 13.17 |
| $\tau = 2.0$ | 0.405 | 0.210 | 2.49 | 0.251 | 2.78 | 16.29 |
| **EvoAb** | 0.401 | **0.217** | 2.65 | 0.247 | **2.45** | **19.20** |

### 4.3. Understanding Iterative Refinement Dynamics

To gain deeper insight into how multi-round optimization improves binding affinity, we analyze the evolution of $\Delta\Delta G$ distributions and design trade-offs across rounds.

**Distribution Shift Toward Low-Energy Regions.** As shown in Figure 3 (a-b), we observe several notable trends in the $\Delta\Delta G$ distribution: (1) As training progresses, while the overall range of $\Delta\Delta G$ remains comparable, the proportion of no-mutation cases (i.e., sequences identical to wild-type) gradually decreases, indicating that the model continuously explores novel mutation combinations through affinity-guided optimization. (2) The $\Delta\Delta G$ distribution progressively shifts toward lower values. Specifically, the proportion of samples with $\Delta\Delta G > 5.0$ kcal/mol decreases substantially, with most samples moving below this threshold. (3) Concurrently, the proportion of stabilizing mutations increases while the minimum $\Delta\Delta G$ continues to decrease, indicating that the model consistently discovers more effective mutations across rounds. These observations demonstrate that M2PO does not merely optimize within a narrow distributional regime, but rather explores the broader low-energy landscape of sequence space.

**Trade-offs Between Affinity and Design Quality.** Multi-round optimization involves inherent trade-offs between binding affinity and other design objectives. As shown in Figure 3 (c-d), when comparing the stabilizing mutation rate ($\%\Delta\Delta G < 0$) against structural quality (DockQ) and sequence naturalness (CAAR), both MF-Design and IgGM are Pareto-dominated by dyMEAN. However, through successive rounds of optimization, EvoAb progressively shifts the trade-off frontier toward more favorable regions. By Round 3, EvoAb approaches and eventually surpasses dyMEAN along the Pareto front, demonstrating that our iterative refinement strategy achieves superior binding affinity without sacrificing structural plausibility.

### 4.4. Case Study: What Does EvoAb Learn During RL?

We select two representative antibody-antigen complexes (PDB: `8D4R`, `8IX3`) to visualize mutation patterns across optimization rounds. To validate that M2PO yields genuine physical energy improvements, we also report FoldX-predicted $\Delta\Delta G$, which serves as an independent physics-based oracle not used during training.

As shown in Figure 5, the base model already exhibits position-specific mutation preferences, and through multi-round optimization, the model progressively refines its mutation strategy. For `8D4R`, mutations concentrate on positions H106 and H108. Across rounds, the model learns to refine H106 from V$\rightarrow$I to V$\rightarrow$M, while dropping the S$\rightarrow$A mutation at H108, ultimately converging to a single optimized substitution. The $\Delta\Delta G$ (StaB-ddG) decreases from 0.03 to $-0.65$ kcal/mol, with FoldX showing a consistent trend (5.84 $\rightarrow$ $-7.73$ kcal/mol).

For `8IX3`, mutations are distributed across positions H99–H114. Several substitutions remain conserved (S$\rightarrow$R at H100, I$\rightarrow$V at H109, Y$\rightarrow$I at H114), while others are refined: N$\rightarrow$D to N$\rightarrow$H at H101, L$\rightarrow$Y to L$\rightarrow$F at H108, and N$\rightarrow$G to N$\rightarrow$W at H111. The $\Delta\Delta G$ (StaB-ddG) decreases from $-0.11$ to $-0.89$ kcal/mol, with FoldX improving from 11.71 to $-3.06$ kcal/mol. Notably, both predictors show consistent trends, validating that M2PO's optimization translates to physically meaningful improvements.

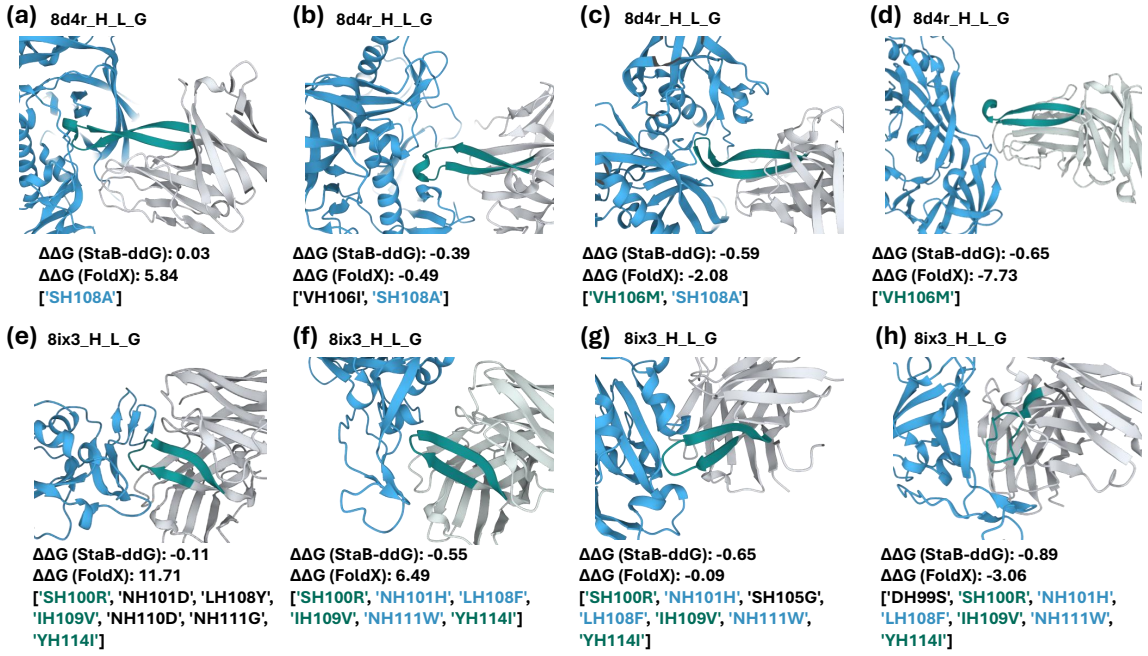


*Figure 5.* **Case Studies for `8d4r` and `8ix3`.** The mutation code represents {WT Amino Acid}{Chain}{Index}{Mutated Amino Acid}.**(a-d)** `8d4r` cases for base model, round 1, round 2, and round 3. **(e-h)** `8ix3` cases for base model, round 1, round 2, and round 3.

### 4.5. Ablations on EvoAb

We investigate the impact of loss configurations on antibody redesign, specifically the anchor loss weight $\lambda$ and the anchor loss cutoff threshold $\tau$. These two hyperparameters control the strength and scope of distribution anchoring, balancing affinity optimization with sequence plausibility. To ensure a controlled comparison, all experiments are conducted with a single round of optimization.

**Anchor Loss Weight.** The anchor loss is designed to prevent reward hacking by regularizing toward high-quality samples. As shown in Table 2 and Figure 4 (b), increasing $\lambda$ leads to higher sequence recovery, lower RMSD, and improved conformational quality (DockQ: $0.25 \rightarrow 0.27$), indicating stronger preservation of natural antibody characteristics. However, a larger anchor weight stabilizes training but limits exploration of low-$\Delta\Delta G$ regions, while a smaller weight ($\lambda = 0.1$) enables aggressive exploration but risks reward hacking. We find that $\lambda = 1.0$ provides an effective balance, achieving the best binding affinity while maintaining reasonable sequence and structural quality.

**Anchor Loss Cutoff Threshold.** The cutoff threshold $\tau$ determines which samples contribute to the anchoring term. We evaluate $\tau \in \{0.0, 1.0, 2.0\}$ kcal/mol, corresponding approximately to the first quartile (Q1), median, and third quartile (Q3) of the training set $\Delta\Delta G$ distribution. As shown in Table 2 and Figure 4 (a), a larger threshold ($\tau = 2.0$) retains a broader distribution, resulting in AAR, CAAR, and RMSD values closer to the original distribution. However, from the perspective of $\Delta\Delta G$ optimization, anchoring to a smaller subset ($\tau = 0.0$) neglects the high-$\Delta\Delta G$ sub-distribution, whereas a larger subset ($\tau = 2.0$) dilutes the optimization signal for low-$\Delta\Delta G$ samples. The default setting $\tau = 1.0$ achieves the best binding affinity by focusing on genuinely improved sequences while maintaining sufficient diversity for stable training.

## 5. Conclusion

We presented EvoAb, a RL framework for high-affinity antibody design that couples a pretrained co-design model with thermodynamics-grounded reward modeling. Our M2PO algorithm combines affinity-weighted learning with distribution anchoring, achieving state-of-the-art binding performance while preserving structural plausibility. Future directions include jointly optimizing multiple CDR regions with backbone flexibility, incorporating ensemble reward models, and experimental validation through binding assays.

## Acknowledgement

We thank for insightful discussion with Authur Deng. The work described in this paper was supported by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project T45-401/22-N.

## Impact Statement

This work aims to accelerate computational antibody design for therapeutic applications, with the potential to benefit the treatment of immune-related diseases and cancer. By enabling efficient *in silico* optimization, EvoAb may reduce the cost and time required for early-stage antibody discovery, potentially democratizing access to antibody engineering capabilities. We note that all designed antibodies require rigorous experimental validation before any clinical application, and we encourage responsible use of such computational tools in conjunction with established safety protocols. This paper presents work whose goal is to advance the field of Machine Learning applied to computational biology; there are many potential societal consequences of our work, none of which we feel must be specifically highlighted here beyond those mentioned above.

## References


Abanades, B., Georges, G., Bujotzek, A., and Deane, C. M. Ablooper: fast accurate antibody cdr loop structure prediction with accuracy estimation. *Bioinformatics*, 38(7): 1877–1880, 2022.

Abanades, B., Wong, W. K., Boyles, F., Georges, G., Bujotzek, A., and Deane, C. M. Immunebuilder: Deep-learning models for predicting the structures of immune proteins. *Communications Biology*, 6(1):575, 2023.

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630 (8016):493–500, 2024.

Agarwal, A. A., Harrang, J., Noble, D., McGowan, K. L., Lange, A. W., Engelhart, E., Lahman, M. C., Adamo, J., Yu, X., Serang, O., et al. Alphabind, a domain-specific model to predict and optimize antibody–antigen binding affinity. In *MAbs*, volume 17, pp. 2534626. Taylor & Francis, 2025.

Bennett, N. R., Watson, J. L., Ragotte, R. J., Borst, A. J., See, D. L., Weidle, C., Biswas, R., Yu, Y., Shrock, E. L., Ault, R., et al. Atomically accurate de novo design of antibodies with rfdiffusion. *Nature*, pp. 1–11, 2025.

Cai, H., Zhang, Z., Wang, M., Zhong, B., Li, Q., Zhong, Y., Wu, Y., Ying, T., and Tang, J. Pretrainable geometric graph neural network for antibody affinity maturation. *Nature communications*, 15(1):7785, 2024.

Cao, H., Zhang, H., Xu, J., Zhang, Z., Shen, L., Sun, M., Liu, G., Xu, J., Li, W.-J., Ni, J., et al. From supervision to exploration: What does protein language model learn during reinforcement learning? *arXiv preprint arXiv:2510.01571*, 2025.

Chen, H., Fan, X., Zhu, S., Pei, Y., Zhang, X., Zhang, X., Liu, L., Qian, F., and Tian, B. H3-opt: Accurate prediction of cdr-h3 loop structures of antibodies with deep learning. *bioRxiv*, pp. 2023–08, 2023.

Chiu, M. L., Goulet, D. R., Teplyakov, A., and Gilliland, G. L. Antibody structure and function: the basis for engineering therapeutics. *Antibodies*, 8(4):55, 2019.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I., Courbet, A., de Haas, R. J., Bethel, N., et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378 (6615):49–56, 2022.

Deng, A., Householder, K., Wu, F., Thrun, S., Garcia, K. C., and Trippe, B. Predicting mutational effects on protein binding from folding energy. *arXiv preprint arXiv:2507.05502*, 2025.

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.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. Cd-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, 2012.

Haakenson, J. K., Huang, R., and Smider, V. V. Diversity in the cow ultralong cdr h3 antibody repertoire. *Frontiers in immunology*, 9:1262, 2018.

Hummer, A. M., Schneider, C., Chinery, L., and Deane, C. M. Investigating the volume and diversity of data needed for generalizable antibody–antigen $\Delta\Delta$G prediction. *Nature Computational Science*, 5(8):635–647, 2025.

Jeon, W. and Kim, D. Abflex: designing antibody complementarity determining regions with flexible cdr definition. *Bioinformatics*, 40(3):btae122, 2024.

Jin, R., Ye, Q., Wang, J., Cao, Z., Jiang, D., Wang, T., Kang, Y., Xu, W., Hsieh, C.-Y., and Hou, T. Attabseq: an attention-based deep learning prediction method for antigen–antibody binding affinity changes based on protein sequences. *Briefings in Bioinformatics*, 25(4): bbae304, 2024.

Jin, W., Wohlwend, J., Barzilay, R., and Jaakkola, T. Iterative refinement graph neural network for antibody sequence-structure co-design. *arXiv preprint arXiv:2110.04624*, 2021.

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. *arXiv preprint arXiv:2208.06073*, 2022.

Kong, X., Huang, W., and Liu, Y. End-to-end full-atom antibody design. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 17409–17429, 2023.

Kong, X., Zhang, Z., Zhang, Z., Jiao, R., Ma, J., Huang, W., Liu, K., and Liu, Y. Unimomo: Unified generative modeling of 3d molecules for de novo binder design. *arXiv preprint arXiv:2503.19300*, 2025.

Lee, S., Seo, Y., Lee, K., Abbeel, P., and Shin, J. Offline-to-online reinforcement learning via balanced replay and pessimistic q-ensemble. In *Conference on Robot Learning*, pp. 1702–1712. PMLR, 2022.

Leem, J., Mitchell, L. S., Farmery, J. H., Barton, J., and Galson, J. D. Deciphering the language of antibodies using self-supervised learning. *Patterns*, 3(7), 2022.

Levine, S., Kumar, A., Tucker, G., and Fu, J. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.

Liu, C., Li, M., Tan, Y., Gou, W., Fan, G., and Zhou, B. Sequence-only prediction of binding affinity changes: a robust and interpretable model for antibody engineering. *Bioinformatics*, 41(8):btaf446, 2025.

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.

Mark, M. S., Ghadirzadeh, A., Chen, X., and Finn, C. Fine-tuning offline policies with optimistic action selection. In *Deep Reinforcement Learning Workshop NeurIPS 2022*.

Nair, A., Gupta, A., Dalal, M., and Levine, S. Awac: Accelerating online reinforcement learning with offline datasets. *arXiv preprint arXiv:2006.09359*, 2020.

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.

Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Somnath, V. R., Getz, N., Portnoi, T., Roy, J., Stark, H., et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *BioRxiv*, 2025.

Paul, S., Konig, M. F., Pardoll, D. M., Bettegowda, C., Papadopoulos, N., Wright, K. M., Gabelli, S. B., Ho, M., van Elsas, A., and Zhou, S. Cancer therapy with antibodies. *Nature Reviews Cancer*, 24(6):399–426, 2024.

Peng, X. B., Kumar, A., Zhang, G., and Levine, S. Advantage-weighted regression: Simple and scalable off-policy reinforcement learning. *arXiv preprint arXiv:1910.00177*, 2019.

Peters, J., Mulling, K., and Altun, Y. Relative entropy policy search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pp. 1607–1612, 2010.

Powles, T., Kockx, M., Rodriguez-Vida, A., Duran, I., Crabb, S. J., Van Der Heijden, M. S., Szabados, B., Pous, A. F., Gravis, G., Herranz, U. A., et al. Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the abacus trial. *Nature medicine*, 25(11):1706–1714, 2019.

Rabia, L. A., Zhang, Y., Ludwig, S. D., Julian, M. C., and Tessier, P. M. Net charge of antibody complementarity-determining regions is a key predictor of specificity. *Protein Engineering, Design and Selection*, 31(11):409–418, 2018.

Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., Magistrelli, G., Farinelli, L., Kosco-Vilbois, M., and Fischer, N. By-passing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic acids research*, 38(21):e193–e193, 2010.

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. Protein structure prediction using rosetta. In *Methods in enzymology*, volume 383, pp. 66–93. Elsevier, 2004.

Ruffolo, J. A., Guerra, C., Mahajan, S. P., Sulam, J., and Gray, J. J. Geometric potentials from deep learning improve prediction of cdr h3 loop structures. *Bioinformatics*, 36(Supplement_1):i268–i275, 2020.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. The foldx web server: an online force field. *Nucleic acids research*, 33(suppl_2):W382–W388, 2005.

Scott, A. M., Wolchok, J. D., and Old, L. J. Antibody therapy of cancer. *Nature reviews cancer*, 12(4):278–287, 2012.

Shanehsazzadeh, A., Alverio, J., Kasun, G., Levine, S., Calman, I., Khan, J. A., Chung, C., Diaz, N., Luton, B. K., Tarter, Y., et al. Igdesign: In vitro validated antibody design against multiple therapeutic antigens using inverse folding. *bioRxiv*, pp. 2023–12, 2023.

Team, C. D., Boitreaud, J., Dent, J., Geisz, D., McPartlon, M., Meier, J., Qiao, Z., Rogozhnikov, A., Rollins, N., Wollenhaupt, P., et al. Zero-shot antibody design in a 24-well plate. *bioRxiv*, pp. 2025–07, 2025a.

Team, L. L., Kenlay, H., Pretorius, D., Crabbé, J., Bridgland, A., Schmon, S. M., Hilmkil, A., Vuckovic, J., Mathis, S., Matteson, T., et al. Drug-like antibodies with low immunogenicity in human panels designed with latent-x2. *arXiv preprint arXiv:2512.20263*, 2025b.

Wang, R., Wu, F., Shi, J., Song, Y., Kong, Y., Ma, J., He, B., Yan, Q., Ying, T., Zhao, P., et al. A generative foundation model for antibody design. *bioRxiv*, pp. 2025–09, 2025.

Weitzner, B. D., Jeliazkov, J. R., Lyskov, S., Marze, N., Kuroda, D., Frick, R., Adolf-Bryfogle, J., Biswas, N., Dunbrack Jr, R. L., and Gray, J. J. Modeling and docking of antibody structures with rosetta. *Nature protocols*, 12 (2):401–416, 2017.

Wen, Y., Xu, C., Hu, J. Y.-C., Ding, K., and Liu, H. Pareto-optimal energy alignment for designing nature-like antibodies. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=6wbykApw7A.

Wu, F. and Li, S. Z. A hierarchical training paradigm for antibody structure-sequence co-design. *Advances in Neural Information Processing Systems*, 36, 2024.

Wu, F., Xuan, W., Qi, H., Cao, H., Chang, H.-J., Zhou, Z., Zhao, H., Jian, M., Ma, C., Cheng, Y.-C., et al. Proteo-r1: Reasoning foundation models for de novo protein design. *arXiv preprint arXiv:2605.02937*, 2026.

Wu, L., Liu, Y., Lin, H., Huang, Y., Zhao, G., Gao, Z., and Li, S. Z. A simple yet effective $\Delta\Delta$G predictor is an unsupervised antibody optimizer and explainer. In *The Thirteenth International Conference on Learning Representations*, 2025.

Wu, Y., Tucker, G., and Nachum, O. Behavior regularized offline reinforcement learning. *arXiv preprint arXiv:1911.11361*, 2019.

Yang, N., Jiang, S., Ma, J., Wu, H., Zheng, S., Jin, W., and Yan, J. Repurposing alphafold3-like protein folding models for antibody sequence and structure co-design. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=M96edY67nS.

Yi, Z., Chan, L., Yiming, M., Wei, Q., Fei, Y., Kexin, Z., Lan, W., Minrui, G., and Quanquan, G. Seedfold: Scaling biomolecular structure prediction. *arXiv preprint arXiv:2512.24354*, 2025.