AF3DESIGN: SELECTIVITY-AWARE NANOBODY DESIGN WITH ALPHAFOLD3

Robin Pan[1*] **Anisha Parsan**[2*] **Caroline Uhler**[2,3] **Wengong Jin**[3,4†]
[1] John A. Paulson School of Engineering and Applied Sciences, Harvard University
[2] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology
[3] Broad Institute of MIT and Harvard
[4] Khoury College of Computer Sciences, Northeastern University

ABSTRACT

Designing antibodies that discriminate between highly similar targets represents a fundamental challenge in computational protein engineering, particularly for T cell receptor-like (TCR-like) nanobodies targeting peptide-MHC complexes. We present *AF3Design*, the first computational framework for nanobody design that integrates binding selectivity directly into the optimization process rather than relying on post-hoc filtering. Our approach leverages AlphaFold3's structure prediction capabilities through a gradient-free genetic algorithm, optimizing a composite fitness function with explicit off-target penalties. These terms include contact-based rewards for discriminating residues and $\Delta\Delta$G-based differential binding predictions. We validate our method on the challenging task of designing nanobodies against peptide-MHC complexes that differ by as little as a single residue. AF3Design significantly improves peptide-facing interface confidence over baselines and enriches binders that preferentially contact the discriminating peptide residue(s), with further $\Delta\Delta$G gains after staged activation. Our results show that optimizing selectivity during search yields highly discriminative candidates suitable for precise molecular recognition.

1 INTRODUCTION

Therapeutic antibodies have transformed modern medicine, particularly in cancer immunotherapy and the treatment of autoimmune diseases. Among therapeutic antibody formats, T-cell receptor(TCR)-like nanobodies represent a clinically important yet exceptionally challenging class. These T-cell receptors are in part characterized by hypervariable loops known as complementarity determining regions (CDRs), which recognize intracellular peptides displayed in peptide-Major Histocompatibility Complex I (pMHC) complexes. TCRs possess high specificity (Du et al. (2025)), and thus TCR-like nanobodies must achieve similar discrimination, distinguishing target from off-target peptides that may differ by as little as a single amino acid to avoid autoimmune toxicity.

Current approaches to protein binder design generally fall under two main paradigms. Generative methods such as RFDiffusion and BoltzGen employ diffusion or flow-based models to generate structures in a single forward pass, followed by post-hoc filtering (Watson et al. (2023), Stark et al. (2025)). Alternatively, hallucination based methods like BindCraft and Germinal directly optimize design objectives through iterative refinement, using structure prediction models as oracles to guide the design process (Pacesa et al. (2025), Mille-Fragoso et al. (2025)). While these paradigms have demonstrated initial success, two critical gaps remain. First, existing methods address binding specificity primarily through post-hoc filtering rather than direct optimization, yielding suboptimal selectivity. Second, no current methods leverage AlphaFold3, which achieves superior antibody structure prediction and confidence metrics (Peng et al. (2025), Hitawala & Gray (2024)), but whose diffusion-based architecture makes gradient-based optimization computationally intractable.

We present *AF3Design*, a computational framework for nanobody design through gradient-free evolutionary search with integrated selectivity optimization. Our method utilizes a composite fitness

*Equal contribution
†w.jin@northeastern.edu

function derived from AlphaFold3 confidence metrics without requiring backprop. Critically, we incorporate binding selectivity directly into the optimization loop through two mechanisms: contact-based rewards and $\Delta\Delta$G-based differential binding predictions. Across KRAS, p53, and WT1 pMHC targets, AF3Design substantially improves peptide-facing interface confidence over baseline and enriches binders that preferentially contact the discriminating residue(s), demonstrating that optimizing selectivity during search can achieve single-residue-level discrimination.

2 METHODS

We design TCR-like nanobodies for pMHC targets under a fixed binder–peptide–MHC template, optimizing amino acid identities over a design mask D (e.g., CDR1–3) while holding the nanobody framework, peptide, and HLA sequence fixed. We consider a *target* peptide and a closely related *off-target* peptide differing by as little as one substitution, and seek binders that (i) form confident peptide-facing interfaces and (ii) avoid peptide-independent contacts on non-canonical MHC surfaces.

AF3DESIGN performs black-box optimization over discrete sequences with AlphaFold3 (AF3) as a structure-and-metrics oracle. Each candidate sequence is predicted in the context of the full complex, evaluated by a forward pass through AF3, and assigned a composite fitness score that jointly rewards fold/interface confidence and selectivity. Selectivity is optimized during search using contact-based terms and a staged $\Delta\Delta$G discrimination term.

2.1 GRADIENT-FREE EVOLUTIONARY OPTIMIZATION WITH AF3

Our genetic algorithm maintains a population of P candidate sequences and iteratively refines them over T generations. At initialization, residues in D are sampled uniformly from the 19 allowed amino acids (excluding Cys). Each generation selects high-fitness parents, applies mutation and single-point crossover within the span of D which generates chimeric offspring with complementary subsequences. Offspring are evaluated with AF3, and the top-K sequences from the combined pool are retained for the next generation. See A.1 for the full algorithm.

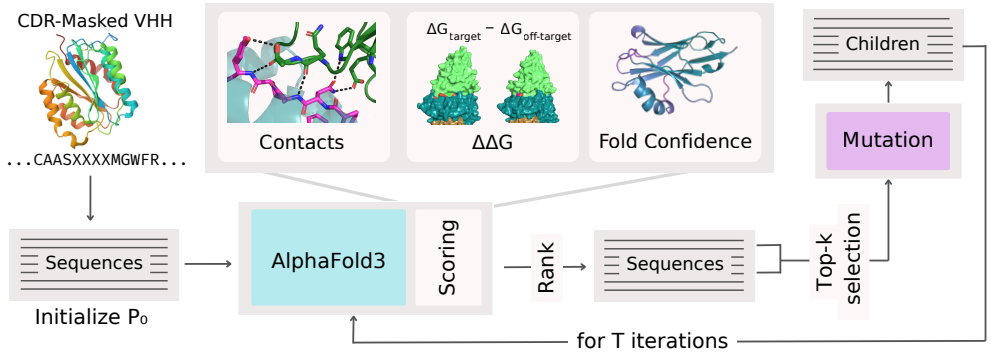

Figure 1: **AF3Design Framework**. AF3Design optimizes nanobody sequences through gradient-free evolutionary search. CDR residues are initialized randomly, then iteratively refined over T generations using mutation and crossover. Candidates are evaluated via AlphaFold3, scoring fold confidence, contact geometry, and predicted $\Delta\Delta$G. Top-k selection retains high-fitness sequences for the next generation.

2.2 MULTI-OBJECTIVE FITNESS SCORING

We assign each candidate sequence s a scalar fitness $F(s)$ that combines fold quality, interface confidence, and geometry-level selectivity terms derived from AF3 outputs. We use a weighted sum:

$$F_t(s) = \sum_i w_i \, \mathcal{L}_i(s) \; + \; \mathbf{1}[t \geq t_{\text{ddG}}] \; w_{\text{ddG}} \, \mathcal{L}_{\text{ddG}}(s). \tag{1}$$

All terms are computed from a single AF3 forward pass, except the staged $\Delta\Delta$G term (Section 2.2). Weights and full definitions are in Table 1 and A.2.

Fold and Interface Confidence. We include confidence-based proxies for plausible folds and well-posed binder–peptide interfaces, such as CDR-restricted pLDDT, intra-chain pAE, and nanobody–peptide interface pAE. We additionally include an interface-specific ipTM computed between the nanobody and peptide chains to avoid dilution from whole-complex metrics.

Geometry-Level Specificity To explicitly drive interactions with specificity-determining peptide residues, we reward contact probability between CDR residues (design mask D) and a predefined set of peptide target positions R. Crucially, R is the residues differing between on-target and off-target peptide sequences. Adapted from Pacesa et al. (2025), contact probabilities are computed from AF3 distograms by integrating probability mass below a distance cutoff. To discourage peptide-independent docking and non-canonical binding on the membrane-proximal $\alpha_3/\beta_2\mathrm{m}$ face, we additionaly penalize close contacts between CDR residues D and a predefined MHC avoidance set M. We use a min-k aggregation to reward the strongest CDR–peptide contacts and a complementary max aggregation to penalize the most confident off-target CDR–MHC interaction.

$\Delta\Delta G$-Based Energetic Specificity Geometry-level objectives encourage engagement with target residues but do not directly penalize binding to off-target peptides. To encode explicit discrimination, we incorporate predicted binding free energy differences $\Delta\Delta G$ using BA-DDG, a pretrained model that estimates changes in binding free energy upon mutation (Jiao et al. (2024)). For each candidate nanobody, we compute the predicted $\Delta\Delta G$ between the target peptide and the off-target peptide. Positive values indicate preferential binding to the target and are rewarded. We enable $\Delta\Delta G$ scoring only after t_{ddG} generations so early search prioritizes establishing plausible binding modes, while later search applies energetic discrimination pressure once interfaces are structured.

3 Experiments

Task. We evaluate AF3DESIGN on three therapeutically relevant pMHC targets from tumor-associated antigens (TP53, KRAS, and WT1), each paired with an off-target peptide differing by 1–3 amino acids. We assess whether our method produces (i) well-formed nanobody–pMHC interfaces and (ii) peptide-discriminative binders. We optimize using the multi-objective fitness described in A.2 and instantiate designs using a fixed nanobody framework scaffold (A.4), designing only CDR residues. Since strong candidates may appear transiently during exploration without surviving to the final population, we output the top-n scoring unique sequences across the full evolutionary trajectory rather than only the final generation.

Baseline. We compare against Germinal as a baseline because it is one of the few open-sourced methods that (i) supports VHH/nanobody-constrained design and (ii) can operate on multi-chain targets. Germinal includes intermediate filtering heuristics that can terminate trajectories early when partial designs fail structural or interface criteria. With these default filters enabled, we observed no successful trajectories under the default weights for our targets given the same compute budget as AF3DESIGN; we therefore disable intermediate filtering. To budget-match across methods, we run both Germinal and AF3DESIGN for $N = 11,000$ total oracle calls, which corresponds to approximately 100 independent trajectories per target in Germinal. We therefore retain the 100 final output structures of these trajectories produced by the Germinal pipeline.

3.1 De Novo Nanobodies Form Confident pMHC Binding Interfaces Across Targets

Figure 2 compares 100 designs from AF3DESIGN to the 100 designs output by Germinal across all three targets. AF3DESIGN significantly improves interface confidence and binding geometry metrics relative to Germinal on each pMHC system, including peptide-facing interface confidence (peptide–binder ipTM, peptide–binder pAE), and CDR-restricted confidence (pLDDT). Specifically, we achieve median peptide–binder ipTM values of 0.83, 0.60, and 0.79, representing $11\times$, $5\times$, and $10\times$ improvements over the baseline. We additionally observe remarkably higher peptide-contact density via CDR3–peptide contacts and improved P-dockQ trends, consistent with more confident, well-packed interfaces. Specificity improvements are further reflected in the $\Delta\Delta G$ discrimination objective. $\Delta\Delta G$ improves on 2 out of 3 targets, with absolute improvement of 1.49.

3.2 AF3DESIGN OPTIMIZES SINGLE-RESIDUE DISCRIMINATION IN BINDING

We plot the $\Delta\Delta G$ predictions of the 100 highest-scoring designs before and after incorporating the $\Delta\Delta G$ prediction term into the objective function (Fig. 3B). Our specificity-aware objective function effectively improves predicted $\Delta\Delta G$ values, though the extent of optimization varies by target. Additional tuning of the specificity weight for individual targets may yield further improvements. To validate design specificity, we quantified contacts to the residue(s) distinguishing on-target and off-target peptides. Specific binders should preferentially contact these residues when bound to the on-target peptide while making fewer or no contacts with the off-target peptide. Consistent with this expectation, Fig. 3A reports a greater proportion of AF3DESIGN binders exhibit more on-target than off-target contacts, whereas many of Germinal-designed binders make zero contacts with the distinguishing residue(s).

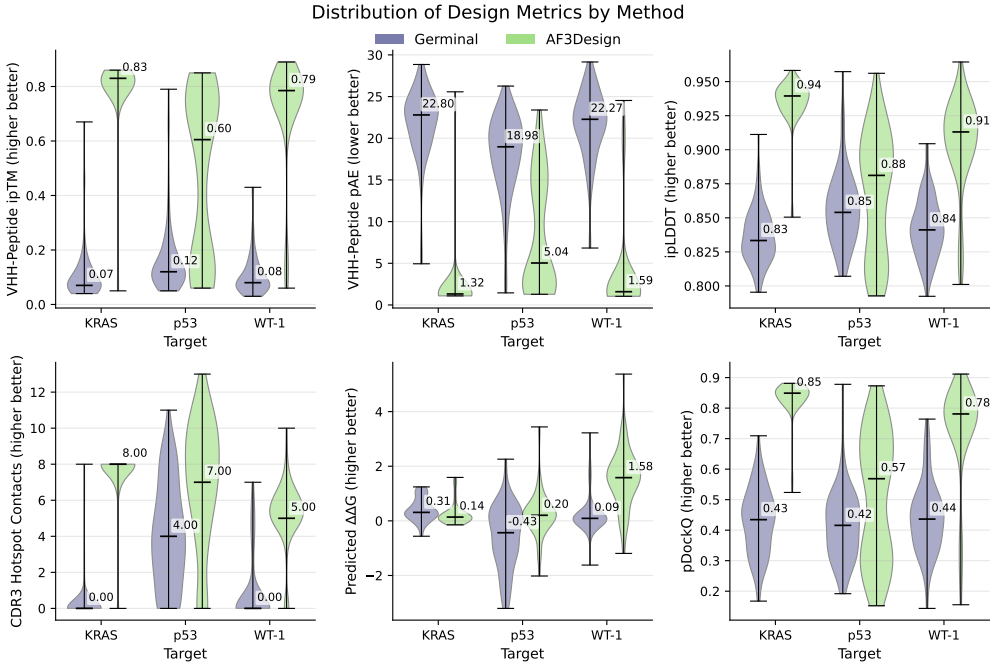

Figure 2: **AF3DESIGN vs baseline performance on relevant in-silico metrics.** For each target, we compare distributions over 100 designs from each method, including VHH-peptide ipTM, ipLDDT, ipAE, pDockQ, and the number of contacts between CDR3 residues and the peptide.

Figure 4 depicts AlphaFold3-predicted structures of selected AF3DESIGN nanobodies bound to KRAS, p53, and WT1 pMHC complexes. These structures illustrate the specificity of our designs, showing more contacts to on-target peptides compared to highly similar off-target candidates.

Across most designs, peptide contacts are primarily mediated by CDR3, a pattern also observed in natural T-cell receptors (Rudolph & Wilson (2002)). The CDR3 of most designs comprises a small helix that contacts the MHC peptide groove and unstructured loops that contact the peptide. We hypothesize this helix-loop combination enables binders to balance the conformational flexibil-

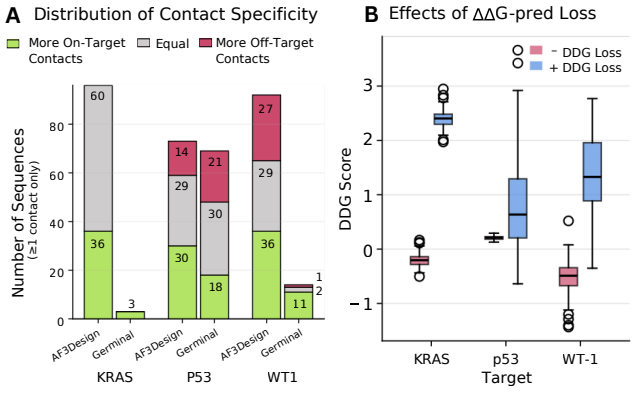

Figure 3: (A) Sequences with more / equal / fewer contacts to the mutating peptide residue(s) in the on-target complex compared to the off-target complex. Only sequences with $\geq$1 contact are visualized. (B) $\Delta\Delta G$ values for the top-100 designs at the final pre-activation iteration versus after $\Delta\Delta G$ loss is enabled.

ity that aids binding in natural TCRs
with the structural stability provided
by the helix, thereby reducing the cross-reactivity associated with highly flexible loops (Scott et al. (2011)).

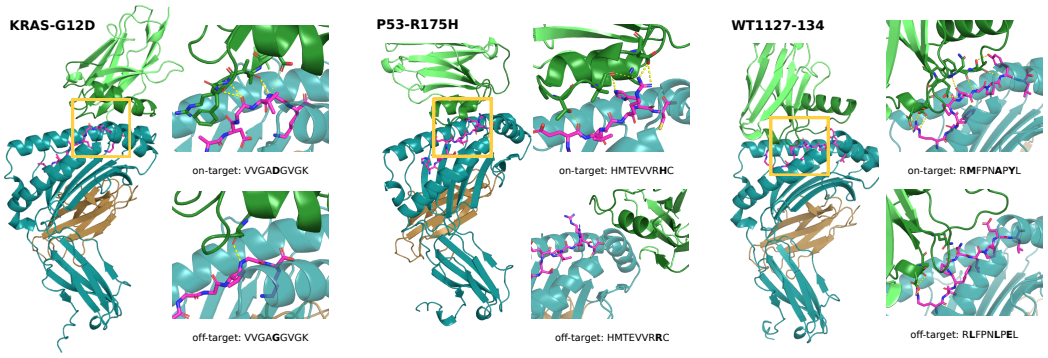

Figure 4: **AlphaFold3-predicted structures of selected AF3Design binders** Nanobody binders are shown in light green with CDRs highlighted in dark green, MHC are teal, and peptides are pink. Dashed yellow lines indicate predicted polar contacts. Designs are predicted to have preferential contacts with on-target peptides compared to off-target peptides.

3.3 Evolutionary Optimization Efficiently Explores CDR Sequence Space

To assess whether our genetic algorithm meaningfully improves binder quality (rather than trading off desirable properties), we track key interface metrics across the evolutionary trajectory and observe a steady distributional shift toward higher-confidence binding over iterations. In Fig. 5A both peptide-facing ipTM, VHH-peptide ipAE, and $\Delta\Delta$G improve monotonically in the population, indicating that selection pressure induced by our AF3-derived objective consistently enriches sequences with desired nanobody–peptide interfaces. Moreover, the first emergence of top-scoring candidates occurs predominantly at later generations Fig. 5B, suggesting that high-quality binders are discovered through effective exploration and exploitation rather than being present at initialization.

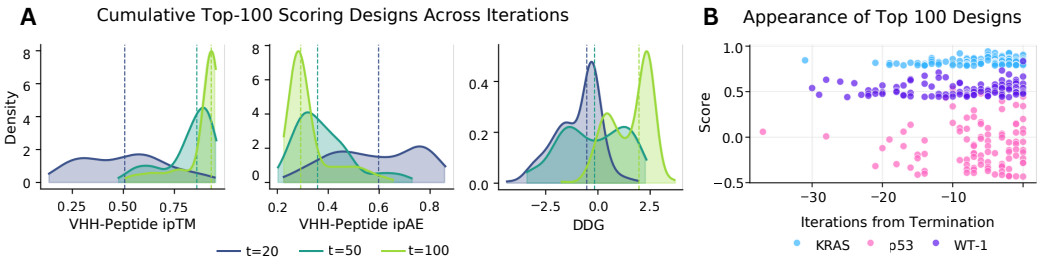

Figure 5: (A) Distributions of ipTM, ipAE, and $\Delta\Delta G$ for the top-100 scoring designs among candidates of population at various iterations. WT-1 excluded from figure as multiple peptide mutations change the BA-$\Delta\Delta G$ scale, preventing 1:1 comparisons. (B) Iteration (relative to termination) at which each of the final top-100 scoring designs first appears during evolution.

4 Conclusion

We present AF3Design, a specificity-aware framework that integrates selectivity directly into gradient-free sequence optimization with AlphaFold3, enabling de-novo TCR-like nanobodies that form confident peptide-facing interfaces while discriminating peptides differing by as little as a single residue. Across KRAS, p53, and WT1, our evolutionary search yields large improvements in interface confidence and peptide-contact geometry, and predicted binding energies. These results

present several promising directions for future work. Although we tested alternative gradient-free optimization approaches, we ultimately favored our implementation for computational efficiency; given these promising results, we expect that adaptive re-weighting schedules or learned proposal policies (RL- or bandit-guided mutations) could further improve exploration of the CDR sequence space. Finally, because our objective is defined over discrete sequences and computed from inexpensive, precomputed signals, the framework can readily accommodate additional low-cost terms. We note that these computational predictions must be validated experimentally, and we are actively developing an in vitro testing campaign to assess binding and specificity.

5 CODE AVAILABILITY

The code for AF3Design and documentation on how to design specifications for nanobody binder design is available at https://github.com/aparsanMIT/AF3Design

REFERENCES

Yehlin Cho, Martin Pacesa, Zhidian Zhang, Bruno E Correia, and Sergey Ovchinnikov. Boltzdesign1: Inverting all-atom structure prediction model for generalized biomolecular binder design. *bioRxiv*, pp. 2025–04, 2025.

H. Du, L. Mallik, D. Hwang, and et al. Targeting peptide antigens using a multiallelic mhc i-binding system. *Nature Biotechnology*, 43:1683–1693, 2025. doi: 10.1038/s41587-024-02505-8.

Fatima N. Hitawala and Jeffrey J. Gray. *What does alphafold3 learn about antigen and nanobody docking, and what remains unsolved?*, Sep 2024. doi: 10.1101/2024.09.21.614257.

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H Moal. Skempi 2.0: An updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, Jul 2018. doi: 10.1093/bioinformatics/bty635.

Xiaoran Jiao, Weian Mao, Wengong Jin, Peiyuan Yang, Hao Chen, and Chunhua Shen. Boltzmann-aligned inverse folding model as a predictor of mutational effects on protein-protein interactions, 2024. URL https://arxiv.org/abs/2410.09543.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, and et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Jul 2021. doi: 10.1038/s41586-021-03819-2.

H.T. Lee and Y.-S. Heo. *High-resolution structure of vWF A1 domain in complex with caplacizumab, the first nanobody-based medicine*, Jul 2021. doi: 10.2210/pdb7eow/pdb.

Bingxu Liu, Nathan F. Greenwood, Julia E. Bonzanini, Amir Motmaen, Jeremy Meyerberg, Tao Dao, Xinyu Xiang, Russell Ault, Jazmin Sharp, Chunyu Wang, and et al. Design of high-specificity binders for peptide–mhc-i complexes. *Science*, 389(6758):386–391, Jul 2025. doi: 10.1126/science.adv0185.

Qiaozhen Meng, Fei Guo, and Jijun Tang. Improved structure-related prediction for insufficient homologous proteins using msa enhancement and pre-trained language model. *Briefings in Bioinformatics*, 24(4), Jun 2023. doi: 10.1093/bib/bbad217.

Luis Santiago Mille-Fragoso, John N. Wang, Claudia L. Driscoll, Haoyu Dai, Talal M. Widatalla, Xiaowe Zhang, Brian L. Hie, and Xiaojing J. Gao. Efficient generation of epitope-targeted de novo antibodies with germinal. *bioRxiv preprint*, bioRxiv:2025.09.19.677421, 2025. arXiv, https://www.biorxiv.org/content/10.1101/2025.09.19.677421v1.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: Making protein folding accessible to all. *Nature Methods*, 19(6):679–682, May 2022. doi: 10.1038/s41592-022-01488-1.

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*, August 27 2025. doi: 10.1038/s41586-025-09429-6.

Chunxiang Peng, Wentao Ni, Quancheng Liu, Gang Hu, and Wei Zheng. A comprehensive benchmarking of the alphafold3 for predicting biomacromolecules and their interactions. *Briefings in Bioinformatics*, 26(6), Nov 2025. doi: 10.1093/bib/bbaf616.

Markus G Rudolph and Ian A Wilson. The specificity of tcr/pmhc interaction. *Current Opinion in Immunology*, 14(1):52–65, Feb 2002. doi: 10.1016/s0952-7915(01)00298-9.

Daniel R. Scott, Oleg Y. Borbulevych, Kurt H. Piepenbrink, Steven A. Corcelli, and Brian M. Baker. Disparate degrees of hypervariable loop flexibility control t-cell receptor cross-reactivity, specificity, and binding mechanism. *Journal of Molecular Biology*, 414(3):385–400, Dec 2011. doi: 10.1016/j.jmb.2011.10.006.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, and et al. *Boltzgen: Toward universal binder design*, Nov 2025. doi: 10.1101/2025.11.20.689494.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, and et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, Jul 2023. doi: 10.1038/s41586-023-06415-8.

A Appendix

A.1 Genetic Algorithm

Algorithm 1 AF3Design Evolutionary Optimization

Require: Design mask D, population P, generations T, mutation probability μ, crossover probability p_c, activate $\Delta\Delta G$ at t_{ddG}

Initialize $P_0 = \{s_p\}_{p=1}^{P}$ by sampling residues on D (exclude Cys); evaluate $F(s)$ via AF3
for $t = 1 \dots T$ **do**
  $P \leftarrow \mathrm{TopK}(P_{t-1})$; $O \leftarrow \emptyset$
  for each $p \in P$ **do**
    $O \leftarrow O \cup \{\mathrm{Mutate}(p; D, \mu)\}$
  end for
  Sample $N_c = \lfloor K p_c \rfloor$ parent pairs (p_1, p_2) from P
  for each pair (p_1, p_2) **do**
    Sample breakpoint x uniformly within span(D); add $\mathrm{Crossover}(p_1, p_2; x)$ to O
  end for
  Subsample O to $|O| = K$; evaluate $F(s)$ for all $s \in O$ (include $\sigma(\Delta\Delta G)$ if $t \geq t_{\mathrm{ddG}}$)
  $P_t \leftarrow \mathrm{TopK}(P \cup O)$
end for
return P_T

A.2 Score Formulation

Notation and Origin Let A denote nanobody residues (chain A), $D \subset A$ the design mask (CDR positions), P peptide residues (chain B), $T \subset P$ the discriminating peptide positions (e.g., the mutated residue(s)), and M the MHC avoidance set (selected HLA residues to discourage non-canonical contacts).

AlphaFold3 provides per-residue pLDDT, a pAE matrix pAE_{ij} (max 31 Å), and a distogram over distance bins for each residue pair (i, j).

Contact probabilities from distograms. We follow Pacesa et al. (2025) and Cho et al. (2025) in defining the probability of a contact within cutoff c for any residue pair (i, j) as

$$P(d_{ij} < c) \;=\; \sum_{b:\, b<c} \mathrm{softmax}(\boldsymbol{\ell}_{ij})_b, \tag{2}$$

where $\boldsymbol{\ell}_{ij}$ are the distogram logits and bins b correspond to distance intervals below c.

Sparse aggregators. Given a multiset of values $\{x_r\}_{r=1}^{R}$ and an integer k:

$$\min\text{-}k(\{x_r\}) = \frac{1}{k} \sum_{r \in \mathcal{I}_k^{\min}} x_r, \quad \mathcal{I}_k^{\min} = \text{indices of the } k \text{ smallest values}, \tag{3}$$

$$\max\text{-}k(\{x_r\}) = \frac{1}{k} \sum_{r \in \mathcal{I}_k^{\max}} x_r, \quad \mathcal{I}_k^{\max} = \text{indices of the } k \text{ largest values}. \tag{4}$$

We use $\min$-k to emphasize the strongest (most confident) desired contacts, and $\max$-k to penalize the worst (most confident) undesired contacts so single pathological interactions are not averaged away.

Full Fitness Function and Weights For the experiments in this work, we maximize a scalar fitness $F(s)$ for candidate sequence s computed by the sum of the terms in 1 multiplied by their respective weightages. The weights for a portion our score terms were based on the default weights for analogous terms in Mille-Fragoso et al. (2025) in the VHH design pipeline. Further exploration and tuning of weights may improve convergence.

Term	Weight
$\mathrm{ipTM}_{A \leftrightarrow B}$	$+0.7$
$\mathcal{L}_{\mathrm{pLDDT}}$	-1.0
$\mathcal{L}_{\text{i-pLDDT}}$	-1.0
$\mathcal{L}_{\text{i-pAE}}$	-0.5
$\mathcal{L}_{\text{pAE-intra}}$	-0.1
$\mathcal{L}_{\text{con-intra}}$	-0.1
$\mathcal{L}_{\mathrm{Rg}}$	-0.1
$\mathcal{L}_{\text{paratope}}$	-0.2
$\mathcal{L}_{\text{target-contact}}$	-0.5
$\mathcal{L}_{\text{peptide-contact}}$	-0.5
$\mathcal{L}_{\text{repulsion}}$	-0.5
$\Delta\Delta G_{\text{pred}}$ (if $t \geq t_{\text{ddG}}$)	$+0.7$

Table 1: Objective components and weights.

A.2.1 Fold and Interface Confidence-related Score Terms

The AF3-derived scores described in A.2.1 are based on the AF2-derived analogs defined as "losses" in Pacesa et al. (2025) and Mille-Fragoso et al. (2025), though they differ in exact implementation due to binning differences in AF3 and AF2.

Confidence & geometry losses. Let A be nanobody residues, B the peptide chain, $D \subseteq A$ the CDR/design mask, and P the peptide residue indices. Define normalized scores $\tilde{p}_i = \mathrm{pLDDT}_i/100$ and $\tilde{e}_{ij} = \mathrm{pAE}_{ij}/31$. Then:

$$\mathcal{L}_{\mathrm{ipTM}} = -\mathrm{ipTM}_{A \leftrightarrow B}, \tag{5a}$$

$$\mathcal{L}_{\text{i-pLDDT}} = 1 - \frac{1}{|D|} \sum_{i \in D} \tilde{p}_i, \qquad \mathcal{L}_{\mathrm{pLDDT}} = 1 - \frac{1}{|A|} \sum_{i \in A} \tilde{p}_i, \tag{5b}$$

$$\mathcal{L}_{\text{i-pAE}} = \frac{1}{|D||P|} \sum_{i \in D} \sum_{j \in P} \tilde{e}_{ij}, \qquad \mathcal{L}_{\text{pAE-intra}} = \frac{1}{|A|^2} \sum_{i,j \in A} \tilde{e}_{ij}. \tag{5c}$$

Radius of gyration penalty. Let $N = |A|$, coordinates $\mathbf{r}_i$ (Cα), and $\bar{\mathbf{r}}$ the centroid:

$$R_g = \sqrt{\frac{1}{N} \sum_{i \in A} \|\mathbf{r}_i - \bar{\mathbf{r}}\|^2}, \tag{6}$$

$$R_g^{\text{expected}} = 2.38 \cdot N^{0.365}, \tag{7}$$

$$\mathcal{L}_{\mathrm{Rg}} = \mathrm{ELU}\big(R_g - R_g^{\text{expected}}\big), \tag{8}$$

Intra-chain contact loss. Using cutoff $c = 14$ Å and excluding near-diagonal pairs $|i - j| \leq 6$:

$$\mathcal{L}_{\text{con-intra}} = \text{min-}k(\{-\log P(d_{ij} < c) : i, j \in A,\ |i - j| > 6\})\,. \tag{9}$$

Paratope loss. We use the same formulation as proposed in Mille-Fragoso et al. (2025). Let $\mathcal{L}_{\text{CDR}}$ and $\mathcal{L}_{\text{framework}}$ denote the CDR–epitope and framework–epitope contact losses, respectively; we define

$$\mathcal{L}_{\text{paratope}} = \frac{\mathcal{L}_{\text{CDR}}^2}{\max(\mathcal{L}_{\text{framework}} - \lambda, \epsilon)}. \tag{10}$$

A.2.2 Specificity-related Score Terms

Targeted discriminating-residue contact. With cutoff $c = 8$ Å and $k_1 = 2$:

$$\mathcal{L}_{\text{target-contact}} = \frac{1}{|T|} \sum_{t \in T} \text{min-}k_1(\{-\log P(d_{ti} < c) : i \in D\})\,. \tag{11}$$

Full peptide contact. With cutoff $c = 8$ Å:

$$\mathcal{L}_{\text{peptide-contact}} = \frac{1}{|P|} \sum_{p \in P} \text{min-}k(\{-\log P(d_{pi} < c) : i \in D\})\,. \tag{12}$$

MHC avoidance / repulsion. With cutoff $c = 8$ Å and $k = 2$:

$$\mathcal{L}_{\text{repulsion}} = \frac{1}{|M|} \sum_{m \in M} \text{max-}k(\{-\log(1 - P(d_{mi} < c)) : i \in D\})\,, \tag{13}$$

which grows as the probability of undesired CDR–MHC contact increases.

Predicted differential binding. BA-DDG provides a predicted binding free energy change between mutant (off-target) and wildtype (on-target) peptides:

$$\Delta\Delta G_{\text{pred}} = \Delta G_{\text{mutant}} - \Delta G_{\text{wildtype}}, \tag{14}$$

where, following SKEMPI convention (Jankauskaitė et al. (2018)), positive $\Delta\Delta G_{\text{pred}}$ indicates preferential binding to the wild-type. Since the magnitude of $\Delta\Delta G_{\text{pred}}$ may vary based on number of mutations between the on-target and off-target peptides, we scale $\Delta\Delta G_{\text{pred}}$ using the sigmoid function, $\frac{1}{1+e^{-\Delta\Delta G_{\text{pred}}}}$, when calculating the total score.

A.3 Run Settings

Genetic algorithm hyperparameters. We use a population size of $K = 100$ and run the evolutionary search for $T = 80$ generations. We apply single-point crossover with rate $p_c = 0.7$ and initialize the population with $N_0 = 100$ sequences. Unless otherwise noted, residues in the design mask are initialized uniformly over the 19 amino acids excluding cysteine. The mutation operator and rate are described in Algorithm 1; in addition to random initialization, the method can be seeded with user-provided sequences to warm-start optimization.

BA-DDG settings We use BA-DDG with its default inference configuration (reporting only non-training parameters):

- `ca_only=False`,
- `hidden_dim=128`,
- `num_layers=3`,
- `backbone_noise=0.0`,
- `num_edges=48`,
- `loss_weight_boltzmann=1.0`

AlphaFold3 Recycles Recycling in AlphaFold has been shown to improve model accuracy (Jumper et al. (2021)). To choose a recycle count that balances runtime and prediction accuracy, we evaluated 400 sequences from an earlier evolution run at multiple recycle settings and compared their resulting fitness values to the 10-recycle baseline. Lower recycle counts correlate substantially less well with the 10-recycle baseline (e.g., $r = 0.44$ at 2 recycles; $r = 0.80$ at 5; $r = 0.94$ at 8, all $p < 0.0001$), while 10 recycles is self-consistent ($r = 1.0$,$p < 0.0001$). We therefore use 10 recycles for all AlphaFold3 forward passes reported in this work.

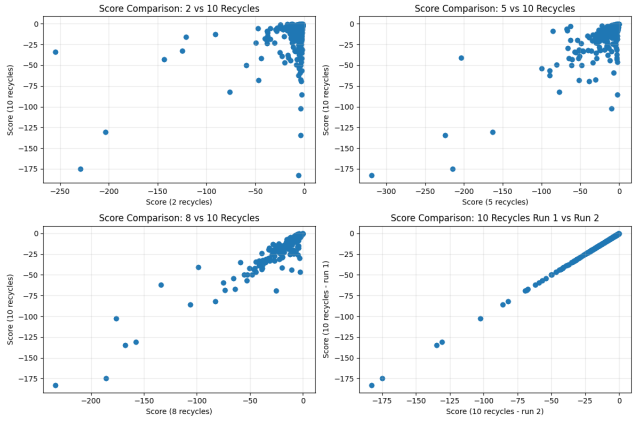

Figure 6: Correlation of fitness scores calculated using 10 recycles compared to 2 recycles, 5 recycles, 8 recycles and a rerun with 10 recycles.

Folding with Binder MSA Folding with MSAs are critical to AlphaFold accuracy (Jumper et al. (2021),Meng et al. (2023)), but is a time-intensive process. To prevent unnecessary MSA searches at every forward pass, we precalculate MSAs using Colabfold-Search (Mirdita et al. (2022)). However, since the nanobody sequence varies, this would typically require an MSA search for every generated sequence. We reasoned that since the majority of the nanobody sequence are framework regions which remain unchanged, the evolutionary information contained in a MSA run on the input input nanobody sequence with X residues in the CDR regions would sufficiently approximate an MSA on the full sequence. This strategy avoids repeated searches while retaining framework-derived evolutionary signal, substantially improving throughput for AF3DESIGN.

A.4 NANOBODY FRAMEWORK AND MHC SEQUENCES

All binders are designed with the following nanobody template from Lee & Heo (2021), with X denoting designable CDR regions:

```
EVQLVESGGGLVQPGGSLRLSCAASXXXXXXXXMGWFRQAPGKGRELVAAXXXXXXXXYYPDSVEGR
FTISRDNAKRMVYLQMNSLRAEDTAVYYCXXXXXXXXXXXXXXXXXXXXXXXXXXWGQGTQVTVSS
```

We use the KRAS G12D variant: `VVGADGVGK`, with the wild-type sequence as the off target: `VVGAGGVGK`. The peptide is folded with HLA-A*1101 of the MHC Class I:

```
GSHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYWDQETRNVK
AQSQTDRVDLGTLRGYYNQSEDGSHTIQIMYGCDVGPDGRFLRGYRQDAYDGKDYIALNEDLRSWTAA
DMAAQITKRKWEAAHAAEQQRAYLEGRCVEWLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCW
ALGFYPAEITLTWQRDGEDQTQDTELVETRPAGDGTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTL
RWELSSQPT
```

We use the TP53 R175H variant: `HMTEVVRHC`, with the wild-type sequence as the off target: `HMTEVVRRC`. The peptide is folded with HLA-A*0201 of the MHC Class I:

```
GSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFDSDAASQRMEPRAPWIEQEGPEYWDGETRKVK
AHSQTHRVDLGTLRGYYNQSEAGSHTVQRMYGCDVGSDWRFLRGYHQYAYDGKDYIALKEDLRSWTAA
DMAAQTTKHKWEAAHVAEQLRAYLEGTCVEWLRRYLENGKETLQRTDAPKTHMTHHAVSDHEATLRCW
```

ALSFYPAEITLTWQRDGEDQTQDTELVETRPAGDGTFQKWAAVVVPSGQEQRYTCHVQHEGLPKPLTL
RWEPSSQPTIPI

We use the WT1127-134 peptide: RMFPNAPYL, with a ARHGEF11 peptide as the off target as identified in Liu et al. (2025): RLFPNLPEL. The peptide is folded with HLA-A*0201 of MHC Class I (see sequence above).

For all analysis, the pMHC complex is also folded with Beta-2-microglobulin (β_2m):

IQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTE
FTPTEKDEYACRVNHVTLSQPKIVKWDRDM