FEEDBACK-EFFICIENT ANTIBODY DESIGN WITHIN DISCRETE CDR SHAPE SPACE

Anonymous authors
Paper under double-blind review

ABSTRACT

Antibody design is typically performed in a joint space of discrete residue identities and continuous atomic coordinates. Yet complementarity-determining regions (CDR) backbones recur within a few canonical conformations (Chothia et al., 1989), suggesting that explicitly leveraging these structural patterns could make antibody design more efficient. We therefore formulate antibody design in a discrete CDR shape space constructed by clustering experimental CDR conformations by type, length and backbone similarity ($n = 772$ shapes). Over this discrete space we train a controllable generative model that, conditioned on a target epitope and a combination of six CDR shapes, produces antibodies that bind the specified epitope while realizing the prescribed conformations. We then perform feedback-guided Bayesian optimization over CDR-shape combinations at test time. Compared to direct generation with test-time scaling in continuous space, our method reduces the CDR structural deviation by $1.87\times$ under the same budget, while matching its performance with $13\times$ fewer evaluations. In short, a compact, biologically grounded search space preserves design capacity while making antibody optimization substantially more feedback-efficient.

1 INTRODUCTION

Antibodies are a major therapeutic modality, but conventional discovery through immunization and library screening offers limited direct control over the targeted epitope. Generative models provide a route to designing antibodies against predefined epitopes (Bennett et al., 2026). Neural generative models learn distributions over antibody sequences and structures to generate new designs (Luo et al., 2022; Kong et al., 2023b; Wang et al., 2025). At test time, learned generative priors can be combined with feedback to steer generation toward desired properties (Didi et al., 2026). Beyond guiding generation through learned distributions and feedback, structural regularities can also be encoded explicitly in the representation used for generation. For example, SaDiT (Mo & Li, 2026) uses discrete structural tokens to construct a compact representation of protein backbones and performs diffusion in the corresponding embedding space, providing a complementary route to efficient protein generation.

Antibodies offer a natural basis for such structural tokenization. Their complementarity-determining regions (CDRs) form well-defined structural units, and recurrent backbone conformations have been observed among CDRs of the same type and length (Chothia et al., 1989; North et al., 2011). These recurring conformations suggest a vocabulary of reusable loop-level structural tokens that specify backbone shapes without prescribing unique amino acid sequences or side-chain conformations.

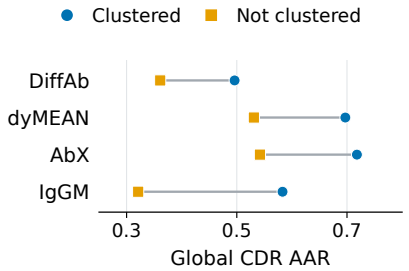


Figure 1: CDR sequence recovery by cluster membership.

To examine the relevance of this structural prior to generative antibody design, we analyze CDRs generated by four SOTA methods, grouping them by membership in recurrent backbone-shape clusters. Figure 1 shows that cluster-assigned CDRs consistently exhibit higher amino acid recovery (AAR) across all four methods, yet account for only 43.2%–59.7% of gener-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ated CDRs. This consistent association motivates making recurrent CDR shapes explicit conditions for antibody generation.

Building on this prior, we define a discrete CDR-shape space using vocabularies constructed from experimentally observed backbone conformations, grouped by CDR type and length. A paired antibody's structural specification consists of one framework token and six CDR-shape tokens (Figure 2). Within this space, we formulate epitope-specific antibody design as the selection and generative realization of CDR-shape combinations. Given an antigen, specified epitope residues, and a fixed framework, the shape tokens can be supplied as user controls or selected through property-guided optimization. The generator then realizes these specifications as antibody sequences and full-atom structures. This discrete space thus provides a compact interface for generation and feedback-guided search, with 2.5×10^8 combinations of six CDR-shape tokens on average, versus approximately 2.8×10^{62} possible sequences for 48 CDR residues (Appendix B.1). This reduction is particularly valuable when candidate evaluation is costly, as in wet-lab validation, where evaluation budgets are often limited to 10^2–10^3 candidates.

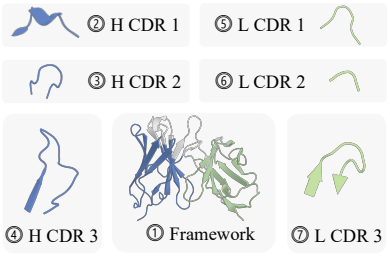


Figure 2: Seven structural tokens.

We introduce *IgSeven* to instantiate this formulation. Its generator, *Ig7G*, supports complete or partial sequence and structural prompts, enabling generation conditioned on specified CDR shapes while leaving unspecified molecular details free to adapt to the target. Its optimizer, *Ig7O*, uses Bayesian optimization over CDR-shape combinations at test time. Each proposed combination is realized by Ig7G and evaluated by external property oracles; the resulting feedback guides subsequent proposals under a limited evaluation budget. This separates the selection of structural motifs from their molecular realization, while using the same structural variables for direct user control and single- or multi-objective optimization without retraining the generator. Experiments show that Ig7G faithfully realizes specified CDR shapes, achieving 1.369 Å CDR RMSD for paired antibodies while preserving high overall design quality. Ig7O enables effective, feedback-efficient optimization, reducing the mean best CDR RMSD from 2.637 Å with continuous-space Best-of-N search to 1.413 Å under the same evaluation budget.

Our contributions are summarized as follows:

- We formulate epitope-specific antibody design as controllable generation and combinatorial optimization over discrete CDR-shape specifications, and construct an empirical shape vocabulary that makes these structural choices explicit.
- We develop Ig7G, a controllable antibody generator that supports complete or partial sequence and structural prompts, enabling faithful realization of specified CDR shapes without prescribing a docking pose.
- We develop Ig7O, a test-time Bayesian optimization framework that uses external feedback to search CDR-shape combinations for single or multiple objectives without retraining the generator.
- We experimentally validate that Ig7G accurately realizes specified CDR shapes and that Ig7O enables effective, feedback-efficient optimization under matched evaluation budgets.

2 Related Work

Existing computational antibody design methods can be broadly categorized by their generation space into sequence design, structure generation, and sequence–structure co-design (Meng et al., 2024; Vecchietti et al., 2025). **Sequence-based methods** operate in amino-acid space, typically through protein language modeling (Kuan & Farimani, 2024; Shuai et al., 2023) or structure-guided hallucination (Mille-Fragoso et al., 2026). **Structure-based methods** directly generate antibody backbones or three-dimensional geometries, as in Ig-VAE (Eguchi et al., 2022), IgDiff (Cutting et al., 2025), and RFantibody (Bennett et al., 2026), with sequences subsequently assigned through inverse folding (Dauparas et al., 2022; Høie et al., 2025). **Sequence–structure co-design** jointly models amino-acid identities and three-dimensional geometry during generation. Early antigen-conditioned approaches, including GNN-based (Kong et al., 2023a) and diffusion-based (Luo et al., 2022; Zhu et al., 2024) methods, typically design CDR sequences and structures *within a pre-positioned anti-*

body–antigen complex. More recent methods (Wang et al., 2025; Kong et al., 2023b) move toward end-to-end target-conditioned generation by jointly modeling antibody design and docking. Recent general-purpose binder design frameworks (Kong et al., 2026; 2025; Stark et al., 2025) also incorporate antibodies or nanobodies as a major design modality alongside other molecular binder classes. Beyond conventional co-design in sequence and 3D structural spaces, we introduce a CDR shape space that encodes representative CDR loop geometries as an explicit structural prior. This representation transforms unconstrained exploration of the high-dimensional structure space into a discrete, biologically grounded search over CDR conformations, providing a more tractable space for guiding sequence–structure generation. We further introduced related studies on protein structural tokenization in Appendix A.

3 PRELIMINARIES

3.1 PROBLEM FORMULATION

Antibodies consist of heavy and light chains, with six complementarity-determining regions (CDRs) forming the primary antigen-binding interface and the surrounding framework providing a relatively conserved structural scaffold. Following Bennett et al. (2026), we focus on *epitope-specific de novo antibody design* with fixed frameworks. We represent an antibody as $\mathcal{B} = (S, X)$ and its target antigen as $\mathcal{A} = (S^{\mathrm{Ag}}, X^{\mathrm{Ag}})$, where S denotes an amino acid sequence and X denotes the corresponding full-atom coordinates. Given a target antigen $\mathcal{A}$, specified epitope residues E, and a framework template $\mathcal{F} = (S^{\mathrm{F}}, X^{\mathrm{F}})$, our goal is to model

$$p_\theta(\mathcal{B} \mid \mathcal{A}, E, \mathcal{F}) = p_\theta(S, X \mid \mathcal{A}, E, \mathcal{F}). \tag{1}$$

The model jointly generates CDR sequences and structures, as well as the antibody–antigen docking pose, while preserving the framework sequence and internal geometry. The coordinates X are expressed in the antigen frame and thus encode both antibody conformation and docking pose. Following previous studies (Bennett et al., 2026; Team & Tang, 2025; Mille-Fragoso et al., 2026; Team et al., 2025b; Martinkus et al., 2023), we use *trastuzumab* and *h_NbBCII10* as framework templates for paired antibodies and nanobodies, respectively, as detailed in Appendix B.2.

3.2 CANONICAL CDR BACKBONE CONFORMATIONS

Despite their substantial sequence diversity, CDRs often adopt recurrent backbone conformations (Chothia et al., 1989; Chothia & Lesk, 1987). Classical studies therefore grouped CDR loops of the same type and length into a limited set of canonical structural classes based on backbone geometry (North et al., 2011; Adolf-Bryfogle et al., 2015; Kelow et al., 2022).

This structural regularity motivates an empirical vocabulary of CDR backbone motifs. These motifs provide reusable structural conditions for antibody generation while allowing variation in amino acid sequences and side-chain conformations, as illustrated in Appendix B.1.

CDR shape space. We collect CDR loops from SAbDab2 (Capel et al., 2026) and partition them into six CDR types $r \in \{\mathrm{H1}, \mathrm{H2}, \mathrm{H3}, \mathrm{L1}, \mathrm{L2}, \mathrm{L3}\}$. For each type, we compute pairwise backbone RMSD between loops of the same length and apply DBSCAN clustering to identify recurrent backbone conformations. The resulting clusters define six CDR-specific shape vocabularies $\mathcal{V}_r$, containing 130, 75, 370, 73, 10, and 114 shape tokens for H1, H2, H3, L1, L2, and L3, respectively. Each token is represented by its cluster medoid and specifies a characteristic loop length and backbone conformation. Details of atlas construction are provided in Appendix B.1.

4 METHODOLOGY

IgSeven operates over a discrete antibody structural space defined by one framework token and six CDR-shape tokens. It comprises two components: *Ig7G*, which generates antibodies conditioned on structural tokens, and *Ig7O*, which uses oracle feedback to optimize CDR-shape combinations at test time. §4.1, §4.2, and §4.3 introduce the structural representation, *Ig7G*, and *Ig7O*, respectively.

4.1 Antibody Generation over a Discrete Structural Space

We specify antibody structural conditions using seven discrete tokens:

$$(t_{\mathrm{F}}, t_{\mathrm{H1}}, t_{\mathrm{H2}}, t_{\mathrm{H3}}, t_{\mathrm{L1}}, t_{\mathrm{L2}}, t_{\mathrm{L3}}).$$

Here, t_{F} indexes the framework template $\mathcal{F}$, including its sequence and structure. For each CDR type r, the token $t_r \in \mathcal{V}_r$ indexes a backbone representative with a specified loop length and shape. The CDR tokens specify structural conditions without prescribing amino acid identities or side-chain conformations. Together, these tokens induce a discrete structural design space. For a fixed framework token t_{F}, we define

$$\Omega_{\mathrm{Ab}} = \{t_{\mathrm{F}}\} \times \mathcal{V}_{\mathrm{H1}} \times \mathcal{V}_{\mathrm{H2}} \times \mathcal{V}_{\mathrm{H3}} \times \mathcal{V}_{\mathrm{L1}} \times \mathcal{V}_{\mathrm{L2}} \times \mathcal{V}_{\mathrm{L3}}, \tag{2}$$

where $\mathcal{V}_r$ denotes the shape vocabulary for CDR type r. Each point in Ω_{Ab} specifies a framework and a combination of CDR lengths and backbone shapes. Conditioned on this specification, the antigen, and the epitope, *Ig7G* generates antibody sequences and full-atom structures while jointly determining the docking pose. The tokens therefore discretize the structural specification, whereas generation proceeds in the continuous coordinate and residue-latent space. For nanobodies, the representation contains one framework token and three heavy-chain CDR tokens, with the light-chain factors omitted from Ω_{Ab}. This space supports both direct structural specification and test-time optimization over CDR-shape combinations.

4.2 *Ig7G*: Controllable Antibody Generation

To realize the structural specifications in Ω_{Ab}, we develop *Ig7G*, a controllable antibody generator that translates the framework and CDR structural tokens into antibody-side controls. Together with the target antigen and epitope conditions, these controls guide the generation of antibody sequences, full-atom structures, and antibody–antigen docking poses.

Generative representation. We build *Ig7G* on the flow-matching generative model and autoencoder of Complexa (Didi et al., 2026). An antibody with N residues is represented by its C_α backbone coordinates $\mathbf{x} \equiv X[C_\alpha] \in \mathbb{R}^{N\times 3}$ and per-residue latent features $\mathbf{z} \in \mathbb{R}^{N\times d}$, where the latent features encode amino acid identities and non-C_α atomic conformations. We denote the denoising network by $\mathcal{N}_\theta$. Given coordinate and latent states at times t_x and t_z, it predicts their velocity fields conditioned on the antigen $\mathcal{A}$ and epitope residues E:

$$(\hat{\mathbf{v}}_x, \hat{\mathbf{v}}_z) = \mathcal{N}_\theta\left(\mathbf{x}_{t_x}, \mathbf{z}_{t_z}, t_x, t_z; \mathcal{A}, E\right). \tag{3}$$

The denoiser is implemented as a Transformer with pair-biased attention. The noisy coordinates and residue latents are embedded to form residue-level features, while the coordinates additionally induce pairwise C_α distances that contribute to the pair representation used in attention.

The generative process $\mathcal{G}_\theta$ uses these predictions to iteratively transform Gaussian reference states into $(\mathbf{x}, \mathbf{z})$. The frozen Complexa decoder then maps this representation to an amino acid sequence S and full-atom coordinates X.

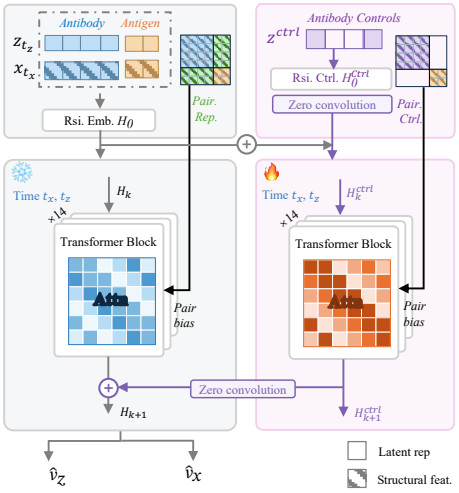


Figure 3: Antibody generation with *Ig7G*.

Antibody-specific adaptation. To specialize the pretrained protein generator to antibody design, we fine-tune $\mathcal{N}_\theta$ on antibody–antigen complexes conditioned on the antigen and epitope. A hotspot curriculum progressively increases the availability of epitope information during training. The resulting denoiser $\mathcal{N}_{\theta_{\mathrm{Ab}}}$ and generator $\mathcal{G}_{\theta_{\mathrm{Ab}}}$ provide the antibody-specific base model for structural control. The training data and adaptation settings are described in Appendices B.3 and C, respectively.

ControlNet-based antibody control. We augment $\mathcal{N}_{\theta_{\mathrm{Ab}}}$ with a trainable control branch following ControlNet (Zhang et al., 2023), while freezing the backbone parameters θ_{Ab}. The branch is parameterized by ϕ and initialized from copies of the corresponding backbone Transformer layers.

We denote the supplied antibody controls by $\mathcal{P}$. At each denoising step, the branch receives the current hidden representation augmented with residue-level controls. Its attention uses the backbone pair representation augmented with pairwise geometric controls. The control inputs are introduced through zero-initialized linear projections. Intermediate branch outputs are added to the corresponding antibody hidden states in the frozen backbone through zero-initialized residual projections. The branch therefore initially contributes no residual correction and learns to modify the denoising trajectory according to the supplied controls. We denote the controlled denoiser by $\mathcal{N}_{\theta_{\mathrm{Ab}},\phi}$.

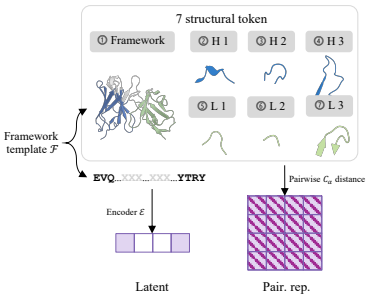


Figure 4: From tokens to controls.

From structural tokens to antibody controls. The antibody controls $\mathcal{P}$ support both structural and sequence conditioning. Structurally, the framework and six CDR tokens allow the backbone shape of each antibody region to be specified separately. For fixed-framework design, sequence conditions provide only the framework amino acid identities, leaving all CDR sequences unknown.

To construct structural controls, we obtain the backbone representative associated with each CDR token and graft it onto the framework template $\mathcal{F}$ by aligning the flanking residues. The assembled backbone serves as a geometric reference for conditioning rather than the final generated structure. We extract pairwise C_α distances within and across the framework and CDR regions and embed them as pairwise control features. These features encode both regional backbone shapes and their relative arrangement.

At framework positions, residue-level controls comprise the known amino acid identities and latent features obtained from the framework template using the frozen encoder $\mathcal{E}$. For shape-conditioned design, neither sequence nor latent controls are supplied at CDR positions. Note that absolute-coordinate controls and antibody–antigen pairwise geometry are excluded, ensuring that structural controls specify only antibody-internal geometry without fixing the docking pose. The model generates CDR sequences, full-atom antibody structures, and docking poses under the antigen, epitope, and antibody conditions:

$$(\mathbf{x}, \mathbf{z}) = \mathcal{G}_{\theta_{\mathrm{Ab}},\phi}\left(\boldsymbol{\epsilon}; \mathcal{A}, E, \mathcal{P}\right), \qquad (S, X) = \mathcal{D}(\mathbf{x}, \mathbf{z}), \tag{4}$$

where $\boldsymbol{\epsilon}$ denotes the randomness used during sampling, and $\mathcal{D}$ is the frozen Complexa decoder. Note that the control interface extends beyond CDR-shape specifications in Ω_{Ab}, accepting sequence and structural prompts directly, as validated in §5.1 and Appendix D.1.

Training objective. Training proceeds in two stages: antibody adaptation and control training. Although the two stages serve different purposes and optimize different parameter subsets, they use the same joint flow-matching objective.

For a clean training example $(\mathbf{x}_1, \mathbf{z}_1)$ extracted from a native antibody–antigen complex, we sample Gaussian reference states $(\mathbf{x}_0, \mathbf{z}_0)$ and independently sample interpolation times t_x and t_z. Following the linear interpolation path, we have $\mathbf{x}_{t_x} = (1-t_x)\mathbf{x}_0 + t_x\mathbf{x}_1$ and $\mathbf{z}_{t_z} = (1-t_z)\mathbf{z}_0 + t_z\mathbf{z}_1$. The model predicts velocity fields $\widehat{\mathbf{v}}_x$ and $\widehat{\mathbf{v}}_z$ from the interpolated states and is optimized by regressing them toward the corresponding linear-flow targets:

$$\mathcal{L} = \mathbb{E}\left[\frac{1}{N}\|\widehat{\mathbf{v}}_x - (\mathbf{x}_1 - \mathbf{x}_0)\|_{\mathrm{F}}^2 + \frac{1}{N}\|\widehat{\mathbf{v}}_z - (\mathbf{z}_1 - \mathbf{z}_0)\|_{\mathrm{F}}^2\right], \tag{5}$$

where N is the number of valid antibody residues. The expectation is taken over training examples, Gaussian reference states, interpolation times.

The two phases share the same prediction targets but differ in their conditioning inputs and trainable parameters. In the adaptation stage, we optimize the base denoiser θ_{Ab} while keeping the autoencoder fixed. The objective trains the denoiser to recover the clean antibody backbone and residue latents from their interpolated states, thereby adapting the pretrained protein generator to the

antibody–antigen distribution. In the control stage, θ_{Ab} is frozen and only the control branch ϕ is optimized. For each training complex, we construct the self-backbone control from the clean antibody backbone and the framework sequence control from the native framework sequence, with a sampled global SE(3) perturbation applied to its pose while preserving its internal geometry. The same flow-matching loss therefore trains the control branch to use the supplied structural and sequence controls to guide the frozen denoiser toward the corresponding clean antibody state. During control training, structural controls are derived from native antibody backbones. At inference, assembled token representatives are converted into the same geometric control features. Detailed training and sampling configurations are provided in Appendix C.

4.3 *Ig7O*: Test-Time Optimization in CDR Shape Space

Given a CDR-shape token combination, *Ig7G* realizes it as an antibody sequence and all-atom structure conditioned on the antigen, epitope, and framework. *Ig7O* focuses on the complementary problem of searching the CDR-shape space for combinations that optimize target properties. To this end, we perform Bayesian optimization at test time, using property feedback to guide the search over CDR-token combinations under a limited evaluation budget.

For a fixed antigen, epitope, and framework token t_{F}, we represent a CDR-shape combination as $\boldsymbol{t} = (t_{\mathrm{H1}}, t_{\mathrm{H2}}, t_{\mathrm{H3}}, t_{\mathrm{L1}}, t_{\mathrm{L2}}, t_{\mathrm{L3}})$, where each entry specifies the structural token of one CDR and $(t_{\mathrm{F}}, \boldsymbol{t}) \in \Omega_{\mathrm{Ab}}$. For nanobodies, $\boldsymbol{t}$ contains only the three heavy-chain CDR tokens. Given a proposed combination $\boldsymbol{t}$, *Ig7G* generates an antibody sequence and structure (S, X), which are evaluated by the selected property oracle(s). We denote the resulting feedback by $\boldsymbol{y}(\boldsymbol{t})$, which is scalar for single-objective optimization and vector-valued for multi-objective optimization. After n evaluations, the accumulated search history is $\mathcal{D}_n = \{(\boldsymbol{t}_i, \boldsymbol{y}(\boldsymbol{t}_i))\}_{i=1}^{n}$. *Ig7O* then fits a surrogate $\hat{\boldsymbol{f}}_n$ on $\mathcal{D}_n$ to predict the property outcomes of unevaluated CDR-shape combinations. The surrogate provides an efficient proxy for the substantially more expensive generation-and-evaluation pipeline, enabling broad exploration of the combinatorial space under a limited evaluation budget. An acquisition function a_n then prioritizes the next combination by balancing exploitation and exploration:

$$\boldsymbol{t}_{n+1} = \arg\max_{\boldsymbol{t} \in \Omega_{\mathrm{Ab}}} a_n\left(\boldsymbol{t}; \hat{\boldsymbol{f}}_n\right). \tag{6}$$

The selected combination is evaluated through *Ig7G* and the property oracle(s), and the resulting observation is incorporated into $\mathcal{D}_{n+1}$.

Search instantiations. *Ig7O* accommodates different surrogate models and acquisition functions, such as tree ensembles or Gaussian processes paired with upper confidence bound (UCB) or expected improvement (EI). We evaluate two single-objective instantiations. **Tree ensemble** uses a random-forest surrogate trained on one-hot-encoded CDR-token combinations and selects candidates using UCB, with the mean and standard deviation of individual-tree predictions providing the predicted score and an empirical uncertainty estimate, respectively. **COMBO** (Oh et al., 2019) uses a Gaussian-process surrogate with a diffusion kernel over the Cartesian product of per-CDR categorical graphs and selects combinations by optimizing EI over this graph. Both instantiations share the same structural-token interface and generation–evaluation loop, with *Ig7G* kept fixed throughout search.

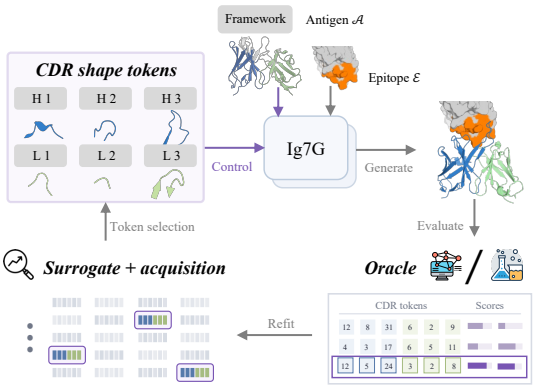


Figure 5: CDR-shape optimization with *Ig7O*.

Pareto-front search. For multiple objectives, *Ig7O* seeks to approximate the Pareto front rather than optimize a fixed aggregate score. We maintain an archive of non-dominated evaluated designs: a design is retained if no other evaluated design is at least as good on every objective and strictly better on at least one. In our multi-objective instantiation, we pair COMBO's graph-based surrogate

with expected hypervolume improvement (EHVI) (Daulton et al., 2020), which prioritizes combinations by their expected increase in the hypervolume dominated by the current frontier relative to a reference point. Each new evaluation updates both the surrogate and the archive, yielding a set of designs that represent different trade-offs among the target properties. Detailed search algorithms are introduced in Appendix E.

5 EXPERIMENTS

We evaluate *IgSeven* in terms of controllable generation and feedback-guided optimization. We first assess sequence-based control through epitope-constrained antibody–antigen co-folding in §5.1. We then examine whether *Ig7G* faithfully realizes specified CDR shapes while maintaining antibody design quality in §5.2. Finally, in §5.3, we evaluate single- and multi-objective optimization with *Ig7O*, comparing continuous-space sampling, discrete random search, and Bayesian optimization to assess the contributions of the discrete shape space and feedback-guided search. Additional feedback-efficiency analyses are provided in Appendix D.2. Detailed descriptions of baselines and evaluation metrics are deferred to Appendices D.3 and D.4, respectively.

Dataset. We use antibody–antigen complexes from SAbDab2 (Capel et al., 2026), following its predefined splits with additional filtering over date, antigen availability, and structural completeness. The resulting dataset contains 4,316 training complexes and 96 test complexes, including 71 paired antibodies and 25 nanobodies in the test set. Data preparation and split details are provided in Appendix B.3.

5.1 EPITOPED-CONSTRAINED FOLDING

Task setup. We first evaluate whether the model can use one modality to guide the generation of the other. For folding, we provide the antibody sequence as the control and generate its structure conditioned on the target antigen and specified epitope. This yields an epitope-constrained co-folding task that evaluates both antibody structural accuracy and antibody-antigen docking.

Baselines. We prioritize folding models that support epitope constraints, including Chai-1 (Chai Discovery, 2024) and Protenix-v0.5.0 (Team et al., 2025a). We also include HelixFold-S1 (Liu et al., 2026), which predicts inter-chain contacts and uses selected contacts as constraints to guide co-folding. We additionally include ESMFold2 (Candido et al., 2026) and Protenix-v2 (Zhang et al., 2026), which do not support specifying target epitopes. Note that all baselines are general-purpose folding models.

Table 1: Epitope-constrained antibody–antigen co-folding.

MSA	Epitopes	Method	Cα RMSD↓	Full RMSD↓	DockQ↑	Success Rate↑
✗	✗	ESMFold2	1.165	<u>1.715</u>	0.160	0.219
✓	✗	Protenix-v2	<u>1.150</u>	1.762	0.351	0.427
✓	✗	HelixFold-S1	1.248	1.823	<u>0.361</u>	<u>0.469</u>
✓	✓	Chai-1	1.295	1.913	0.315	0.458
✓	✓	Protenix-v0.5.0	1.227	1.856	0.329	0.427
✗	✓	*IgSeven* (Ours)	**0.595**	**0.628**	**0.643**	**0.719**

Results. Table 1 presents the results of epitope-constrained antibody–antigen co-folding, regarding RMSD and docking quality using DockQ. When conditioned on antibody sequences, *IgSeven* generates accurate antibody structures and docking poses, achieving the best performance across all reported metrics. Compared with the strongest baseline for each metric, *IgSeven* reduces Cα RMSD from 1.150 to 0.595 Å, increases DockQ from 0.361 to 0.643, and improves docking success from 46.9% to 71.9%. These results support the effectiveness of sequence-based control in *IgSeven*. Although not specifically designed for folding, the model accurately generates structures from sequence prompts, suggesting a strong alignment between the sequence and structure modalities and providing a foundation for controllable antibody design.

Conversely, controlling the antibody backbone enables *IgSeven* to perform inverse folding, generating amino-acid sequences compatible with the specified structure. *IgSeven* achieves the best average rank on sequence generation performance, as detailed in Appendix D.1.

5.2 REALIZING SPECIFIED CDR SHAPES

Task setup. We evaluate whether *Ig7G* can faithfully realize specified CDR shapes while maintaining high antibody design quality. To isolate generation quality from shape selection, we assign each CDR token to the cluster containing its experimentally observed conformation. Baselines do not use these oracle shape labels and are therefore included *only as references* for existing antibody design quality rather than as input-matched competitors. All results are reported from a single generation run. We evaluate both paired antibodies and nanobodies using the framework templates described in Appendix B.2.

Baselines. We consider antibody design methods that jointly generate antibodies and determine their antibody–antigen docking poses, including RFantibody (Bennett et al., 2026), DyMEAN (Kong et al., 2023b), and BoltzGen (Stark et al., 2025). The released DyMEAN checkpoint supports paired antibody design only. We additionally include AbX (Zhu et al., 2024) and DiffAb (Luo et al., 2022), which require a predetermined docking pose. For these two methods, we use the Kabsch algorithm to align the framework template to the antibody framework in the ground-truth antibody–antigen complex, thereby providing its position and orientation relative to the antigen.

Table 2: Realizing specified CDR shapes in fixed-framework antibody design. *Ig7G* uses oracle CDR-shape labels, whereas the baselines do not. Bold and underlined values denote the best and second-best results, respectively, among methods without a fixed docking pose.

		Paired Antibodies				Nanobodies			
Fixed Dock	Method	CDR RMSD ↓	MolProbity Score ↓	CDR AAR ↑	CDR SeqNat ↑	CDR RMSD ↓	MolProbity Score ↓	CDR AAR ↑	CDR SeqNat ↑
✓	AbX	2.937	1.313	0.447	0.491	3.451	1.192	0.396	0.764
✓	DiffAb	3.048	1.060	0.354	0.517	2.964	0.797	0.349	0.743
✗	RFantibody	3.594	1.001	0.159	0.060	<u>2.772</u>	0.874	0.220	0.383
✗	DyMEAN	<u>2.768</u>	2.417	<u>0.376</u>	**0.601**	–	–	–	–
✗	BoltzGen	3.693	<u>0.923</u>	0.288	0.386	3.333	<u>0.752</u>	<u>0.274</u>	<u>0.577</u>
✗	***Ig7G* (Oracle)**	**1.369**	**0.822**	**0.438**	<u>0.504</u>	**2.092**	**0.743**	**0.337**	**0.640**

Results. Table 2 reports the generation quality for paired antibodies and nanobodies, respectively. We evaluate structural quality using CDR RMSD and MolProbity (Chen et al., 2010), and sequence quality using CDR amino-acid recovery (AAR) and CDR SeqNat (Ramon et al., 2024). *Ig7G* faithfully realizes the specified CDR shapes while maintaining high overall design quality. For paired antibodies, *Ig7G* achieves a CDR RMSD of 1.369 Å and a MolProbity score of 0.822; for nanobodies, the corresponding values are 2.092 Å and 0.743. These are the lowest CDR RMSD and MolProbity values among the reported methods in both settings. *Ig7G* also achieves strong sequence quality, with CDR AARs of 0.438 and 0.337 and CDR SeqNat scores of 0.504 and 0.640 for paired antibodies and nanobodies, respectively. Together, these results show that *Ig7G* can precisely realize specified CDR shapes without compromising the structural or sequence quality.

5.3 FEEDBACK-GUIDED OPTIMIZATION OVER CDR SHAPE SPACE

Task setup. We evaluate whether the discrete CDR-shape space supports effective optimization through direct search over structural configurations. In test time, *Ig7O* optimizes CDR shapes by searching over shape-token combinations with external oracle feedback, while keeping *Ig7G* fixed. We instantiate *Ig7O* with a tree-ensemble surrogate and COMBO for single-objective optimization, and Pareto optimization for multiple objectives, as described in §4.3. The objectives are ΔG, CDR AAR and RMSD, reflecting interface energetics, sequence recovery, and structural fidelity. We compare *IgSeven* against Complexa θ_{Ab} in §4.2 using Best-of-N selection in its continuous latent space. We also compare *IgSeven* with random search in discrete space. For each metric in Table 3, we report the best value achieved by each method for that specific objective; the gain rows are computed accordingly on a per-metric basis. For reference, we report the best and second-best baseline performance for each metric from §5.2; these results are obtained from single-design evaluation without feedback-based optimization.

Guided by feedback from the corresponding oracle, *IgSeven* steers antibody generation toward the specified design objectives. For paired antibodies, separately optimizing computed ΔG, CDR AAR,

Table 3: Feedback-guided antibody optimization.

Method	Search Protocol	Paired Antibodies			Nanobodies		
		$\Delta G \downarrow$	CDR AAR ↑	CDR RMSD ↓	$\Delta G \downarrow$	CDR AAR ↑	CDR RMSD ↓
Best baseline		-25.763	0.376	2.768	-35.492	0.274	2.772
Second-best baseline		-18.156	0.288	3.594	-23.057	0.220	3.333
Complexa θ_{Ab}	Best-of-N	-36.143	0.442	2.764	**-42.792**	0.404	2.227
IgSeven	Random	-31.377	0.456	1.908	-29.770	0.464	2.057
	Tree ensemble	**-38.244**	0.511	1.403	-32.681	0.496	**1.830**
	COMBO	-35.311	**0.520**	**1.305**	-30.661	**0.498**	1.870
	Pareto	-33.067	0.503	1.393	-28.099	0.482	1.921
Gain vs. Complexa θ_{Ab}		+5.81%	+17.63%	+52.79%	-23.63%	+23.26%	+17.82%
Gain vs. random *Ig7O*		+21.89%	+13.99%	+31.59%	+9.78%	+7.37%	+11.02%

and CDR RMSD yields -38.244, 0.520, and 1.305 Å, respectively. For nanobodies, the corresponding values are -32.681, 0.498, and 1.830 Å. The results further highlight two complementary sources of improvement:

- **From continuous to discrete space.** Under the same base denoiser $\mathcal{N}_{\theta_{\mathrm{Ab}}}$, searching over the discrete CDR-shape space improves over Best-of-N search in the continuous latent space of Complexa θ_{Ab} on five of the six metrics. For paired antibodies, *Ig7O* improves ΔG, CDR AAR, and CDR RMSD by 5.81%, 17.63%, and 52.79%, respectively. For nanobodies, it improves CDR AAR and CDR RMSD by 23.26% and 17.82%, while Complexa remains stronger on ΔG. These results suggest that the discrete structural interface provides a more effective search space for controllable antibody generation.
- **From random search to Bayesian optimization.** Within the same discrete CDR-shape space, Bayesian optimization notably improves over random exploration. Relative to random Ig7 search, *Ig7O* improves all six metrics, with gains of 21.89%, 13.99%, and 31.59% for paired antibodies, and 9.78%, 7.37%, and 11.02% for nanobodies on ΔG, CDR AAR, and CDR RMSD, respectively. This shows that the improvement is not solely due to discretization: feedback-guided optimization can exploit the resulting search space more effectively than random sampling.

Across objectives, changing the oracle changes the property favored by the search, while Pareto optimization further enables trade-offs among different objectives. Together, these results show that *Ig7O* provides an effective interface for both single- and multi-objective antibody optimization.

We further analyze the contributions of the discrete CDR-shape space and Bayesian optimization to feedback efficiency in Appendix D.2. Under matched oracle-evaluation budgets, *Ig7O* achieves lower CDR RMSD and higher reference-free ESMFold2 rewards than continuous-space Best-of-N sampling, demonstrating improved feedback efficiency across both types of feedback.

6 Conclusion

We constructed a discrete CDR-shape space from recurrent backbone conformations and formulated epitope-specific antibody design as the selection and generative realization of CDR-shape combinations. We introduced *IgSeven* to instantiate this formulation. Its generator, *Ig7G*, supports controllable generation from flexible sequence and structural prompts, while its optimizer, *Ig7O*, performs feedback-guided Bayesian optimization over CDR-shape combinations without retraining the generator. Experiments on paired antibodies and nanobodies demonstrate faithful realization of specified CDR shapes and effective single- and multi-objective optimization. Under matched evaluation budgets, CDR-shape search improves over continuous-space Best-of-N sampling, while Bayesian optimization further improves over random search within the same discrete space. More broadly, *IgSeven* provides a new perspective on antibody generation: rather than designing directly in sequence space or continuous 3D structural space, it leverages antibody structural priors through a compact discrete representation, enabling efficient and controllable exploration of the antibody design space.

AI USE STATEMENT

We used generative AI tools solely for language polishing and improving the clarity of the manuscript. All AI-assisted edits were reviewed by the authors.

ETHICS STATEMENT

This work is computational and does not involve human subjects or private data. We acknowledge that computational antibody design may have downstream biomedical applications and should be used with appropriate experimental validation.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the model architecture, training procedure, and dataset construction and preprocessing in the main paper and appendix. Code will be released to facilitate reproducibility.

REFERENCES

Jared Adolf-Bryfogle, Qifang Xu, Benjamin North, Andreas Lehmann, and Roland L. Jr. Dunbrack. Pyigclassify: a database of antibody cdr structural classifications. *Nucleic Acids Research*, 43 (D1):D432–D438, 2015. doi: 10.1093/nar/gku1106.

Sankar Basu and Björn Wallner. DockQ: A quality measure for protein-protein docking models. *PLOS ONE*, 11(8):e0161879, 2016. doi: 10.1371/journal.pone.0161879.

Nathaniel R Bennett, Joseph L Watson, Robert J Ragotte, Andrew J Borst, DéJenaé L See, Connor Weidle, Riti Biswas, Yutong Yu, Ellen L Shrock, Russell Ault, et al. Atomically accurate de novo design of antibodies with rfdiffusion. *Nature*, 649(8095):183–193, 2026.

Salvatore Candido, Thomas Hayes, Alexander Derry, Roshan Rao, Zeming Lin, Robert Verkuil, Bryan Z. Wu, Jin Sub Lee, Elise S. Bruguera, Jehan A. Keval, Mykhailo Kopylov, John E. Pak, Wesley Wu, Neil Thomas, Samson Mataraso, Alvin Hsu, Ashton C. Trotman-Grant, Kilian Fatras, Allan dos Santos Costa, Rohil Badkundri, Halil Akın, Deniz Oktay, Jonathan Deaton, Elizabeth Montabana, Hrishita Sitwala, Yue Yu, Marius Wiggert, Dylan Alexander Carlin, Anthony W. Goering, Tomasz Blazejewski, McCullen Sandora, Michael Hla, Tina Z. Jia, Leon H. Kloker, Nicholas J. Sofroniew, Masatoshi Uehara, Jassi Pannu, Sharrol Bachas, Daniel S. Liu, Tom Sercu, and Alexander Rives. Language modeling materializes a world model of protein biology. *bioRxiv*, 2026. doi: 10.64898/2026.06.03.729735.

Henriette L. Capel, Odysseas Vavourakis, Benjamin H. Williams, Christopher R. Taylor, and Charlotte M. Deane. Sabdab2: The structural antibody database in the age of machine learning. *bioRxiv*, 2026. doi: 10.64898/2026.06.16.732554.

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

Vincent B Chen, W Bryan Arendall, Jeffrey J Headd, Daniel A Keedy, Robert M Immormino, Gary J Kapral, Laura W Murray, Jane S Richardson, and David C Richardson. Molprobity: all-atom structure validation for macromolecular crystallography. *Biological crystallography*, 66(1): 12–21, 2010.

Cyrus Chothia and Arthur M Lesk. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of molecular biology*, 196(4):901–917, 1987.

Cyrus Chothia, Arthur M Lesk, Anna Tramontano, Michael Levitt, Sandra J Smith-Gill, Gillian Air, Steven Sheriff, Eduardo A Padlan, David Davies, William R Tulip, et al. Conformations of immunoglobulin hypervariable regions. *Nature*, 342(6252):877–883, 1989.

Daniel Cutting, Frédéric A Dreyer, David Errington, Constantin Schneider, and Charlotte M Deane. De novo antibody design with se (3) diffusion. *Journal of Computational Biology*, 32(4):351–361, 2025.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9851–9864, 2020.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations*, 2026.

Frédéric A Dreyer, Daniel Cutting, Constantin Schneider, Henry Kenlay, and Charlotte M Deane. Inverse folding for antibody sequence design using deep learning. *arXiv preprint arXiv:2310.19513*, 2023.

Raphael R Eguchi, Christian A Choe, and Po-Ssu Huang. Ig-vae: Generative modeling of protein structure by direct 3d coordinate generation. *PLoS computational biology*, 18(6):e1010271, 2022.

Ada Fang, Robert G Alberstein, Simon Kelow, and Frédéric A Dreyer. Tokenizing loops of antibodies. In *Mabs*, volume 18, pp. 2645271. Taylor & Francis, 2026.

Thomas Hayes, Roshan Rao, Halil Akin, Nicholas J Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q Tran, Jonathan Deaton, Marius Wiggert, et al. Simulating 500 million years of evolution with a language model. *Science*, 387(6736):850–858, 2025.

Magnus Haraldson Høie, Alissa M Hummer, Tobias H Olsen, Broncio Aguilar-Sanjuan, Morten Nielsen, and Charlotte M Deane. Antifold: improved structure-based antibody design using inverse folding. *Bioinformatics Advances*, 5(1):vbae202, 2025.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. *ICML*, 2022. doi: 10.1101/2022.04.10.487779.

Simon Kelow, Bulat Faezov, Qifang Xu, Mitchell Parker, Jared Adolf-Bryfogle, and Roland L. Jr. Dunbrack. A penultimate classification of canonical antibody cdr conformations. *bioRxiv*, 2022. doi: 10.1101/2022.10.12.511988.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Conditional antibody design as 3d equivariant graph translation. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023a.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17409–17429. PMLR, 2023b.

Xiangzhe Kong, Zishen Zhang, Ziting Zhang, Rui Jiao, Jianzhu Ma, Wenbing Huang, Kai Liu, and Yang Liu. Unimomo: Unified generative modeling of 3d molecules for de novo binder design. In *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025.

Xiangzhe Kong, Junwei Chen, Ziting Zhang, Gaodeng Li, Qingyuan Zhu, Lei Wei, Mingyu Li, Yan Shi, Weiyang Dai, Zishen Zhang, et al. Programming biomolecular interactions with all-atom generative model. *bioRxiv*, pp. 2026–03, 2026.

Desmond Kuan and Amir Barati Farimani. Abgpt: de novo antibody design via generative language modeling. *arXiv preprint arXiv:2409.06090*, 2024.

Chingwei V. Lee, Wei-Ching Liang, Mark S. Dennis, Charles Eigenbrot, Sachdev S. Sidhu, and Germaine Fuh. High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *Journal of Molecular Biology*, 340(5):1073–1093, 2004. doi: 10.1016/j.jmb.2004.05.051.

Lihang Liu, Yang Liu, Xianbin Ye, Shanzhuo Zhang, Yuxin Li, Kunrui Zhu, Yang Xue, Jingbo Zhou, Xiaonan Zhang, and Xiaomin Fang. Reshaping biomolecular structure prediction through strategic conformational exploration with helixfold-s1. *Nature Machine Intelligence*, 8(7):1115–1126, 2026. ISSN 2522-5839. doi: 10.1038/s42256-026-01264-2.

Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances in Neural Information Processing Systems*, 35:9754–9767, 2022.

Karolis Martinkus, Jan Ludwiczak, Wei-Ching Liang, Julien Lafrance-Vanasse, Isidro Hotzel, Arvind Rajpal, Yan Wu, Kyunghyun Cho, Richard Bonneau, Vladimir Gligorijevic, et al. Abdiffuser: full-atom generation of in-vitro functioning antibodies. *Advances in Neural Information Processing Systems*, 36:40729–40759, 2023.

Fanxu Meng, Na Zhou, Guangchun Hu, Ruotong Liu, Yuanyuan Zhang, Ming Jing, and Qingzhen Hou. A comprehensive overview of recent advances in generative models for antibodies. *Computational and structural biotechnology journal*, 23:2648–2660, 2024.

Luis S Mille-Fragoso, Claudia L Driscoll, John N Wang, Haoyu Dai, Talal Widatalla, Jim L Zhang, Xiaowei Zhang, Bing Rao, Liang Feng, Brian L Hie, et al. Efficient generation of epitope-targeted antibodies with germinal. *Nature biotechnology*, pp. 1–10, 2026.

Shentong Mo and Lanqing Li. Sadit: Efficient protein backbone design via latent structural tokenization and diffusion transformers. *arXiv preprint arXiv:2602.06706*, 2026.

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

Changyong Oh, Jakub M. Tomczak, Efstratios Gavves, and Max Welling. Combinatorial bayesian optimization using the graph cartesian product. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

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

Aubin Ramon, Montader Ali, Misha Atkinson, Alessio Saturnino, Kieran Didi, Cristina Visentin, Stefano Ricagno, Xing Xu, Matthew Greenig, and Pietro Sormanni. Assessing antibody and nanobody nativeness for hit selection and humanization with abnativ. *Nature Machine Intelligence*, 6(1):74–91, 2024.

Richard W Shuai, Jeffrey A Ruffolo, and Jeffrey J Gray. Iglm: Infilling language modeling for antibody sequence design. *Cell systems*, 14(11):979–989, 2023.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, et al. Boltzgen: Toward universal binder design. *BioRxiv*, 2025.

Jin Su, Chenchen Han, Yuyang Zhou, Junjie Shan, Xibin Zhou, and Fajie Yuan. Saprot: Protein language modeling with structure-aware vocabulary. In *International Conference on Learning Representations*, volume 2024, pp. 6987–7009, 2024.

Michael Sun, Weize Yuan, Gang Liu, Wojciech Matusik, and Marinka Zitnik. Protein structure tokenization via geometric byte pair encoding. *International Conference on Learning Representations, ICLR*, 2026.

BioGeometry Team and Jian Tang. Rapid de novo antibody design with geoflow-v3. *bioRxiv*, pp. 2025–10, 2025.

ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, Shenghao Wu, Kuangqi Zhou, Yanping Yang, Zhenyu Liu, Lan Wang, Bo Shi, Shaochen Shi, and Wenzhi Xiao. Protenix - advancing structure prediction through a comprehensive alphafold3 reproduction. *bioRxiv*, 2025a. doi: 10.1101/2025.01.08.631967.

Latent Labs Team, Henry Kenlay, Daniella Pretorius, Jonathan Crabbé, Alex Bridgland, Sebastian M Schmon, Agrin Hilmkil, James Vuckovic, Simon Mathis, Tomas Matteson, et al. Drug-like antibodies with low immunogenicity in human panels designed with latent-x2. *arXiv preprint arXiv:2512.20263*, 2025b.

Michel Van Kempen, Stephanie S Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron LM Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with foldseek. *Nature biotechnology*, 42(2):243–246, 2024.

Luiz Felipe Vecchietti, Bryan Nathanael Wijaya, Azamat Armanuly, Begench Hangeldiyev, Hyunkyu Jung, Sooyeon Lee, Meeyoung Cha, and Ho Min Kim. Artificial intelligence-driven computational methods for antibody design and optimization. In *MAbs*, volume 17, pp. 2528902. Taylor & Francis, 2025.

Rubo Wang, Fandi Wu, Xingyu Gao, Jiaxiang Wu, Peilin Zhao, and Jianhua Yao. Iggm: A generative model for functional antibody and nanobody design. In *International Conference on Learning Representations*, volume 2025, pp. 100801–100827, 2025.

Sheng Xu, Qiantai Feng, Lifeng Qiao, Hao Wu, Tao Shen, Yu Cheng, Shuangjia Zheng, and Siqi Sun. Benchmarking all-atom biomolecular structure prediction with FoldBench. *Nature Communications*, December 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-67127-3.

Xinyu Yuan, Zichen Wang, Marcus D. Collins, and Huzefa Rangwala. Protein structure tokenization: Benchmarking and new recipe. In *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pp. 3813–3824. IEEE, 2023. doi: 10.1109/ICCV51070.2023.00355.

Xingyi Zhang, Kun Xie, Ningqiao Huang, Wei Liu, Peilin Zhao, Sibo Wang, Kangfei Zhao, and Biaobin Jiang. Fast and accurate antibody sequence design via structure retrieval. *arXiv preprint arXiv:2502.19395*, 2025.

Yuxuan Zhang, Chengyue Gong, Jinyuan Sun, Jiaqi Guan, Milong Ren, Song Xue, Hanyu Zhang, Wenzhi Ma, Zhenyu Liu, Xinshi Chen, and Wenzhi Xiao. Protenix-v2: Broadening the reach of structure prediction and biomolecular design. *bioRxiv*, 2026. doi: 10.64898/2026.04.10.717613.

Tian Zhu, Milong Ren, and Haicang Zhang. Antibody design using a score-based diffusion model guided by evolutionary, physical and geometric constraints. In *Forty-first International Conference on Machine Learning*, 2024.

A Protein structure tokenization methods

Protein structure tokenization seeks to discretize continuous protein geometries into a finite vocabulary of structural states, providing compact and reusable representations for structural modeling (Su

et al., 2024; Van Kempen et al., 2024; Yuan et al., 2025; Hayes et al., 2025), and generation (Dauparas et al., 2022; Mo & Li, 2026). However, most existing methods remain *residue-aligned*, assigning approximately one structural token to each residue. Very recent work (Sun et al., 2026) has extended structure tokenization beyond the residue level to variable-length structural motifs, but it is designed for general proteins rather than antibody-specific structures.

Compared with general proteins, antibodies offer a particularly natural setting for higher-level structural tokenization: their CDRs are well-defined, variable-length structural units that often adopt a limited repertoire of recurrent backbone conformations. Classical CDR canonical clustering (Chothia et al., 1989; Kelow et al., 2022; Adolf-Bryfogle et al., 2015) captures this regularity by grouping loops with similar backbone geometries into discrete conformational classes, providing a natural basis for a CDR-level structural vocabulary. More recently, learning-based approaches have revisited this idea. IgSeek (Zhang et al., 2025) learns geometric representations of CDR backbones to retrieve structurally similar natural loops and uses the retrieved sequences to guide CDR sequence design, while IGLOO (Fang et al., 2026) jointly encodes CDR sequences and backbone dihedral angles into continuous and discrete loop-level tokens and organizes the latent space according to structural similarity. However, to our knowledge, prior work has not used discrete CDR conformations as a generative vocabulary for antibody design. We build on this idea by retrieving and composing CDR structural motifs to guide antibody generation.

B DATASETS

B.1 CANONICAL CDR CLUSTERS

Following the density-based clustering strategy of Kelow et al. (2022), we use DBSCAN to identify recurrent CDR conformations while treating structural outliers as noise. CDRs are first grouped by type and loop length, and pairwise C_α RMSD is used as the structural distance within each group. We sweep the neighborhood radius ϵ and minimum number of points minPts, and evaluate the resulting cluster compactness and vocabulary size. We select $\epsilon = 0.35\,\text{Å}$ and $\mathrm{minPts} = 3$, which provide a reasonable balance between the two. CDRs that do not belong to any sufficiently dense neighborhood are treated as noise and excluded from the structural vocabulary.

Figure 6 compares the pairwise Cα RMSD distributions within and across clusters for CDR loops of the same type and length. Intra-cluster distances are consistently smaller than inter-cluster distances, indicating that the generated clusters capture structurally compact and distinct backbone conformations. Figure 7 further shows the distribution of shape tokens across CDR types and loop lengths. The resulting vocabularies vary considerably in size, ranging from 10 tokens for L2 to 370 for H3, with H3 exhibiting the broadest diversity across loop lengths. For each of the six CDR types, we select the most common loop length and visualize the five most frequent shape tokens of that length in Figures 8 and 9.

We further characterize the CDR combinatorial space statistics. For each test antibody–antigen complex, we compute the full set of CDR-token combinations compatible with its CDR lengths. In parallel, we count the unique CDR-token combinations actually realized by antibodies in SAbDab2. As summarized in Table 4, the mean size of the full combinatorial space for test samples is 2.540×10^8 for paired antibodies and 52,078 for VHHs, while 1,926 and 608 unique CDR-token combinations are observed by antibodies in SAbDab2, respectively. For reference, using representative common IMGT CDR lengths of $[8, 8, 14]$ for CDR-H1/H2/H3 and $[6, 3, 9]$ for CDR-L1/L2/L3, a paired antibody has 48 CDR residues, corresponding to an unconstrained amino-acid sequence space of $20^{48} \approx 2.8 \times 10^{62}$.

Table 4: Combinatorial CDR design space.

	Paired antibody	Nanobody
Mean space per test target	2.540×10^8	52,078
Unique combinations in dataset	1,926	608

Flexibility within CDR structural motifs Our CDR-shape tokens characterize loop length and backbone geometry rather than prescribing a unique sequence or full-atom conformation. Consequently, CDRs assigned to the same token can preserve nearly identical backbone conformations while forming substantially different interfaces with the antigen.

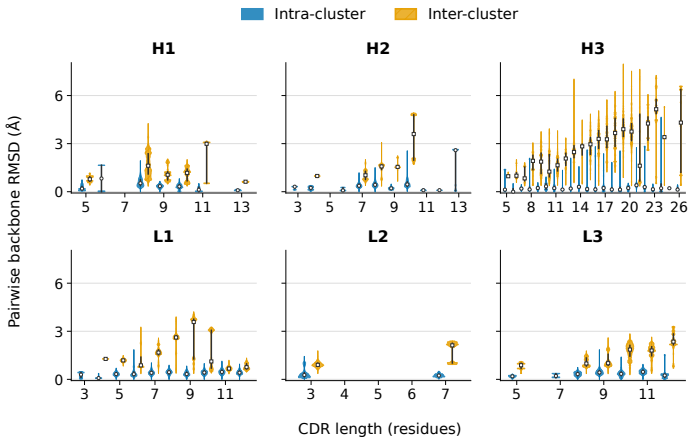


Figure 6: Intra- and inter-cluster Cα RMSD. Loops are compared within each CDR type and length.

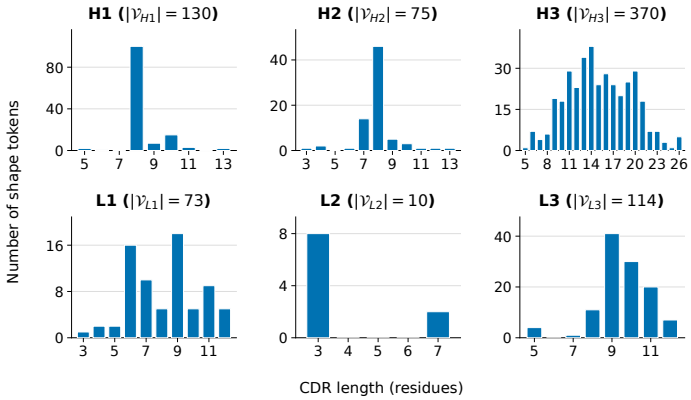


Figure 7: CDR-shape counts by type and length.

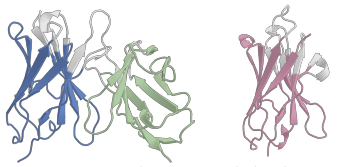
The HCDR3 loops of 7PS1 and 7QF0 belong to `H3_len11_cluster27` (backbone RMSD: 0.269 Å) but differ at three sequence positions (`ARDLVVYGMDV` versus `ARDGDYYGMDV`). They engage largely overlapping epitopes, yet 7QF0 forms an Asp–Lys417 salt bridge (2.54 Å) absent from 7PS1, where the corresponding Val/Asn pair makes no direct contact. Thus, a shared shape token accommodates distinct side-chain chemistry and antigen interactions.

The LCDR3 loops of 8HPF and 9CPV share `L3_len9_cluster1` (backbone RMSD: 0.231 Å) but differ at three sequence positions (`QQFGDSPFT` versus `QQYGSSPRT`) and contact non-overlapping LCDR3-specific epitopes. Their local shape-complementarity scores are 0.697 and 0.759, illustrating how the same backbone motif supports complementary interfaces with distinct epitopes. The distinct HCDR3 conformations are accompanied by different binding orientations and local interaction environments around LCDR3. These structural differences provide a rationale for context-dependent LCDR3 sequence compatibility beyond its shared backbone shape. We present these two examples in Figure 10.

B.2 Choice of Antibody Frameworks

We use *trastuzumab* (hu4D5-8) as the fixed framework for paired H/L antibodies and *h_NbBCII10* for nanobodies. Both scaffolds have been repeatedly adopted in recent antibody design studies and are supported by experimental precedent.

For paired antibodies, the hu4D5-8/trastuzumab framework is used by RFantibody Bennett et al. (2026) and Germinal Mille-Fragoso et al. (2026). Synthetic libraries built on this scaffold have yielded functional antibodies against diverse anti-



(a) trastuzumab (b) h_NbBCII10

Figure 11: Frameworks templates.

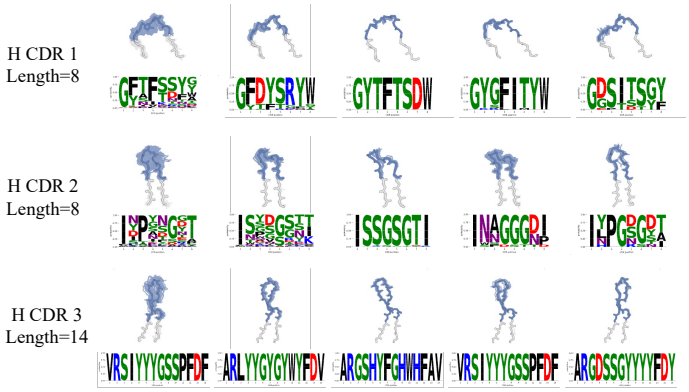


Figure 8: Representative heavy-chain CDR shapes. Each row shows backbone conformations and sequence logos for five clusters at the most common loop length.

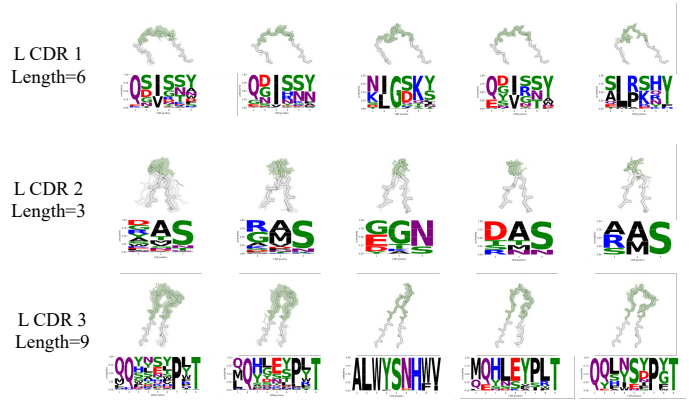


Figure 9: Representative light-chain CDR shapes. Each row shows backbone conformations and sequence logos for five clusters at the most common loop length.

gens through CDR diversification, supporting its use in fixed-framework design Lee et al. (2004).

For nanobodies, h-NbBcII10_FGLA is used by RFantibody Bennett et al. (2026), Germinal Mille-Fragoso et al. (2026), GeoFlow-V3 Team & Tang (2025), and Latent-X2 Team et al. (2025b), with experimental validation across multiple antigens and epitopes.

B.3 Train/test splits

We use SAbDab2 (Capel et al., 2026), a weekly updated database of publicly available antibody structures that provides cleaned data and predefined train/test splits for AI/ML applications. During pre-cleaning, SAbDab2 retains only X-ray or EM structures with resolutions better than 3.5 Å and removes entries containing non-standard or unknown amino acids. For data splitting, samples are clustered separately by antibody CDRH3, CDRH1–3, and CDRL1–3, also antigen sequences. Samples sharing any antibody or antigen cluster are then merged through connected components, and the resulting clusters are split into training and test sets.

Table 5: Dataset statistics.

	Train	Test	Total
Paired antibody	3,278	71	3,349
Nanobody	1,038	25	1,063
Total	4,316	96	4,412

We apply *additional filtering* for antigen-specific antibody generation. Structures without antigens are removed, stricter structural completeness checks are performed, and, to further reduce potential data leakage, the *test set is restricted to structures released after 2025*. The final dataset statistics are summarized in Table 5.

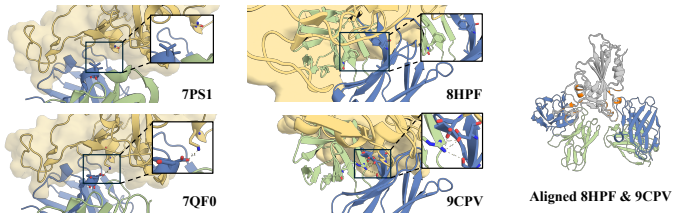


Figure 10: Flexibility within CDR structural motifs.

Table 6: Inverse folding on paired antibodies and nanobodies. Average rank is computed across the four metrics. Bold and underlined values indicate the best and second-best results, respectively.

Method	Full AAR ↑	CDR AAR ↑	Full scRMSD (Å) ↓	CDR scRMSD (Å) ↓	Avg. Rank ↓	Antibody Data
AntiFold	**0.836**	**0.650**	1.445	2.412	<u>2.25</u>	149k
AbMPNN	0.813	<u>0.623</u>	**1.370**	<u>2.090</u>	**2.00**	151k
ProteinMPNN	0.525	0.414	1.541	2.225	3.75	–
ESM-IF1	0.505	0.409	1.834	2.693	5.00	–
Ig7G	<u>0.819</u>	0.575	<u>1.389</u>	**2.079**	**2.00**	4k

C Training configurations

Backbone adaptation. We adapt the pretrained Complexa denoiser to SAbDab2 antibody–antigen complexes in three consecutive stages of 5, 5, and 10 epochs, with learning rates of 2×10^{-5}, 10^{-5}, and 5×10^{-6}, respectively. Hotspot dropout is progressively reduced from 0.80 to 0.50 and 0.15. Each stage is initialized from the final EMA checkpoint of the preceding stage. The autoencoder remains frozen, and antibody-side control channels are disabled throughout adaptation.

Control-branch training. The control branch is initialized from the adapted backbone, after which the backbone is frozen and only the control modules are optimized for 40 epochs with a learning rate of 5×10^{-5}. Training primarily uses framework-level latent and sequence conditions together with structural controls that capture intra- and inter-CDR geometry. To expose the model to stronger conditioning, full-antibody residue identities and absolute-coordinate controls are each provided with probability 0.05.

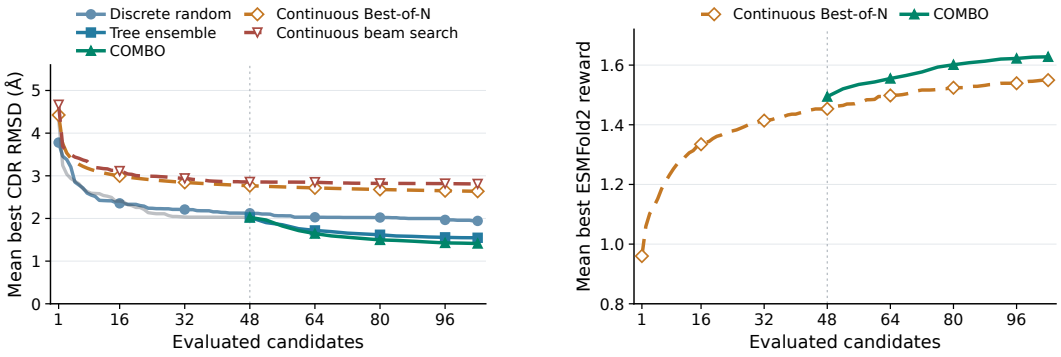
Sampling. We use the final EMA checkpoint with self-conditioning. For fixed-framework design, we provide framework-specific latent and sequence conditions together with structural controls derived from the selected CDR representatives. Full-antibody residue identities and absolute-coordinate controls are disabled. Backbone coordinates and residue latents are initialized from Gaussian noise, and no external pre-docking is used.

D Additional experiments

D.1 Inverse folding

Task setup. In this section, we evaluate whether structural prompts can guide *IgSeven* to generate antibody sequences. Specifically, we provide the antibody structure as a prompt and generate its sequence conditioned on the target antigen. The generated sequences are then refolded by Protenix-v2, best folding model regarding RMSD, as shown in Table 1. We report the amino acid recovery (AAR) and self-consistency RMSD (scRMSD) over the full antibody and CDR regions separately. We compare *IgSeven* with AntiFold (Høie et al., 2025), AbMPNN (Dreyer et al., 2023), ProteinMPNN (Dauparas et al., 2022) and ESM-IF1 (Hsu et al., 2022). For each method, we evaluate a single prediction per target, without filtering or candidate selection.

Results. Table 6 presents the inverse-folding results on all 96 paired-antibody and nanobody targets. Overall, *IgSeven* delivers highly competitive inverse-folding performance, achieving the best average rank of 2.00 across the four evaluation metrics, tied with the strongest specialized baseline. In particular, *IgSeven* attains the best CDR scRMSD of 2.079 Å, together with the second-



(a) Reference-based CDR RMSD feedback.

(b) Reference-free ESMFold2 feedback.

Figure 12: Search efficiency under matched oracle-evaluation budgets. Curves report mean best-so-far scores; lower is better in (a), and higher is better in (b). Initialization evaluations are included in the total budget.

highest full-antibody AAR of 81.93% and the second-lowest full-antibody scRMSD of 1.389 Å. This is particularly notable as *IgSeven* is neither designed as a dedicated inverse-folding model nor explicitly trained to optimize antibody sequence recovery, while using only about 4k antibody training samples, compared with approximately 149k and 151k used by AntiFold and AbMPNN, respectively. Despite the substantially smaller amount of antibody-specific training data, *IgSeven* remains competitive with these specialized models. These results demonstrate that *IgSeven* can effectively recover sequences consistent with a given antibody structure, further supporting the strong sequence–structure alignment learned by the model.

D.2 Search efficiency

We evaluate search efficiency using two complementary oracles: reference-based CDR RMSD for controlled analysis and a reference-free ESMFold2 confidence reward to assess whether the efficiency advantage extends beyond reference-based diagnostics.

Evaluation protocol. We compare continuous-space search with the antibody-adapted Complexa generator $\mathcal{G}_{\theta_{\mathrm{Ab}}}$ against search over discrete CDR-shape combinations realized by *Ig7G*. The generators share the same base denoiser $\mathcal{N}_{\theta_{\mathrm{Ab}}}$, while *Ig7G* additionally uses its trained control branch to realize the selected structural conditions. We report mean best-so-far oracle scores under matched evaluation budgets, including initialization evaluations. All generator parameters remain fixed throughout search.

Reference-based structural feedback. We first use CDR RMSD as a controlled diagnostic to examine the contributions of the search interface and optimization strategy. Figure 12a compares Best-of-N sampling and beam search in the continuous generation space with random search, tree-ensemble Bayesian optimization, and COMBO in the discrete CDR-shape space.

At 104 evaluations, discrete random search achieves a mean best CDR RMSD of 1.943 Å, compared with 2.637 Å for continuous-space Best-of-N sampling. This shows that the shape-conditioned search interface improves search performance even without model-guided token selection. Within the same discrete space and evaluation budget, tree-ensemble Bayesian optimization and COMBO further reduce mean best CDR RMSD to 1.547 Å and 1.413 Å, respectively. This within-space comparison demonstrates that Bayesian optimization uses the evaluation budget more effectively than random exploration. Notably, *Ig7O* matches the mean best CDR RMSD of Complexa Best-of-N using only 8 initialization evaluations versus 104, reducing the evaluation budget by a factor of 13.

Reference-free confidence feedback. We next examine whether the efficiency advantage extends to feedback that does not require a native antibody reference. We compare COMBO with

continuous-space Best-of-N sampling using the same ESMFold2 reward, which assigns equal weight to complex-level pTM and antibody–antigen interface ipTM.

As shown in Figure 12b, the discrete search initialization already achieves a mean best reward approximately 2.82% higher than Best-of-N after 48 evaluations. Once acquisition-guided search begins, COMBO widens this advantage: the relative improvement reaches 3.79%, 5.05%, and 5.40% at 64, 80, and 96 evaluations, respectively. Thus, the initial advantage precedes Bayesian optimization, while subsequent COMBO search yields progressively larger gains over continuous-space sampling. These results show that the efficiency advantage also holds under reference-free confidence feedback.

Together, these experiments demonstrate improved oracle-evaluation efficiency under both reference-based structural feedback and reference-free confidence feedback. The RMSD experiment further shows that Bayesian optimization improves over random search within the same discrete CDR-shape space.

D.3 Baselines

Epitope-constrained co-folding. In §5.1, we evaluate the following folding baselines:

- **ESMFold2** (Candido et al., 2026) builds on ESMC protein language model representations to infer atomic-resolution structures. We evaluate both the full and fast checkpoints and report results from the fast checkpoint, which performs better in our evaluation.
- **Protenix-v2** (Zhang et al., 2026) scales up the AlphaFold3-style architecture of Protenix through wider internal representations while retaining its Pairformer trunk and all-atom diffusion module.
- **Chai-1** (Chai Discovery, 2024) adopts an AlphaFold3-like diffusion architecture that integrates sequence, evolutionary, template, and restraint features.
- **HelixFold-S1** (Liu et al., 2026) augments HelixFold3 with contact prediction and conditioning modules for guided structural sampling.
- **Protenix-v0.5.0** (Team et al., 2025a) combines a Pairformer trunk with an all-atom diffusion module and supports contact and pocket conditioning through a constraint-enabled checkpoint.

For epitope conditioning, we use inter-chain restraints in Chai-1 and pocket constraints in Protenix-v0.5.0. HelixFold-S1 uses predicted inter-chain contacts to guide co-folding, while ESMFold2 and Protenix-v2 do not support specifying target epitopes. We omit antigen structural templates for Protenix-v0.5.0 because its constraint-enabled checkpoint does not support template inputs.

Realizing specified CDR shapes. In §5.2, we include the following methods as references for antibody design quality. Only *Ig7G* receives oracle CDR-shape labels; the reference methods do not use these labels.

- **RFantibody** (Bennett et al., 2026) combines antibody-fine-tuned RFdiffusion for backbone generation, ProteinMPNN for sequence design, and antibody-fine-tuned RoseTTAFold2 for structural assessment in a three-stage pipeline.
- **DyMEAN** (Kong et al., 2023b) uses an adaptive multi-channel equivariant graph network to iteratively update residue identities and full-atom coordinates, with a shadow paratope representation connecting antibody modeling to antigen docking. Its released checkpoint supports paired antibody design only.
- **BoltzGen** (Stark et al., 2025) extends the Boltz-2 architecture with a geometric encoding of amino acid identities, enabling joint sequence and all-atom structure generation within a continuous diffusion process.
- **AbX** (Zhu et al., 2024) adopts a score-based diffusion formulation for sequence–structure co-design, incorporating evolutionary priors and physical and geometric constraints.
- **DiffAb** (Luo et al., 2022) couples discrete diffusion over amino acid identities with continuous diffusion over residue positions and orientations, using a geometric denoising network conditioned on the surrounding structural context.

Both AbX and DiffAb require a structurally positioned framework–antigen complex as input. Although DiffAb provides an HDOCK-based workflow using a complete antibody template, this introduces template CDRs into the docking step. In our setting, only the framework is specified.

Therefore, for these two methods, we place the framework template in the reference docking pose by Kabsch alignment to the ground-truth framework.

Feedback-guided optimization. In §5.3, we use two search baselines to assess the contributions of the discrete CDR-shape space and feedback-guided candidate selection:

- **Continuous-space Best-of-N** samples candidates from the antibody-adapted Complexa generator $\mathcal{G}_{\theta_{\mathrm{Ab}}}$ without CDR-shape control and selects the best candidate according to the chosen oracle.
- **Discrete random search** randomly samples CDR-shape combinations, generates candidates using *Ig7G*, and selects the best evaluated candidate according to the same oracle. Oracle feedback is not used to guide subsequent candidate selection.

The search-efficiency experiments additionally include continuous-space beam search with $\mathcal{G}_{\theta_{\mathrm{Ab}}}$ under matched oracle-evaluation budgets (Appendix D.2). The tree-ensemble, COMBO, and Pareto variants of *Ig7O* are described in Appendix E.

For reference, the best and second-best baseline rows in Table 3 report single-design results without feedback-guided optimization. These values are selected separately for each metric among RFantibody, DyMEAN, and BoltzGen, the design baselines that do not require a fixed docking pose; DyMEAN is included only for paired antibodies.

Antibody inverse folding. For the antibody inverse-folding task in Appendix D.1, we include the following baselines:

- **AntiFold** (Høie et al., 2025) fine-tunes ESM-IF1 on antibody structures for backbone-conditioned sequence design.
- **AbMPNN** (Dreyer et al., 2023) adapts ProteinMPNN through antibody-specific fine-tuning for structure-conditioned sequence design.
- **ProteinMPNN** (Dauparas et al., 2022) is a general-purpose inverse-folding model that combines message passing over backbone geometry with autoregressive sequence decoding.
- **ESM-IF1** (Hsu et al., 2022) is a structure-conditioned inverse-folding model that combines invariant geometric processing based on geometric vector perceptrons (GVPs) with an autoregressive encoder–decoder Transformer to generate protein sequences from backbone coordinates.

Each method is evaluated using a single prediction per target, without filtering or candidate selection.

D.4 METRICS

We describe the evaluation metrics used for each task below.

Epitope-constrained co-folding. In §5.1, we evaluate antibody structural accuracy and antibody–antigen docking quality using:

- **Cα RMSD** and **Full-atom RMSD** measure deviations from the reference antibody structure over Cα atoms and all heavy atoms, respectively, after alignment using the Kabsch algorithm. Lower values indicate greater structural accuracy.
- **DockQ** (Basu & Wallner, 2016) assesses docking quality by combining native contact recovery, interface RMSD, and ligand RMSD. Following Foldbench (Xu et al., 2025), we assess only the interaction between antibody chains and antigen chains.
- **Docking success rate** reports the fraction of predicted complexes with DockQ > 0.23, as suggested by Basu & Wallner (2016).

Realizing specified CDR shapes. In §5.2, we evaluate structural and sequence quality using four metrics:

- **CDR RMSD** measures the deviation of designed CDR backbone coordinates from the reference after framework alignment. Lower values indicate greater structural fidelity.
- **MolProbity** (Chen et al., 2010) assesses stereochemical quality through steric clashes, rotamer outliers, and Ramachandran statistics. Lower scores indicate better stereochemical quality.

- **CDR amino acid recovery (CDR AAR)** reports the fraction of designed CDR residues matching the reference sequence.
- **CDR sequence nativeness (CDR SeqNat)** measures the nativeness of designed CDR sequences using AbNatiV (Ramon et al., 2024), with higher scores indicating greater resemblance to natural antibody sequences.

Feedback-guided optimization. In §5.3, we use CDR AAR and CDR RMSD, as defined above, as reference-based diagnostic objectives, and computed ΔG as an energy-based objective. Single-objective search maximizes CDR AAR or minimizes CDR RMSD or ΔG, while Pareto search considers all three objectives jointly.

- **Computed interface energy** (ΔG) measures the Rosetta interface energy in Rosetta energy units (REU). Following the protocol in BindCraft (Pacesa et al., 2025), we use InterfaceAnalyzerMover after constrained FastRelax with a common all-atom energy function. Backbone and side chains are relaxed while inter-partner jumps are fixed, and coordinate restraints preserve the predicted pose. Lower values are preferred.
- **ESMFold2 reward** is used in the additional search-efficiency experiments in Appendix D.2. It combines the complex-level pTM and antibody–antigen interface ipTM from ESMFold2 with equal weights. Higher rewards are preferred.

Table 3 reports the best observed value for each objective separately; the values in one row need not correspond to the same generated design. Search-efficiency curves track the best observed objective value as a function of the number of oracle evaluations. The evaluation budget includes initialization, and reported designs are selected from stored, evaluated candidates, as described in Appendix E.

Antibody inverse folding. For the task in Appendix D.1, we evaluate sequence recovery and structural self-consistency using:

- **Amino acid recovery (AAR)** reports the fraction of generated residues matching the reference sequence.
- **Self-consistency RMSD (scRMSD)** measures the deviation between the refolded structure of the generated sequence and the input structure after structural alignment. We use Protenix-v2 for refolding.

Both AAR and scRMSD are reported over the full antibody and the CDR regions separately for paired antibodies and nanobodies. The average rank in Table 6 is computed across these four metrics, with lower ranks indicating better overall performance.

E Search Algorithms for *Ig7O*

E.1 Shared notation and evaluation protocol

We retain the notation of the main text: $\boldsymbol{t}$ denotes a CDR-token combination, $\boldsymbol{y}(\boldsymbol{t})$ its observed feedback, $\mathcal{D}_n$ the history after n evaluations, $\hat{\boldsymbol{f}}_n$ the fitted surrogate, and a_n the acquisition function. For a single objective, we abbreviate the sole components of $\boldsymbol{y}$ and $\hat{\boldsymbol{f}}_n$ as y and $\hat{f}_n$, respectively. The active CDR types are indexed by $\mathcal{R} = \{\mathrm{H1}, \mathrm{H2}, \mathrm{H3}, \mathrm{L1}, \mathrm{L2}, \mathrm{L3}\}$ for paired antibodies and $\mathcal{R} = \{\mathrm{H1}, \mathrm{H2}, \mathrm{H3}\}$ for nanobodies. Since Ω_{Ab} includes the fixed framework token, the unevaluated CDR combinations are

$$\mathcal{U}_n = \{\boldsymbol{t} : (t_{\mathrm{F}}, \boldsymbol{t}) \in \Omega_{\mathrm{Ab}}, \quad \boldsymbol{t} \notin \{\boldsymbol{t}_i\}_{i=1}^{n}\}\,. \tag{7}$$

Any prescribed loop lengths are enforced by restricting the eligible tokens at the corresponding CDR positions. The set $\mathcal{U}_n$ and the product graph are represented implicitly; neither requires enumerating every token combination.

Algorithm 1 *Ig7O* with RF+UCB

Input: Fixed antigen $\mathcal{A}$, epitope E, framework token t_{F}, and design space Ω_{Ab}; frozen *Ig7G* and a property oracle.
Input: Initial history $\mathcal{D}_{n_0}$ and stored designs $\{(S_i, X_i)\}_{i=1}^{n_0}$; total evaluation budget $B \geq n_0 \geq 1$; number of trees $K \geq 1$; candidate-pool size $C \geq 1$; exploration coefficient $\beta \geq 0$.
$n \leftarrow n_0$; initialize the unevaluated set $\mathcal{U}_n$ by Eq. (7).
while $n < B$ **and** $\mathcal{U}_n \neq \varnothing$ **do**
  Fit an RF surrogate $\hat{f}_n$ with K trees to the one-hot-encoded observations in $\mathcal{D}_n$.
  Sample $\mathcal{C}_n \subseteq \mathcal{U}_n$ with $|\mathcal{C}_n| = \min(C, |\mathcal{U}_n|)$.
  Compute $\mu_n(\boldsymbol{t})$ and $\sigma_n(\boldsymbol{t})$ for $\boldsymbol{t} \in \mathcal{C}_n$ using Eq. (9).
  $a_n(\boldsymbol{t}; \hat{f}_n) \leftarrow \mu_n(\boldsymbol{t}) + \beta\sigma_n(\boldsymbol{t})$.
  $\boldsymbol{t}_{n+1} \in \arg\max_{\boldsymbol{t} \in \mathcal{C}_n} a_n(\boldsymbol{t}; \hat{f}_n)$.
  $(S_{n+1}, X_{n+1}, \boldsymbol{y}(\boldsymbol{t}_{n+1})) \leftarrow$ EVALUATE($\boldsymbol{t}_{n+1}$).
  Store (S_{n+1}, X_{n+1}).
  $\mathcal{D}_{n+1} \leftarrow \mathcal{D}_n \cup \{(\boldsymbol{t}_{n+1}, \boldsymbol{y}(\boldsymbol{t}_{n+1}))\}$.
  $\mathcal{U}_{n+1} \leftarrow \mathcal{U}_n \setminus \{\boldsymbol{t}_{n+1}\}$; $n \leftarrow n + 1$.
end while
$i^\star \in \arg\max_{1 \leq i \leq n} y(\boldsymbol{t}_i)$.
return $(\boldsymbol{t}_{i^\star}, S_{i^\star}, X_{i^\star}, \boldsymbol{y}(\boldsymbol{t}_{i^\star}))$.

Algorithm 2 *Ig7O* with COMBO and expected improvement

Input: Fixed conditioning and initial evaluated designs as in Algorithm 1; budget B; per-CDR graphs $\{G_r\}_{r \in \mathcal{R}}$; candidate-pool size $C \geq 1$; restart count $1 \leq J \leq C$.
Define the product graph $G = \square_{r \in \mathcal{R}} G_r$ and precompute the eigensystem of each graph Laplacian L_r.
$n \leftarrow n_0$; initialize $\mathcal{U}_n$ by Eq. (7).
while $n < B$ **and** $\mathcal{U}_n \neq \varnothing$ **do**
  Fit the GP surrogate $\hat{f}_n$ to $\mathcal{D}_n$ using the graph diffusion kernel in Eq. (11).
  $y_n^\star \leftarrow \max_{1 \leq i \leq n} y(\boldsymbol{t}_i)$.
  Define $a_n(\boldsymbol{t}; \hat{f}_n)$ as EI using Eq. (12).
  $\boldsymbol{t}_{n+1} \leftarrow$ GRAPHSEARCH($a_n(\cdot; \hat{f}_n), G, \mathcal{U}_n, C, J$).
  $(S_{n+1}, X_{n+1}, \boldsymbol{y}(\boldsymbol{t}_{n+1})) \leftarrow$ EVALUATE($\boldsymbol{t}_{n+1}$).
  Store (S_{n+1}, X_{n+1}).
  $\mathcal{D}_{n+1} \leftarrow \mathcal{D}_n \cup \{(\boldsymbol{t}_{n+1}, \boldsymbol{y}(\boldsymbol{t}_{n+1}))\}$.
  $\mathcal{U}_{n+1} \leftarrow \mathcal{U}_n \setminus \{\boldsymbol{t}_{n+1}\}$; $n \leftarrow n + 1$.
end while
$i^\star \in \arg\max_{1 \leq i \leq n} y(\boldsymbol{t}_i)$.
return $(\boldsymbol{t}_{i^\star}, S_{i^\star}, X_{i^\star}, \boldsymbol{y}(\boldsymbol{t}_{i^\star}))$.

function GRAPHSEARCH($a, G, \mathcal{U}, C, J$)
  Sample $\mathcal{C} \subseteq \mathcal{U}$ with $|\mathcal{C}| = \min(C, |\mathcal{U}|)$.
  Let $\mathcal{S}$ contain the $\min(J, |\mathcal{C}|)$ candidates in $\mathcal{C}$ with highest acquisition values.
  $\mathcal{M} \leftarrow \mathcal{S}$.
  for each $\boldsymbol{s} \in \mathcal{S}$ **do**
    $\boldsymbol{u} \leftarrow \boldsymbol{s}$.
    while true do
      $\mathcal{W} \leftarrow \{\boldsymbol{u}\} \cup \big(\mathcal{N}_G(\boldsymbol{u}) \cap \mathcal{U}\big)$.
      $\boldsymbol{v} \in \arg\max_{\boldsymbol{w} \in \mathcal{W}} a(\boldsymbol{w})$.
      if $a(\boldsymbol{v}) \leq a(\boldsymbol{u})$ **then**
        break
      end if
      $\boldsymbol{u} \leftarrow \boldsymbol{v}$.
    end while
    $\mathcal{M} \leftarrow \mathcal{M} \cup \{\boldsymbol{u}\}$.
  end for
  return a maximizer of a over $\mathcal{M}$.
end function

Algorithm 3 *Ig7O* with COMBO+EHVI for Pareto-front search

Input: Fixed conditioning and initial evaluated designs as in Algorithm 1, now with m property oracles; budget B; graphs $\{G_r\}_{r\in\mathcal{R}}$; candidate-pool size $C \geq 1$; restart count $1 \leq J \leq C$; hypervolume reference point $\boldsymbol{r}_{\mathrm{ref}}$.

Define $G = \square_{r\in\mathcal{R}} G_r$ and precompute the eigensystems of $\{L_r\}_{r\in\mathcal{R}}$.
$n \leftarrow n_0$; initialize $\mathcal{U}_n$ by Eq. (7).
$\mathcal{I}_n \leftarrow \mathrm{ND}(\mathcal{D}_n)$ using Eq. (14).
while $n < B$ **and** $\mathcal{U}_n \neq \varnothing$ **do**
    Fit the graph-GP surrogate $\hat{\boldsymbol{f}}_n$ to the vector-valued observations in $\mathcal{D}_n$.
    $\mathcal{Y}_n^\star \leftarrow \{\boldsymbol{y}(\boldsymbol{t}_i) : i \in \mathcal{I}_n\}$.
    Define $a_n(\boldsymbol{t}; \hat{\boldsymbol{f}}_n)$ as EHVI using Eq. (16).
    $\boldsymbol{t}_{n+1} \leftarrow \textsc{GraphSearch}(a_n(\cdot; \hat{\boldsymbol{f}}_n), G, \mathcal{U}_n, C, J)$.
    $(S_{n+1}, X_{n+1}, \boldsymbol{y}(\boldsymbol{t}_{n+1})) \leftarrow \textsc{Evaluate}(\boldsymbol{t}_{n+1})$.
    Store (S_{n+1}, X_{n+1}).
    $\mathcal{D}_{n+1} \leftarrow \mathcal{D}_n \cup \{(\boldsymbol{t}_{n+1}, \boldsymbol{y}(\boldsymbol{t}_{n+1}))\}$.
    $\mathcal{I}_{n+1} \leftarrow \mathrm{ND}(\mathcal{D}_{n+1})$.
    $\mathcal{U}_{n+1} \leftarrow \mathcal{U}_n \setminus \{\boldsymbol{t}_{n+1}\}$; $n \leftarrow n+1$.
end while
return $\{(\boldsymbol{t}_i, S_i, X_i, \boldsymbol{y}(\boldsymbol{t}_i)) : i \in \mathcal{I}_n\}$.

Generation and oracle evaluation. The shared EVALUATE operation converts a proposed combination into the antibody controls $\mathcal{P}(\boldsymbol{t})$ described in §4.2, runs *Ig7G*, and decodes its output:

$$\begin{aligned}(\mathbf{x}, \mathbf{z}) &= \mathcal{G}_{\theta_{\mathrm{Ab}},\phi}\big(\boldsymbol{\epsilon}; \mathcal{A}, E, \mathcal{P}(\boldsymbol{t})\big),\\ (S, X) &= \mathcal{D}(\mathbf{x}, \mathbf{z}),\\ \boldsymbol{y}(\boldsymbol{t}) &= \big(q_1(S, X), \ldots, q_m(S, X)\big),\end{aligned} \tag{8}$$

where q_j is the selected property oracle, $\boldsymbol{\epsilon}$ is the Gaussian initialization noise, and m is the number of optimization objectives. All scores are oriented so that larger values are preferred; minimization objectives are negated before surrogate fitting and acquisition evaluation. The generator parameters θ_{Ab} and ϕ, as well as the decoder $\mathcal{D}$, remain fixed throughout the search. Each observation refers to the generated realization stored with it, rather than a deterministic outcome of the token combination.

The initial history contains n_0 distinct feasible combinations, evaluated through the same operation, and counts toward the total budget B. The algorithms use the single-candidate update $n \mapsto n+1$ of the main text and terminate when the budget is exhausted or no unevaluated combination remains. Final outputs are selected from stored, oracle-evaluated designs; the selected combinations are not regenerated for reporting.

E.2 TREE ENSEMBLE

Algorithm 1 instantiates tree-ensemble search with a random-forest surrogate and a UCB acquisition function. Each CDR-token combination is represented by the concatenation $e(\boldsymbol{t})$ of its CDR-wise one-hot encodings. Let $h_{n,k}(e(\boldsymbol{t}))$ denote the prediction of tree k in the random forest fitted to $\mathcal{D}_n$. The ensemble mean and spread are

$$\begin{aligned}\mu_n(\boldsymbol{t}) &= \frac{1}{K}\sum_{k=1}^{K} h_{n,k}(e(\boldsymbol{t})),\\ \sigma_n(\boldsymbol{t}) &= \left[\frac{1}{K}\sum_{k=1}^{K}\big(h_{n,k}(e(\boldsymbol{t})) - \mu_n(\boldsymbol{t})\big)^2\right]^{1/2}.\end{aligned} \tag{9}$$

The acquisition score is

$$a_n(\boldsymbol{t}; \hat{f}_n) = \mu_n(\boldsymbol{t}) + \beta\sigma_n(\boldsymbol{t}), \tag{10}$$

where $\beta \geq 0$ controls the preference for exploration. The tree-wise spread serves as an empirical uncertainty heuristic rather than a calibrated posterior standard deviation.

At each iteration, tree-ensemble search ranks a sampled pool of unevaluated combinations by their acquisition scores and selects the highest-ranked candidate for generation and oracle evaluation.

This separates inexpensive surrogate screening from expensive design evaluation without enumerating the full combinatorial space.

E.3 COMBO

Algorithm 2 instantiates *Ig7O* with COMBO (Oh et al., 2019), which combines a graph-based Gaussian-process surrogate with expected improvement. Each factor graph G_r has the eligible tokens in $\mathcal{V}_r$ as vertices. Their Cartesian product G represents joint CDR-token assignments, with each edge changing one CDR token along an edge of its corresponding factor graph. For categorical factors, G_r is a complete graph. Writing L_r for its graph Laplacian, the diffusion kernel takes the factorized form

$$K_r = \exp(-\tau_r L_r), \qquad k(\boldsymbol{t}, \boldsymbol{t}') = \sigma_f^2 \prod_{r \in \mathcal{R}} [K_r]_{t_r, t'_r}, \tag{11}$$

where $\tau_r \geq 0$ is a CDR-specific diffusion scale and $\sigma_f^2 > 0$ is the kernel amplitude. This construction evaluates covariances from the factor graphs without constructing the full product-graph kernel matrix.

Given the scalar predictive distribution $p_n(v \mid \boldsymbol{t})$ supplied by $\hat{f}_n$, EI measures improvement over the best observed score $y_n^\star = \max_{i \leq n} y(\boldsymbol{t}_i)$:

$$a_n(\boldsymbol{t}; \hat{f}_n) = \mathbb{E}_{v \sim p_n(\cdot \mid \boldsymbol{t})}\big[(v - y_n^\star)_+\big], \qquad (u)_+ = \max(u, 0). \tag{12}$$

For a Gaussian predictive distribution with mean $\mu_n(\boldsymbol{t})$ and standard deviation $\sigma_n(\boldsymbol{t})$, this becomes

$$\begin{aligned} a_n(\boldsymbol{t}; \hat{f}_n) = \delta_n(\boldsymbol{t}) \Phi_{\mathrm{N}}\left(\frac{\delta_n(\boldsymbol{t})}{\sigma_n(\boldsymbol{t})}\right) \\ + \sigma_n(\boldsymbol{t}) \varphi_{\mathrm{N}}\left(\frac{\delta_n(\boldsymbol{t})}{\sigma_n(\boldsymbol{t})}\right), \end{aligned} \tag{13}$$

where $\delta_n(\boldsymbol{t}) = \mu_n(\boldsymbol{t}) - y_n^\star$, and Φ_{N} and φ_{N} are the standard normal cumulative distribution and density functions. At zero predictive variance, EI is $(\delta_n(\boldsymbol{t}))_+$. When GP hyperparameters are marginalized, Eq. (12) is averaged over their posterior rather than evaluated using only moment-matched predictions.

COMBO approximately maximizes EI using the GRAPHSEARCH routine in Algorithm 2. Starting from multiple high-scoring candidates, the routine performs greedy moves within the graph neighborhood $\mathcal{N}_G(\boldsymbol{u})$. Only unevaluated feasible combinations are eligible, and a move is accepted only when it strictly increases the acquisition value. The best terminal candidate across the restarts is selected for generation and oracle evaluation. These local searches use surrogate predictions only and do not consume oracle evaluations or guarantee a global acquisition maximizer.

E.4 PARETO-FRONT SEARCH

Algorithm 3 uses the same graph representation while replacing scalar improvement with expected hypervolume improvement (EHVI) (Daulton et al., 2020). For maximization-oriented feedback, $\boldsymbol{u}$ dominates $\boldsymbol{v}$, written $\boldsymbol{u} \succ \boldsymbol{v}$, if $u_j \geq v_j$ for every objective j and $u_j > v_j$ for at least one. The non-dominated archive is indexed by

$$\begin{aligned} \mathcal{I}_n &= \mathrm{ND}(\mathcal{D}_n) \\ &= \{i \in \{1, \ldots, n\} : \nexists k \in \{1, \ldots, n\},\ \boldsymbol{y}(\boldsymbol{t}_k) \succ \boldsymbol{y}(\boldsymbol{t}_i)\}, \end{aligned} \tag{14}$$

with objective-space frontier $\mathcal{Y}_n^\star = \{\boldsymbol{y}(\boldsymbol{t}_i) : i \in \mathcal{I}_n\}$. Both the archive and the frontier contain evaluated outcomes, not surrogate predictions.

For a fixed reference point $\boldsymbol{r}_{\mathrm{ref}}$, define

$$\mathrm{HV}(\mathcal{Y}; \boldsymbol{r}_{\mathrm{ref}}) = \lambda_m \left(\bigcup_{\boldsymbol{y} \in \mathcal{Y}} \{\boldsymbol{v} : \boldsymbol{r}_{\mathrm{ref}} \preceq \boldsymbol{v} \preceq \boldsymbol{y}\} \right), \tag{15}$$

where λ_m is m-dimensional volume and $\preceq$ denotes component-wise comparison. The reference point is specified in the same maximization-oriented score coordinates and held fixed during a search.

Let $p_n(\boldsymbol{v} \mid \boldsymbol{t})$ be the predictive law supplied by the graph-based surrogate $\hat{\boldsymbol{f}}_n$. EHVI is

$$a_n(\boldsymbol{t}; \hat{\boldsymbol{f}}_n) = \mathbb{E}_{\boldsymbol{v} \sim p_n(\cdot \mid \boldsymbol{t})} \Big[\mathrm{HV}\big(\mathcal{Y}_n^\star \cup \{\boldsymbol{v}\}; \boldsymbol{r}_{\mathrm{ref}}\big) - \mathrm{HV}\big(\mathcal{Y}_n^\star; \boldsymbol{r}_{\mathrm{ref}}\big) \Big]. \tag{16}$$

The expectation accounts for predictive uncertainty; evaluating hypervolume improvement only at the predicted mean is not EHVI. For Monte Carlo evaluation, the expectation is approximated by averaging the hypervolume increments of samples from $p_n(\cdot \mid \boldsymbol{t})$.

The selected combination is generated and scored on all objectives. Its observed feedback updates the surrogate training history and the non-dominated archive. The final output is a set of evaluated antibody designs with different objective trade-offs, rather than a single winner under a fixed weighted sum. This is an empirical Pareto approximation over the evaluated designs, not a claim of global Pareto optimality over Ω_{Ab}.

Search configurations. All search methods use the same evaluation schedule: $n_0 = 48$ evaluations in round 0, followed by four additional evaluations per round. Thus, R post-initialization rounds correspond to a total budget of $B = 48 + 4R$ evaluations, including initialization. Each evaluation generates a single antibody and computes its feedback; surrogate-based candidate screening does not consume this budget. The single-candidate notation in Algorithms 1–3 indexes individual evaluations rather than experimental rounds.