Agent-Guided De Novo Design of Nanobody Binders Against a Novel Cancer Target

Yue Zhao [*1] **Melih Yilmaz** [*1] **Edward Lee** [*1] **Chuanyui Teh** [*1] **Lan Guo** [1]
Kemal Sonmez [1] **Luca Giancardo** [1] **Gordon Trang** [1] **Fangda Xu** [1] **Madelyn Espinosa-Cotton** [2]
Nai-Kong V Cheung [†2] **Jiwon Kim** [†1] **Xinyun (Nina) Cheng** [†1]

Abstract

Therapeutic antibody discovery remains slow and resource-intensive, with traditional methods providing limited control over epitope selection. We present an agent-guided workflow for de novo nanobody design against a novel Desmoplastic Small Round Cell Tumor target, encompassing epitope identification via our hotspot recommendation agent, de novo generation using three independent methods (RFantibody, IgGM, mBER) across multiple predicted antigen structures, multi-metric scoring including structural and sequence-based binding affinity metrics, and high-throughput yeast surface display screening followed by surface plasmon resonance (SPR) characterization. From 288,000 designs spanning eight epitope regions and three VHH frameworks, Pareto-based filtering selected 100,000 candidates; of 116 enriched candidates advanced to SPR, 46 (39.7%) produced reliable kinetic fits ($R_{\max} \geq 30$ RU) with K_D values from 0.66 nM to 305 nM (median 31.7 nM). These results demonstrate that an agent-guided computational workflow can design nanomolar to sub-nanomolar nanobody binders against a novel target without experimental structure or prior antibody information.

1. Introduction

Monoclonal antibodies have become the dominant modality in biopharmaceutical development, with over 100 FDA-approved antibody drugs and well-established manufacturing, regulatory, and clinical development pathways (Carter & Rajpal, 2022). Specialized formats such as single-domain VHHs (nanobodies) further expand therapeutic possibilities through their favorable developability characteristics, including high solubility, robust expression yields, low aggregation propensity, excellent thermal and chemical stability, and ease of engineering (Jovčevska & Muyldermans, 2020).

Yet conventional antibody discovery remains fundamentally constrained. Animal immunization, phage display, and synthetic library screening have proven success but impose inherent restrictions on accessible sequence space and demand months of iterative optimization for affinity, stability, and immunogenicity to ensure fitness for purpose (Georgiou et al., 2014; Chiu & Gilliland, 2016). Most critically, these methods provide minimal prospective control over epitope selection, sequence composition, and paratope architecture, which can ultimately determine therapeutic success. Cryptic epitopes often prove intractable, and initial hit sets offer no guarantee of viable clinical candidates.

Machine learning has transformed antibody design capabilities. Structure prediction methods including AlphaFold (Jumper et al., 2021; Abramson et al., 2024), Boltz (Passaro et al., 2025), Chai (Chai Discovery, 2024), and IntelliFold (Qiao et al., 2025) now enable accurate modeling of protein structures, while specialized tools like NanobodyBuilder2 (Abanades et al., 2023) predict VHH conformations with high confidence. For de novo design, diffusion-based frameworks including RFdiffusion (Watson et al., 2023) and BindCraft (Pacesa et al., 2024) have designed novel protein binders with explicit epitope control. Complementary approaches such as IgGM (Wang et al., 2024), which jointly optimizes CDR sequences and structures through diffusion, and mBER (Swanson et al., 2025), which leverages backpropagation through structure prediction models, expand the design toolkit. Sequence-based predictors utilizing protein language models like ESM2 (Lin et al., 2023) provide orthogonal binding validation independent of structural assumptions. However, integrating diverse ML models into cohesive discovery pipelines remains a bottleneck, and reported experimental hit rates range from 0.1% to 39%

[*]Equal contribution [1]Amazon Web Services, Applied AI Solutions, Life Sciences [2]Memorial Sloan Kettering Cancer Center. Correspondence to: Nai-Kong V Cheung <cheungn@mskcc.org>, Jiwon Kim <kimaaj@amazon.com>, Xinyun (Nina) Cheng <cxinyun@amazon.com>.

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

(Laustsen et al., 2021; Bennett et al., 2025; Mille-Fragoso et al., 2025; Nabla Bio, 2025), though direct comparison is difficult due to varying hit rate definitions, target difficulty, and screening stringency.

We test these approaches against a target that amplifies these difficulties: a novel Desmoplastic Small Round Cell Tumor target identified through transcriptomic analysis (see Appendix: Target Identification). Without a wet-lab-resolved protein structure, epitope identification and interface modeling must rely entirely on computationally predicted structures, which can introduce uncertainties in surface-exposed loop conformations, side-chain orientations, and local electrostatic features critical for defining druggable binding sites. These uncertainties are compounded by the target's multi-domain architecture, flexible interdomain regions, and potential glycosylation sites that are poorly captured by static predictions.

No public antibody information exists for this target, eliminating conventional starting points such as paratope-epitope template transfer, homology-based CDR grafting, or affinity maturation trajectories from prior campaigns. The absence of any antigen-antibody interaction data means that neither the de novo design models (RFantibody, IgGM, mBER) nor the scoring models have encountered this antigen during training. The challenge lies in whether generalized priors from other PDB complexes can produce functional binders against an unseen surface. Most de novo campaigns have targeted well-characterized antigens; validation against truly novel targets has only recently emerged (InventCures, 2026).

Here, we report results from the first cycle of designing nanobodies against this target through an integrated workflow: (1) target structure prediction and agent-guided epitope identification; (2) de novo design of 288,000 nanobody candidates; (3) multi-metric scoring and filtering; (4) *in vitro* validation of 100,000 candidates via yeast display, FACS, and SPR. We demonstrate that this workflow produces nanomolar to sub-nanomolar binders against an uncharacterized target, providing a foundation for iterative DBTL cycles.

2. Results

2.1. De novo designed nanobodies bind a novel cancer target with nanomolar to sub-nanomolar affinity

Following the computational design and scoring pipeline (Figure 1), we identified epitope residues using our internally developed hotspot recommendation agent and de novo designed 288,000 nanobody candidates, which were filtered down to 100,000 using our candidate selection agent performing multi-property optimization. The 100,000 candidates were screened by yeast surface display (with a 90.6% VHH expression rate in yeast library) with 2 rounds of FACS enrichment where antigen-specific binders were identified by flow cytometry using individually stained yeast clones. 116 candidates were advanced to SPR characterization based on mean fluorescence intensity (MFI) at the 2nd round of FACS (Figure S1). 116 candidates expressed successfully, with a median yield of 184 mg/L (range: 34.5–200 mg/L) at the 2 mL culture scale, with 97% of the library exceeding 100 mg/L.

Specificity testing confirmed no binding to an irrelevant antigen, transferrin receptor protein (TfR1), across all 116 candidates. Of the 116 candidates characterized, 46 (39.7%) produced binding responses to the target antigen with both measurable K_D and $R_{\max} \geq 30$ RU, providing reliable kinetic measurements. These 46 binders spanned K_D values from 0.66 nM to 305 nM, with a median K_D of 31.7 nM. The top candidate, PRJ266_044, achieved K_D = 0.66 nM. The second-ranked candidate, PRJ266_080, measured K_D = 2.3 nM (Figure 2). An additional 29 candidates shared detectable but low-amplitude SPR responses ($R_{\max}$ ¡ 30 RU). Low $R_{\max}$ values can arise from multiple factors including low active concentration of the binder, partial binder immunoreactivity, and the reduced reliability of kinetic parameter estimation with low signals. The fitted K_D values for these candidates should therefore be considered provisional estimates requiring orthogonal confirmation. Among them, 14 candidates yielded K_D values in the sub-nanomolar range, including PRJ266_104 (K_D = 0.13 nM, $R_{\max}$ = 12.1 RU). SPR sensorgrams for the three candidates PRJ266_044, PRJ266_080, and PRJ266_104 are shown in Figure S2.

We examined the influence of design parameters on binding affinity among the 46 high-signal binders (Figure 3). Of the three nanobody frameworks, only two yielded confirmed binders in the high-signal group, and the distribution was heavily skewed: Framework B accounted for 45 of 46 binders (Figure 3a). This dominance indicates that framework properties can determine whether designed sequences translate into strong binders and highlights the value of diversifying frameworks to avoid missing productive scaffolds. Both IgGM (n = 33, median 28.0 nM) and mBER (n = 13, median 43.9 nM) produced binders, though the difference was not significant (p = 0.311) (Figure 3b). All three RFantibody binders showed sub-nanomolar to low-nanomolar K_D (0.13, 0.62, and 5.5 nM) but with uniformly low $R_{\max}$ (11.1–16.4 RU), below the 30 RU threshold (Figure 3b). No significant differences were detected across the five structure prediction methods (Kruskal-Wallis p = 0.330, Figure 3c) or eight hotspot groups (p = 0.725, Figure 3d), though small per-group sample sizes limit statistical power.

Binders originating from all eight hotspot-conditioned design groups were confirmed by SPR, with median K_D values ranging from 10.3 nM (Hotspot G) to 43.9 nM

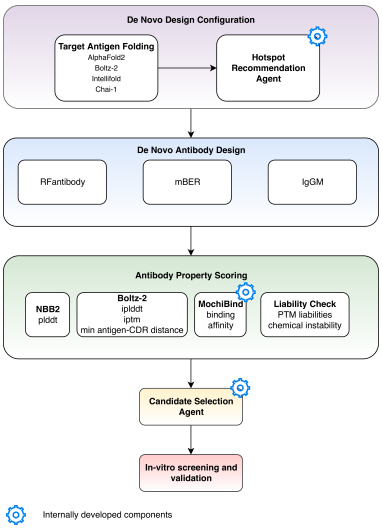

Figure 1. Integrated computational and experimental pipeline for de novo nanobody design. The workflow proceeds through five stages: (1) De Novo Design Configuration's (purple) hotspot recommendation agent predicts target antigen structures using four folding methods and recommends 8 binding epitopes. (2) De novo design (blue) generates 288,000 nanobody candidates using three complementary methods (RFantibody, mBER, IgGM) across multiple epitopes and frameworks. (3) Multi-property scoring (green) evaluates all 288,000 candidates using structural metrics (NanobodyBuilder2 pLDDT, Boltz-2 complex quality and CDR-antigen distance), sequence-based binding affinity prediction (MochiBind (Anonymous, 2026), see Methods section for a more detailed description of this model), and antibody sequence liability checks. (4) Candidate Selection Agent (yellow) performs multi-objective optimization to select 100,000 top candidates. (5) *In vitro* screening and validation (pink) experimentally validates binding activity, with 116 candidates advanced to SPR characterization, yielding 46 confirmed binders ($R_{\max} \geq 30$ RU) with measurable kinetics. Gear icons denote internally developed components that automate key decision-making steps in the workflow.

(Hotspot B). Boltz-2 co-folding independently confirms that designs concentrate contacts near agent-recommended hotspots (Figure 4b), though the mapping between conditioning hotspot and predicted binding site is not strictly one-to-one—sequences may dock at neighboring hotspots sharing a contiguous surface. SPR confirms target binding but does not resolve epitope; experimental mapping would be needed.

2.2. Target Structure Prediction

Pairwise TM-scores (Zhang & Skolnick, 2004) among the five predicted structures confirmed that all methods recovered the same overall fold (all TM ¿ 0.5; Table S1). Boltz-2, Chai-1, and IntelliFold formed a tight cluster (TM 0.78–0.91), while AlphaFold2 was more divergent (TM 0.62–0.66), likely reflecting differences in MSA construction and model version, as this structure was retrieved from the AlphaFold Database (Varadi et al., 2024) rather than generated with controlled parameters.

2.3. De Novo Design – Hotspot recommendation

Our hotspot recommendation agent orchestrates seven bioinformatics tools and synthesizes their outputs through an LLM (Claude Sonnet 4) to identify candidate epitopes (Figure 4a; see Methods). By grounding the LLM's recommendations in deterministic, verifiable tool outputs—computed surface accessibility, IEDB alignment hits, PFAM domain boundaries—rather than relying on parametric knowledge, this architecture constrains recommendations to evidence-supported regions. The agent prioritizes broad epitope coverage: in early-stage design against an uncharacterized target, missing a viable epitope outweighs the cost of testing additional candidates.

The agent was validated on SAbDab (Dunbar et al., 2014) antibody-antigen complexes. Using a holdout set (n=76) of immunologically relevant targets clustered at 70% identity, the agent proposes k=5 non-overlapping 10-amino-acid regions per antigen and achieves ∼80% top-5 accuracy (hit if any proposed region overlaps a true epitope residue).

2.4. De Novo Design – VHH binder generation

Using the three generative models described in Methods, we varied design parameters combinatorially: eight hotspots, three nanobody frameworks, CDR III lengths 4–13 residues, and five predicted antigen conformations from distinct structure prediction methods (Jumper et al., 2021; Wohlwend et al., 2024; Chai Discovery, 2024; Passaro et al., 2025; Qiao et al., 2026). Designing against multiple conformations reduces overfitting to any single model's artifacts. This strategy generated 288,000 nanobody candidates (Figure 1).

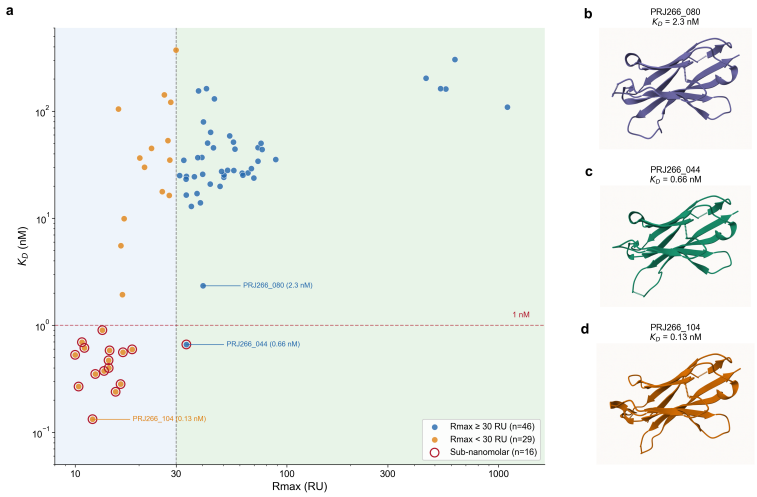

Figure 2. Binding affinity landscape and predicted structures of top de novo designed VHH antibodies. (a) Equilibrium dissociation constant (K_D) plotted against SPR maximum response ($R_{\max}$) for all 75 candidates with measurable target binding. Blue points denote candidates with $R_{\max} \geq 30$ RU (n = 46), a threshold (dashed vertical line) above which kinetic fits benefit from favorable signal-to-noise; orange points denote candidates below this threshold (n = 29). Red open circles highlight 16 candidates with sub-nanomolar affinity (K_D ¡ 1 nM; dashed horizontal line). Values for candidates with $R_{\max}$ ¡ 30 RU carry substantially higher uncertainty due to limited signal amplitude and should be interpreted as provisional estimates. Three top-performing candidates are annotated: PRJ266_044 (K_D = 0.66 nM), the highest-affinity confirmed binder with $R_{\max} \geq 30$ RU; PRJ266_080 (K_D = 2.3 nM), the second-ranked binder in that group; and PRJ266_104 (K_D = 0.13 nM), the lowest measured K_D overall, though this estimate carries substantial uncertainty due to low $R_{\max}$. (b–d) Predicted structures of the three annotated candidates, generated using NanobodyBuilder2.

2.5. Hotspot adherence using Boltz-2 to predict complex structure

To assess whether designed sequences encode hotspot preference recoverable by an independent method, we performed Boltz-2 co-folding, which predicts complex structures from antibody and antigen sequences alone without access to original hotspot constraints. We note that Boltz-2 and the design models share PDB training data, so this tests whether hotspot specificity is encoded in the designed sequences and recoverable by co-folding, rather than providing fully independent validation. Designs targeting agent-recommended hotspots concentrate contacts near the intended residues (Figure 4b), with Hotspots F, G, and H achieving high peak CDR contact frequencies. Hotspot B appears to form a discontinuous epitope shared with F and G, accounting for overlapping contact signals.

Contact frequency profiles for SPR-confirmed binders (Figure 4c) broadly mirror pre-filtering distributions. Modest shifts are observed upon filtering: designs targeting Hotspots B, F, and G appear enriched among the lowest-K_D binders, though this observation is based on small per-group sample sizes and should be treated as hypothesis-generating. These results suggest that designs conditioned on Hotspots B, F, and G may engage a shared or overlapping binding surface. However, the SPR data confirms binding but does not determine the epitope. Experimental epitope mapping is needed to validate whether the computationally predicted binding sites correspond to the true epitopes.

2.6. *In Silico* Scoring, Filtering, and Candidate Selection

Using the five-metric scoring framework and Pareto-based selection described in Methods, the candidate selection agent reduced 288,000 designs to 100,000 for experimental validation (metrics distribution in Figure S3). The three design methods produced non-overlapping sequence sets. Among the four design dimensions analyzed (epitope, framework, antigen structure, method), the choice of de novo method produced the most pronounced *in silico* score differences: RFantibody showed lower median ipTM and CDR-antigen proximity compared to mBER and IgGM, while performing comparably on monomer quality and predicted binding affinity (Figure S3). Sequence diversity analysis confirmed broadly distributed amino acid usage across CDR positions, with longer CDR H3 loops exhibiting greater positional diversity (Figures S4 to S6).

3. Discussion

Traditional antibody discovery requires 6–12 months (Kelley, 2020) from target identification to lead selection, with animal immunization consuming 8–12 weeks (Zhang, 2023) and phage display requiring 3–5 rounds of bio-panning over 4–8 weeks (Huang et al., 2016) followed by affinity maturation (Lee et al., 2004; Hoogenboom, 2005). Critically, these methods provide minimal prospective control over epi-

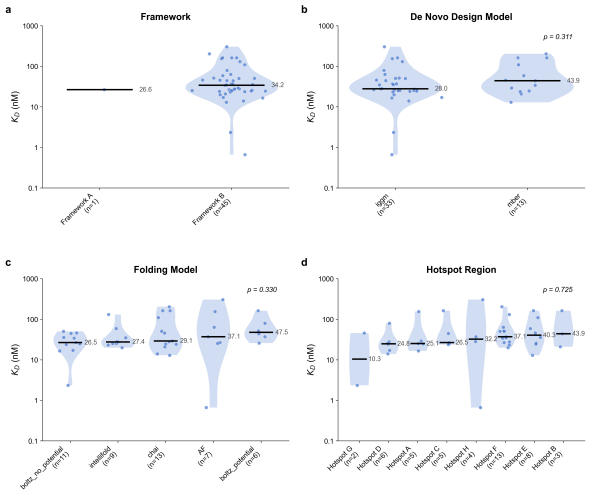

Figure 3. Binding affinity distribution by design parameters. Violin and strip plots of K_D for 46 confirmed binders ($R_{\max} \geq 30$ RU) grouped by (a) structural framework, (b) de novo generative model, (c) structure prediction model, and (d) target hotspot region. Black lines indicate group medians (nM). p-values from Kruskal–Wallis test.

tope selection. In contrast, our *in silico* workflow generated 288,000 candidates with explicit epitope targeting across eight binding regions using three complementary generative methods (RFantibody (Bennett et al., 2025), IgGM (Wang et al., 2024), mBER (Swanson et al., 2025)) conditioned on agent-predicted hotspots, multiple target conformations, and varied frameworks. Computational pre-filtering via structure prediction (Boltz-2 (Passaro et al., 2025)) and sequence-based binding affinity estimation enabled resource-efficient experimental validation of 100,000 candidates, shifting the discovery bottleneck from experimental screening throughput to the quality of generative models and scoring functions.

The Hotspot recommendation agent considers both biophysical and biochemical accessibility to binding as well as antibody function. The agent synthesized inputs from IEDB, PFAM, surface accessibility, secondary structure, and hydrophobicity profiles to propose eight distinct hotspot regions, each supported by complementary lines of evidence. For example, Hotspot B was selected for its high surface accessibility (up to 90%) and cysteine content, while Hotspot G was prioritized based on its location within a functionally critical domain with favorable charged-residue composition. Crucially, these recommendations are not opaque rankings; the agent provides biophysical and functional rationale for each hotspot, grounding its selections in interpretable biophysical properties such as surface exposure, electrostatic favorability, and domain relevance. By narrowing the design space to structurally and functionally motivated epitopes, the agent provided a rational basis for de novo design against a target with no prior antibody data. The subsequent hotspot adherence analysis (Figure 4b-c) confirms that agent-recommended hotspots are recoverable by independent structure prediction, and SPR provided confirmation of binding.

Using data from both the adherence study and SPR data, the eight hotspots do not appear to function as fully independent epitopes. Hotspot B, despite being specified as a distinct region, shares overlapping contact signals with Hotspots F and G in the Boltz-2 co-folding analysis, suggesting these three regions have formed a spatially contiguous binding surface on the folded antigen. Among the confirmed binders, designs contacting a potential discontinuous epitope (B+G) achieved some of the lowest K_D values in the panel (e.g. candidate PRJ266_080 conditioned on hotspot G, 2.3 nM). This combinatorial pattern suggests that the agent's individual hotspot recommendations, when viewed collectively, can reveal higher-order epitope architectures that no single hotspot definition would capture.

Filtering de novo candidates based on *in silico* metrics remains a challenge: median versus top-percentile performance can yield different conclusions about method efficacy. RFantibody-designed nanobodies were underrepresented in the filtered group on average, yet top designs showed a balanced distribution across all three methods. We attribute this to: (1) RFantibody's structure-first paradigm producing distinct score distributions from sequence-conditioned approaches (IgGM, mBER); (2) a known index offset issue in RFantibody's hotspot conditioning; and (3) score variance from using both ProteinMPNN (Dauparas et al., 2022) and AbMPNN (Shanehsazzadeh et al., 2023) for sequence design within the RFantibody pipeline.

While *in silico* metrics enabled efficient filtering for library generation, we observed little correlation between most *in silico* folding metrics (pLDDT, ipTM) and experimental

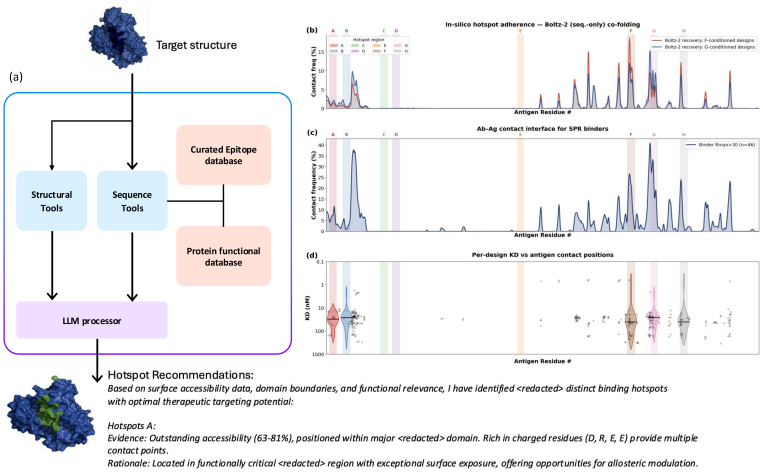

Figure 4. Hotspot Recommendation agent. (a) The agent analyzes the structures of the target (PDB) and recommends hotspot residues that are used to condition binder design to the target. (b) *In silico* hotspot adherence. The recommended hotspots are supplied as input to de novo structure-based design models (RFantibody, IgGM, mBER) which generate VHH sequences conditioned on each hotspot. The designed sequences are then co-folded with the antigen using Boltz-2, a sequence-only structure prediction model without previous hotspot and structural knowledge. The overlay shows that Boltz-2 recovers the hotspot-conditioned interface: designs targeting the agent's proposed hotspots engage near the intended targets. Nonetheless, we observe epitopes that are independent of hotspot conditioning as some epitopes are highly surface accessible and energetically favorable. (c) Antibody–antigen contact interface after Boltz-2 complex folding showing CDR contact frequency (% of binder designs contacting the i-th antigen residue) for 46 SPR-confirmed binders (K_D measurable, $R_{\max} \geq 30$ RU). Contacts are defined as any CDR heavy atom within 7 Å of any antigen heavy atom in Boltz-2 predicted structures. Shaded vertical bands mark the 8 agent-defined hotspot regions. (d) Per-design K_D vs antigen contact positions. Each design has a single K_D value (y-axis) but many designs contact multiple antigen residues simultaneously; each contacted residue is plotted as a dot at the design's K_D. Density estimates in (d) are shown for binders whose contacts fall within ± 10 residues of each hotspot range.

binding outcomes in the SPR data (Figure S7). This highlights that current structure prediction confidence scores, while valuable for assessing model quality, do not directly translate to robust binding affinity predictions, particularly for novel target antigens like the one used in this study, as observed by others in the literature, including mBER paper (Swanson et al., 2025) where an increased ipTM filtering threshold for de novo designed nanobodies doesn't translate to a higher experimental hit rate for various target antigens. This underscores the continued necessity of experimental validation at scale and points to opportunities for developing dedicated binding affinity predictors in future work.

Our approach has several limitations that we are actively working to address: (1) Limited assessment of biophysical properties, current reliance on *in silico* metrics (e.g. Boltz-2 co-folding, MochiBind affinity estimates) may not fully capture critical characteristics such as conformational stability or dynamic instabilities; we are developing molecular dynamics-based refinement to validate structural stability. (2) Limited scaffold diversity, the workflow used only three generative methods (RFantibody, IgGM, mBER), potentially constraining exploration of underrepresented scaffold classes that might offer superior developability; we are expanding scaffold diversity through additional generative models.

To integrate *in silico* and experimental workflows, we have designed a lab-in-the-loop framework (Frey et al., 2025) progressing through DBTL cycles (Figure S8). The first Design-Build-Test cycle is complete: 116 candidates have been characterized by SPR. Cell-based validation and developability characterization (thermal stability, polyreactivity) are planned to establish ground-truth fitness values. Characterized binders will then serve as training data for target-specific ML models to guide directed evolution campaigns (Romero et al., 2013; Yang et al., 2019; Wu et al., 2019; 2020; Wittmann et al., 2021) in subsequent cycles.

4. Methods

4.1. Hotspot Recommendation Agent and Tools

To systematically identify candidate binding hotspots on the target antigen, we developed an AI agent that orchestrates seven complementary bioinformatics tools and synthesizes their outputs through an LLM (Claude Sonnet 4). The agent operates through a deterministic workflow; an example of its recommendation is shown in Table 1. The seven tools are:

Unique Regions Analyzer. The user provides the target antigen and negative targets; the analyzer performs sequence alignment to ensure recommended epitopes are exclusive to

the positive target.

IEDB Epitope Database Search. Aligns the antigen sequence against 500k curated epitopes from the Immune Epitope Database using MMseqs2 with relaxed E-value thresholds tuned to recover discontinuous conformational epitopes and linear B-cell epitope fragments.

PFAM Domain Annotation. Identifies functional domain boundaries via MMseqs2-based PFAM lookup, replacing the InterPro/HMMer3 pipeline (reducing latency from ¿10 minutes to ¡1 minute).

Solvent-Accessible Surface Area. Computes per-residue SASA with chain isolation; for complex PDB inputs, existing antibody chains are removed before calculation to avoid artificially reduced accessibility scores.

Secondary Structure. Three-state classification (helix, sheet, coil) informs epitope likelihood, as loop regions and beta-turns are overrepresented in known epitopes.

Hydrophobicity Profiling. Per-residue scores characterize the physicochemical landscape, identifying hydrophobic patches that may drive binding affinity and flanking hydrophilic regions contributing to specificity.

Interface Analyzer. For complex structures, computes pairwise distances across all residue pairs to identify interface contacts (not used in this study as no prior complex existed).

By grounding the LLM's recommendations in these deterministic, verifiable tool outputs rather than parametric knowledge, this architecture constrains the recommendation space to evidence-supported regions.

Table 1. Representative hotspot recommendations from the agent. Three of eight recommended epitope regions are shown. Sensitive protein information is redacted with ⟨redacted⟩.

HS	Evidence	Rationale
A	Accessibility 63–81%, within major ⟨redacted⟩ domain. Rich in charged residues (D, R, E, E).	Functionally critical region with exceptional surface exposure.
B	Accessibility 60–90%, 81.82% uniqueness. Contains cysteine residues.	Highly accessible with unique cysteine content for conformational specificity.
C	Accessibility 75–94%, predominantly coil, 80% uniqueness. Flexible linker region.	Maximum surface exposure in structurally flexible regions ideal for binding.

4.2. De Novo Antibody Design

We employed three generative models with distinct design principles, each designing CDR I–III on three nanobody frameworks, conditioned on eight hotspots, with CDR III lengths 4–13 residues, against five predicted antigen conformations from AlphaFold2 (Jumper et al., 2021), Boltz-2 (Passaro et al., 2025), Chai-1 (Chai Discovery, 2024), and IntelliFold (Qiao et al., 2025):

RFantibody (Bennett et al., 2025) uses denoising diffusion probabilistic models to generate nanobody CDR backbones on a fixed framework, with hotspot residues specified as soft constraints biasing the diffusion trajectory toward the intended epitope.

IgGM (Wang et al., 2024) jointly optimizes CDR sequences and structure through diffusion, taking a framework with masked CDRs and antigen structure as input. We completed 960 runs (3 frameworks $\times$ 5 structures $\times$ 8 CDR III lengths $\times$ 8 hotspots), each generating 100 sequences.

mBER (Swanson et al., 2025) uses gradient-based optimization of sequence logits through AlphaFold-Multimer, guided by protein language model priors, to identify CDR sequences predicted to form stable bound complexes. The same 960-run parameterization was used.

Each method generated 96,000 candidates (288,000 total).

4.3. Scoring and Candidate Selection

Candidates were scored on five metrics: NanobodyBuilder2 pLDDT (monomer fold quality), Boltz-2 complex ipTM and ipLDDT (interface quality), minimum CDR-antigen distance, and MochiBind (Anonymous, 2026) predicted binding affinity (a sequence-based predictor using ESM-2 embeddings, orthogonal to structural metrics). Developability was assessed via LAP-adapted liability checks (Satława et al., 2024) screening for N-linked glycosylation, deamidation, fragmentation, and oxidation motifs. The candidate selection agent applied Pareto optimization (Efficient Non-dominated Sort) to stratify 288,000 candidates into successive fronts without objective weights; the first 9 ranks plus sampling from rank 10 yielded 100,000 candidates for experimental validation.

5. Conclusions

We designed nanobody binders de novo against a novel cancer target with neither an experimental structure nor prior antibody information. Starting from sequence alone, our agent-guided workflow identified eight candidate epitope regions, generated 288,000 nanobody designs across three generative methods (RFantibody, IgGM, mBER), five predicted antigen conformations, and three VHH frameworks, and selected 100,000 candidates via Pareto-based filtering

for yeast surface display screening. Of 116 candidates characterized by SPR, 46 produced reliable kinetic measurements ($R_{\max} \geq 30$ RU) with K_D values from 0.66 nM to 305 nM (median 31.7 nM). Both IgGM and mBER contributed confirmed binders, and binders originating from all eight hotspot-conditioned design groups were represented among confirmed hits. Specificity testing confirmed target-specific binding, with no candidates showing detectable interaction with the off-target control antigen.

The hotspot recommendation agent provided the epitope definitions conditioning all downstream design. Binders were recovered from designs conditioned on each of the eight hotspot regions, and Boltz-2 co-folding showed that designs concentrated predicted contacts near the intended residues. The lowest-K_D binders were enriched among designs targeting Hotspots B, F, G, and H, though experimental epitope mapping is needed to confirm true binding sites. The confirmed binders and NGS-derived labels will provide training data for target-specific ML models to guide subsequent DBTL cycles, with the goal of improving binding affinity and developability through iterative optimization.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning applied to therapeutic antibody discovery. The computational workflow described here aims to accelerate the development of nanobody binders against cancer targets, which could have positive societal impact through improved therapeutic options. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

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

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

Anonymous. MochiBind, 2026. Submission under review.

Bausch-Fluck, D., Goldmann, U., Müller, S., van Oostrum, M., Müller, M., Schubert, O. T., et al. The *in silico* human surfaceome. *Proceedings of the National Academy of Sciences*, 115:E10988–E10997, 2018.

Bennett, N. R. et al. Atomically accurate de novo design of antibodies with RFdiffusion. *bioRxiv*, 2025.

Carter, P. J. and Rajpal, A. Designing antibodies as therapeutics. *Cell*, 185:2789–2805, 2022.

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

Chiu, M. L. and Gilliland, G. L. Engineering antibody therapeutics. *Current Opinion in Structural Biology*, 38:163–173, 2016.

Dauparas, J., Anishchenko, I., Bennett, N., Baek, M., Luber, S., DiMaio, F., and Baker, D. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378:49–56, 2022.

Dunbar, J. et al. SAbDab: the structural antibody database. *Nucleic Acids Research*, 2014.

Frey, N. C. et al. Lab-in-the-loop therapeutic antibody design with deep learning. *Genentech/Roche*, 2025.

Georgiou, G., Ippolito, G. C., Beausang, J., Busse, C. E., Wardemann, H., and Quake, S. R. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology*, 32:158–168, 2014.

Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nature Biotechnology*, 23:1105–1116, 2005.

Huang, J., Ru, B., Zhu, P., Nie, F., Yang, J., and Huang, W. Advancement and applications of peptide phage display technology in biomedical science. *Journal of Biomedical Science*, 23:8, 2016.

InventCures. Comparative analysis of de novo antibody design platforms: JAM-2, Chai-2, Origin-1, and RFAntibody. Version 7, 2026. URL https://inventcures.github.io/files/v7_denovo_antibody_design_comparison.pdf.

Jovčevska, I. and Muyldermans, S. The therapeutic potential of nanobodies. *BioDrugs*, 34:11–26, 2020.

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

Kelley, B. Developing therapeutic monoclonal antibodies at pandemic pace. *Nature Biotechnology*, 38:540–545, 2020.

Laustsen, A. H., Greiff, V., Karatt-Vellatt, A., Muyldermans, S., and Jenkins, T. P. Animal immunization, *in vitro* display technologies, and machine learning for antibody discovery. *Trends in Biotechnology*, 39:1263–1273, 2021.

Lee, C. V., Sidhu, S. S., and Fuh, G. Bivalent antibody phage display mimics natural immunoglobulin avidity and permits direct isolation of high-affinity antibodies. *Nature Biotechnology*, 22:269–274, 2004.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379:1123–1130, 2023.

Mille-Fragoso, L. S. et al. Efficient generation of epitope-targeted de novo antibodies with Germinal. *bioRxiv*, 2025.

Nabla Bio. JAM-2: Fully computational design of drug-like antibodies with high success rates. Nabla Bio Research, 2025.

Pacesa, M. et al. BindCraft: one-shot design of functional protein binders. *bioRxiv*, 2024.

Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Jing, B., Portnoi, T., Nayfeh, S., and Jaakkola, T. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707.

Qiao, L., Yan, H., Liu, G., Guo, G., and Sun, S. IntelliFold-2: Surpassing AlphaFold 3 via architectural refinement and structural consistency. 2026. doi: 10.64898/2026.02.09.704787.

Qiao, L. et al. IntFold: A controllable foundation model for general and specialized biomolecular structure prediction. *arXiv*, 2025. arXiv:2507.02025.

Romero, P. A., Krause, A., and Arnold, F. H. Navigating the protein fitness landscape with Gaussian processes. *Proceedings of the National Academy of Sciences*, 110:E193–E201, 2013.

Satława, T. et al. Liability antibody profiling (LAP): a method for assessing the developability of therapeutic antibodies. *Communications Biology*, 7:102, 2024.

Shanehsazzadeh, A. et al. Unlocking de novo antibody design with generative artificial intelligence. *bioRxiv*, 2023. doi: 10.1101/2023.01.08.523187.

Swanson, E. et al. mBER: Controllable de novo antibody design with million-scale experimental screening. *bioRxiv*, 2025. doi: 10.1101/2025.09.26.678877.

The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45:580–585, 2013.

Varadi, M. et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52:D368–D375, 2024.

Wang, R. et al. IgGM: A generative model for functional antibody and nanobody design. *bioRxiv*, 2024. doi: 10.1101/2024.09.19.613838.

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

Wittmann, B. J., Johnston, K. E., Wu, Z., and Arnold, F. H. Advances in machine learning for directed evolution. *Current Opinion in Structural Biology*, 69:11–18, 2021.

Wohlwend, J. et al. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*, 2024.

Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., and Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences*, 116:8852–8858, 2019.

Wu, Z. et al. Machine learning-assisted directed evolution navigates a combinatorial epistatic fitness landscape with minimal screening burden. *bioRxiv*, 2020.

Yang, K. K., Wu, Z., and Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nature Methods*, 16:687–694, 2019.

Zhang, Y. Evolution of phage display libraries for therapeutic antibody discovery. *mAbs*, 15:2213793, 2023.

Zhang, Y. and Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.

A. Methods

A.1. Target Identification

RNA sequencing was performed on 47 desmoplastic small round cell tumor (DSRCT) specimens collected from patients undergoing surgery at a major cancer center. All patients provided consent for their tumors to be used in research. We sought to identify highly expressed cell-surface targets for antibody-based therapy that are restricted in normal tissues to minimize on-target, off-tumor effects. To this end, we performed differential gene expression analysis to identify genes upregulated in our tumor samples compared to five normal tissue samples taken from the peritoneum of healthy donors. We narrowed this list to include only genes upregulated by EWSR1::WT1, the pathognomonic transcription factor present in all DSRCT and absent in normal tissue, and then to targets predicted to be expressed on the cell surface according to the *in silico* human surfaceome (Bausch-Fluck et al., 2018). Finally, we excluded any genes with more than minimal expression in any normal tissue according to the Genotype-Tissue Expression (GTEx) database (The GTEx Consortium, 2013). The target we selected was the most differentially upregulated gene remaining after these filters were applied to our dataset.

A.2. Stacks and Compute

To systematically generate and evaluate nanobody designs at scale, we implemented a batch processing framework using a cloud-based genomics service. A master manifest was created to orchestrate the design generation process, where each manifest entry specified a unique combination of design parameters including framework selection, target structure, hotspot definition, and other relevant variables. Each entry in the manifest corresponds to an individual Omics workflow execution that generated a predetermined number of designs, with the collective output across all runs totaling 96,000 candidates per design module (RFantibody, IgGM, and mBER).

This approach allowed for parallel execution of multiple design jobs while maintaining traceability of design parameters. The manifest-driven architecture facilitated efficient GPU allocation and provided a structured framework for tracking and aggregating results from multiple concurrent design runs. Similar batch processing strategies were subsequently employed in the scoring phase, where structural prediction and evaluation tools were executed across the consolidated design pool.

The computational infrastructure consisted of 2,000 GPUs distributed across two cloud regions (1,000 GPUs each). This multi-region strategy provided robust computational capacity while maintaining flexibility to accommodate varying instance availability patterns.

A.3. Antigen structure generation

5 different antigen structures were used in the de novo design. Structure A is retrieved from AlphaFold Protein Structure Database (Varadi et al., 2024). Structure B,C,D,E are predicted using open-source protein folding models with the same key parameters including recycling steps set to 10, number of diffusion samples set to 15, and multiple sequence alignment (MSA) enabled. Structure B and C are generated by Boltz-2 (Passaro et al., 2025), with structure B enabling argument use_potentials and structure C disabling argument use_potentials. Structure D is generated by Chai-1 (Chai Discovery, 2024), and structure E is generated by IntelliFold (Qiao et al., 2025).

A.4. De Novo Antibody Design

We used 3 different protein design methods including RFantibody, IgGM and mBER to design CDR I-III regions of 3 different nanobody frameworks for the target antigen. In addition to diversifying de novo design methods, target hotspots and nanobody frameworks, we designed against 5 different predicted target structures (each from a different structure prediction model), systematically varied CDR III length and ran de novo methods with multiple random seeds to increase the diversity of our nanobody designs. For each de novo method, we generated 96,000 nanobody designs by varying the parameters and designed 288,000 nanobody sequences in total.

A.4.1. RFantibody for de novo antibody design

We used RFantibody, a structure-based generative model that leverages denoising diffusion probabilistic models (DDPMs), to design nanobody CDR backbones on a fixed framework. RFantibody operates by iteratively denoising random 3D coordinates into biophysically plausible backbone structures. For antibody-antigen applications, the model conditions CDR loop design on the fixed antigen structure. The hotspot residue set is an input parameter and is specified as soft constraints

that bias the diffusion trajectory toward generating conformations to the intended epitope.

A.4.2. Optimized Implementation for RFantibody

We developed an accelerated implementation of RFantibody that achieved 2 to 7x speedup over the vanilla RFantibody release with minimal performance degradation. Our optimizations target two computational bottlenecks: a) SE(3) Transformer Quantization. We identified layers within the SE(3)-equivariant transformer architecture that exhibit low sensitivity to reduced numerical precision. The weight layers and input/output activations inside the transformer were selectively quantized to BF16/FP16, substantially reducing memory bandwidth requirements and enabling increased batch throughput without measurable impact on design quality metrics (interface RMSD, pLDDT, PAE). b) IGSO(3) Distribution Caching. The diffusion process over rotational degrees of freedom requires sampling from the IGSO(3) distribution, the isotropic Gaussian distribution on SO(3). Computing these distributions at runtime introduces significant latency. We precompute and cache IGSO(3) probability densities across the relevant parameter ranges, eliminating redundant computation during inference.

To assess the quality of the optimized implementation, we performed two levels of validation. First, we measured the per-layer quantization error $\|W_q \cdot a_q - W \cdot a\|$ between the quantized weights W_q and original weights W applied to representative activations a from the previous layer, and confirmed that the error remained within the 16-bit machine epsilon bound, consistent with the theoretical precision floor of 16-bit reduced-mantissa formats. Second, we evaluated design quality end-to-end by running both the original and optimized RFantibody on generated samples across a diverse set of therapeutic targets and nanobody frameworks. The designed sequences from each implementation were independently folded using RF2 to recover backbone structures. The distributions of interaction PAE, predicted LDDT, and framework-aligned CDR RMSDs were near-equivalent between the two implementations, confirming that the 2–7x speedup does not affect design quality.

A.4.3. IgGM for de novo antibody design

We used IgGM, a structure-conditioned deep generative model, to design nanobody CDR sequences on a fixed framework given the predicted 3D structure of the target antigen. IgGM uses the diffusion process to jointly optimize CDR regions on nanobody sequences and structure of the nanobody-antigen complex. It takes nanobody framework with CDR residues masked and predicted antigen structure as inputs and outputs a nanobody sequence with filled-in CDR I-III regions. When designing with IgGM, we create a design run to generate 100 nanobody sequences, each with a random seed for the diffusion process. We complete a total of 960 runs, where each run is parametrized by 3 different nanobody frameworks, 5 predicted antigen structures, 8 different CDR III lengths (6 - 13 residues) and 8 target hotspots ($3 \times 5 \times 8 \times 8 = 960$). We generate a total of 96,000 nanobody designs with IgGM.

A.4.4. mBER for de novo antibody design

We used mBER, a structure- and sequence-conditioned binder-design framework, to design nanobody CDRs on a fixed VHH scaffold given the 3D structure of the target antigen. mBER leverages a fixed nanobody framework, masks the CDR regions, and then uses gradient-based optimization of sequence logits through a structure prediction model (AlphaFold-Multimer), guided by sequence priors from a protein language model, to iteratively optimize the variable regions for binding. Rather than updating model weights, mBER backpropagates structure prediction losses to the input sequence representation to identify CDR sequences predicted to form stable bound complexes. The method outputs fully specified nanobody sequences (within the defined framework) that are predicted to adopt stable bound structures against the target antigen epitope. Similar to IgGM, we use the same parameters to complete 960 design runs and generate a total of 96,000 nanobody designs with mBER.

A.5. Antibody Scoring Workflow

Prior to implementing the scoring phase, we consolidated all nanobody designs from multiple design methods (RFantibody, IgGM, and mBER) into a unified dataset. This aggregation step involved collecting designs generated against different predicted target structures, various hotspots, and across different nanobody frameworks. Each design entry was standardized to include the full nanobody sequence, design method parameters, and target epitope information to facilitate subsequent systematic evaluation. This consolidated dataset of 288,000 initial candidates was then prepared for comprehensive scoring using multiple structural prediction tools including NanobodyBuilder2 and Boltz-2.

For optimal computational resource utilization, we implemented a partitioned batch processing strategy. The 288,000

candidates were divided into smaller batches optimized for concurrent execution across the available GPU resources. Within each batch, sequences were processed sequentially to maintain computational stability. This batching approach balanced the need for high-throughput processing with the constraints of limited GPU availability, enabling efficient scoring of the large candidate pool.

A.5.1. NanoBodyBuilder2 Antibody Folding

For structural validation of the nanobody designs, we employed NanobodyBuilder2 to evaluate the structural integrity of individual candidates. Prior to scoring, all design candidates were consolidated into a single dataset. Each nanobody sequence was processed through NanobodyBuilder2, which generates predicted structural metrics including pLDDT scores. These scores serve as a proxy for structural plausibility and stability of the designed nanobodies in their unbound state. This structural validation step formed one component of our multi-metric scoring approach used to evaluate and filter candidate designs.

A.5.2. Boltz-2 Antigen-Antibody Co-Folding

Complex structure prediction and validation were performed using Boltz-2. For each nanobody-antigen pair, Boltz-2 was used to predict the structure of the complex, generating key metrics including complex iPTM and complex ipLDDT scores. These metrics served as indicators of binding plausibility and complex stability.

Additionally, from the predicted complex structures, we calculated minimum CDR distances by identifying the complementarity-determining regions (CDRs) using Abnumber and measuring the minimum atomic distances between CDR residues and the antigen.

A.5.3. Parameters Used (Default Settings)

The following parameters were configured for the Boltz-2 prediction:

`recycling_steps`: 3 The number of recycling steps to use for prediction.

`sampling_steps`: 200 The number of sampling steps to use for prediction.

`diffusion_samples`: 1 The number of diffusion samples to use for prediction.

`max_parallel_samples`: 5 Maximum number of samples to predict in parallel.

`step_scale`: 1.638 The step size is related to the temperature at which the diffusion process samples the distribution. The lower the value, the higher the diversity among samples (recommended between 1 and 2).

A.5.4. MochiBind Antigen-Antibody Binding Affinity Prediction

MochiBind (Anonymous, 2026) is our internally developed sequence-based binding affinity predictor to estimate binding affinities of our designs to the target antigen to get a final scoring metric that's orthogonal to structural metrics discussed earlier. MochiBind uses ESM2 to embed the nanobody-antigen complex and applies a predictor head on the complex embedding to predict the binding affinity of the nanobody sequence. Our validation results indicate that this method outperforms structure-based proxies for binding affinity on out-of-distribution datasets of unseen antigen-antibody pairs. Using MochiBind at the scoring stage, we predicted the binding affinity of each nanobody design against our target antigen and used this metric for downstream filtering. MochiBind manuscript is currently under review.

A.5.5. Antibody Diversity and Liability Analysis

Sequence diversity module employed IMGT-numbered positional analysis to determine amino acid frequencies across Complementarity Determining Regions (CDRs) and Framework Regions (FRs). Sequence-based clustering was performed using MMseqs2 for enhanced performance and accuracy, complemented by traditional metrics like substitution analysis and pairwise distance calculations to quantify sequence variability. The liability analysis module systematically checked for a range of potential developability risks, adopting checks from the LAP methodology (Satława et al., 2024). This check was designed to look for critical sequence motifs for high-risk liabilities such as N-linked glycosylation, high-risk deamidation, and fragmentation within CDRs and FRs, along with other concerns like isomerization, cysteine bond issues, and oxidation

risks, providing an essential assessment of potential manufacturing and stability challenges. Antibody candidates which have red flags from the liability check were filtered out.

A.6. Candidate Selection

Computationally designed nanobody candidates must be evaluated across multiple competing structural and functional quality metrics before advancing to experimental validation. Selecting the top 100,000 candidates from a large, filtered library requires balancing these objectives without imposing arbitrary relative weights, preserving diverse high-quality candidates rather than collapsing to a single composite score. The 288,000 de novo designed VHH candidate sequences as well as metrics from the scoring modules were selected through multi-property optimization through Pareto ranks to select the 100,000 best candidates that perform well in five objectives shown in Table 2.

Each candidate was scored on five normalized properties (higher is better):

Table 2. Properties used for antibody candidate selection.

Property	Description
CDR Min Distance	Minimum distance between CDR loops and the target, inverted so that tighter engagement scores higher
Predicted Binding Affinity	MochiBind predicted binding affinity (inverted), reflecting stronger target interaction
Protein iPTM	Interface predicted TM-score from structure prediction, measuring confidence in the binding interface
Complex ipLDDT	Interface per-residue local distance difference test, assessing local structural accuracy at the interface
Nanobody Avg pLDDT	Average pLDDT across the nanobody, reflecting overall fold confidence

We applied a modified version of the Efficient Non-dominated Sort (ENS) algorithm to identify non-dominated antibodies to stratify all candidates into successive Pareto fronts without requiring objective weights:

- Pareto Front Identification. A candidate belongs to Rank 0 (the true Pareto front) if no other candidate is simultaneously equal or better on all five objectives and strictly better on at least one. This front represents the set of candidates where improving any single metric necessarily comes at the cost of another.
- Iterative Peeling. After removing Rank 0, the same non-dominated sorting is repeated on the remaining candidates to identify Rank 1, then Rank 2, and so on. This produces an ordered layering of the full library by multi-objective quality.
- Pareto Rank Thresholding. We computed the minimum number of Pareto ranks that resulted in our target library size of 100,000. Specifically, we selected the first 9 Pareto ranks completely and randomly sampled from the Pareto Rank 10 set until we had the required number of antibodies.

A.7. Yeast Library Construction and Quality Control

Linear DNA libraries were synthesized by Twist Bioscience and transformed into the *Saccharomyces cerevisiae* strain EBY100. Library quality control was performed using both Sanger sequencing and Next-Generation Sequencing (NGS) with the MiSeq 300×300 kit (Illumina), adhering to the manufacturer's recommended protocol.

A.8. Yeast Display Induction and Staining

Transformed yeast cells were cultured in SDCAA medium (Teknova) and subsequently induced at an OD_{600} of 1.0 for 48 hours at 18°C in SGCAA medium (Teknova). Following induction, cells were washed with PBS supplemented with 0.5% BSA (PBSA). The cells were incubated with the target antigen for 1 hour at 4°C with gentle shaking. Post-incubation, cells were centrifuged at 300×g for 5 minutes and washed with PBSA.

For fluorescent labeling, cells were resuspended in 1 mL of buffer containing secondary antibodies: 1:200 of 0.5 mg/mL anti-Flag-647 (GenScript) for antigen detection and 1:200 of 0.5 mg/mL anti-V5-488 (GenScript) for VHH display quantification.

The mixture was incubated for 1 hour at 4°C. Cells were washed three times with PBSA and finally resuspended in 1 mL of PBSA for analysis.

A.9. Flow Cytometry and Sorting

Sorting was performed using an SH800S Cell Sorter (Sony Biotechnology). The gating strategy isolated the double-positive population (indicating both VHH display and antigen binding) as well as the display-positive/antigen-negative population. A minimum of 10-fold coverage of the library diversity was bulk sorted and propagated for the subsequent round of selection.

In Round 1, cells were sorted against 400 nM target antigen. In Round 2, high non-specific background binding to the secondary detection antibodies was observed. To mitigate this, a negative sorting step was introduced prior to positive selection: the induced library was incubated with all secondary detection reagents without antigen, and the non-binding population was collected to deplete clones exhibiting non-specific adhesion. Positive sorting was then performed at 200 nM target antigen, with both higher and lower binding subpopulations within Q2 collected separately.

A.10. SPR

The 116 selected VHH candidates were recombinantly expressed in HEK293 cells and purified via protein A chromatography. All 116 candidates expressed successfully, with a median yield of 368 μg per 2 $\times$ 1 mL culture (range: 69–400 μg; 97% above 200 μg). Binding kinetics were characterized by surface plasmon resonance (SPR) on a Carterra LSA instrument. The VHH-fc was immobilized on the sensor surface, and the target antigen ECD were injected as analytes at multiple concentrations to determine association rate (k_{on}), dissociation rate (k_{off}), and equilibrium dissociation constant (K_D). Specificity was confirmed by testing all candidates against an irrelevant control antigen transferrin receptor protein (TfR1); no binding was detected for any candidate against the control, confirming target-specific interaction. Candidates with measurable binding responses and $R_{\max} \geq 30$ RU were considered to have reliable kinetic fits for affinity determination.

B. Supplementary Information

Table S1. Pairwise structural similarity (TM-score) of predicted models for the target protein.

	AlphaFold2	Boltz-2	Boltz-2 + potential	Chai-1	IntelliFold
AlphaFold2	—	0.627	0.632	0.663	0.622
Boltz-2		—	0.909	0.850	0.775
Boltz-2 + potential			—	0.864	0.825
Chai-1				—	0.893
IntelliFold					—

TM-scores were computed on Cα atoms after optimal superposition. A TM-score > 0.5 indicates the same global fold; 1.0 denotes identical structures.

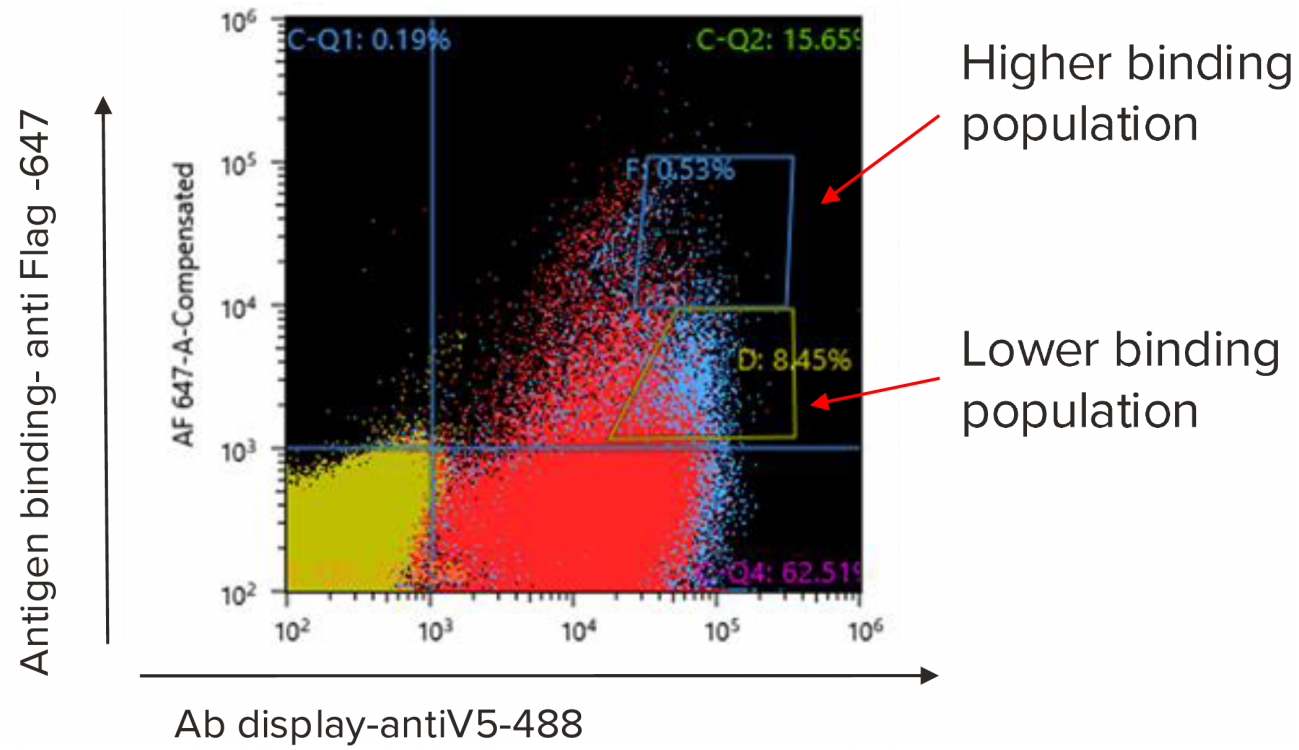

Figure S1. Round 2 FACS enrichment and selection of 116 candidates for SPR. Following the 2nd round of FACS enrichment, individual yeast clones from the binding-positive population were arrayed and rescreened by flow cytometry. Binding specificity for each clone was assessed using mean fluorescence intensity (MFI) measurements. Clones exceeding a predefined MFI threshold were classified as specific binders, yielding 116 candidates that were advanced to recombinant expression, purification, and SPR characterization.

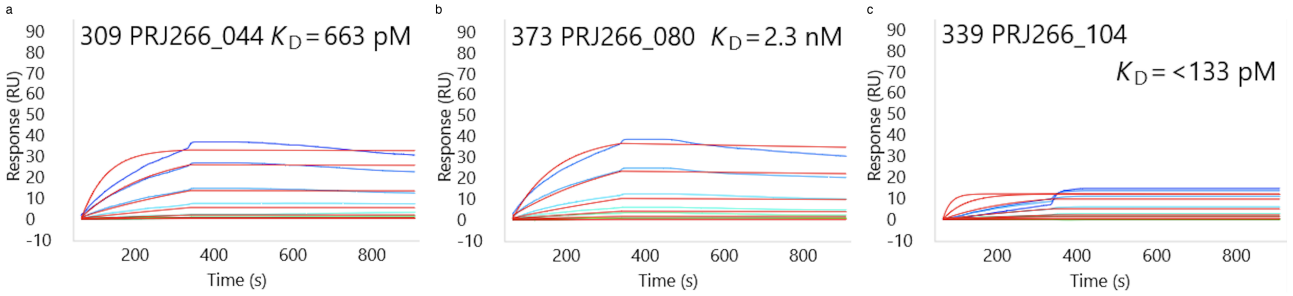

Figure S2. SPR sensorgrams of three top de novo designed VHH binders against the target antigen. Multi-concentration kinetic binding curves showing association and dissociation phases for (a) PRJ266_044, (b) PRJ266_080, and (c) PRJ266_104. Colored traces represent experimental data at different analyte concentrations; fitted curves are overlaid.

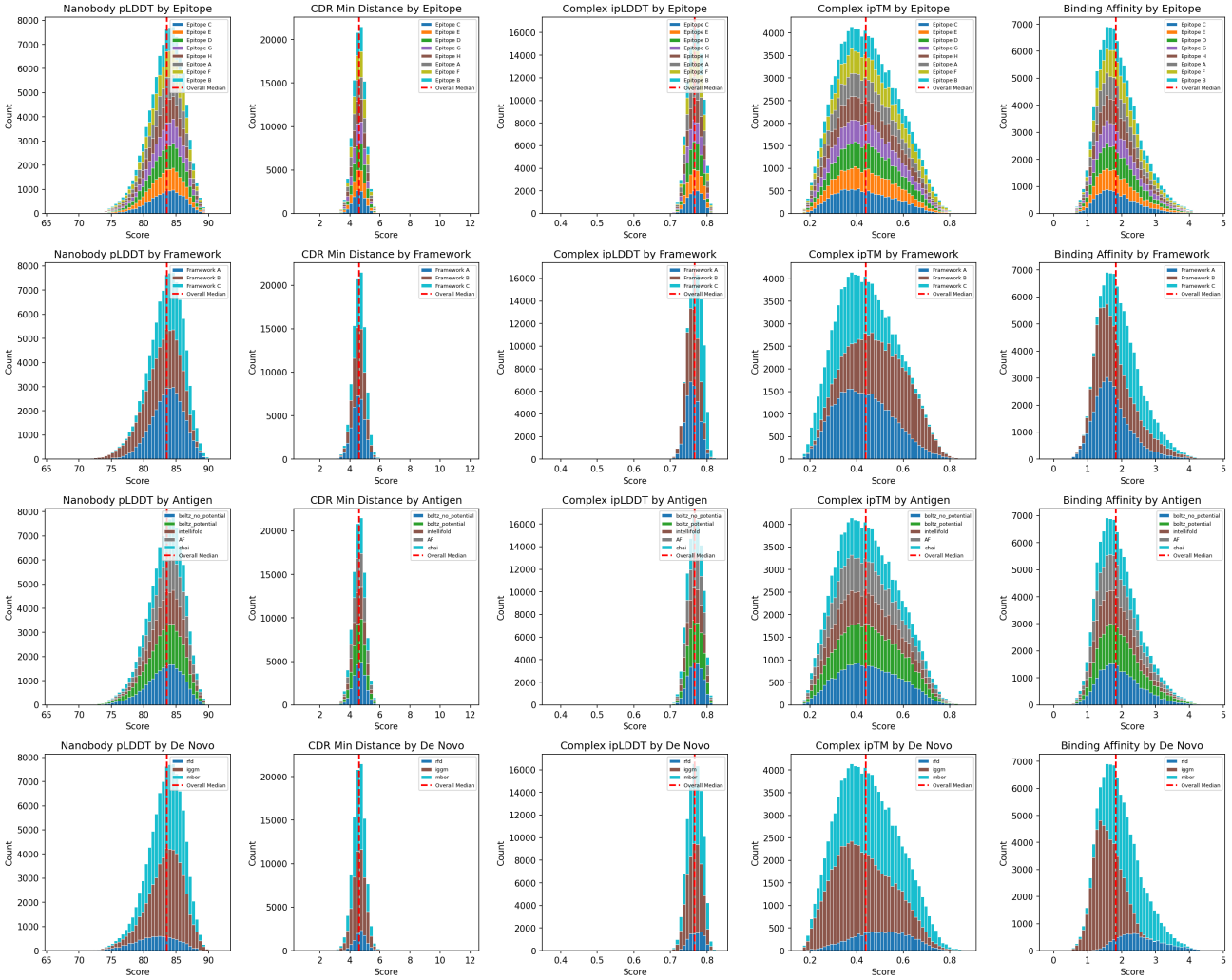

Figure S3. Distribution of *in silico* metrics for 100,000 candidates selected for *in vitro* experiments. Rows correspond to experiment parameters (from top to bottom): target epitope, nanobody framework, antigen structure prediction method, de novo design method. Columns correspond to *in silico* metrics (left to right): nanobody monomer folding pLDDT score, minimum CDR-antigen distance, complex ipLDDT, ipTM and predicted binding affinity. Higher scores indicate better designs, except in minimum CDR distance and predicted binding.

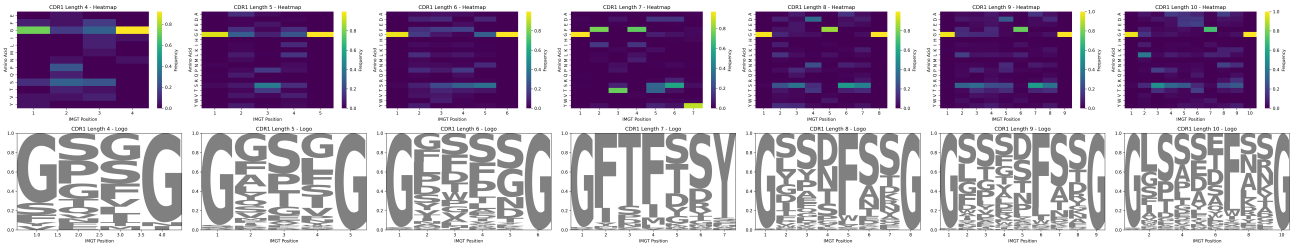

Figure S4. Library sequence diversity analysis for CDR H1 region positional amino acid frequencies for different CDR lengths. Heatmaps and sequence logos visualize the same underlying frequencies for a given CDR length.

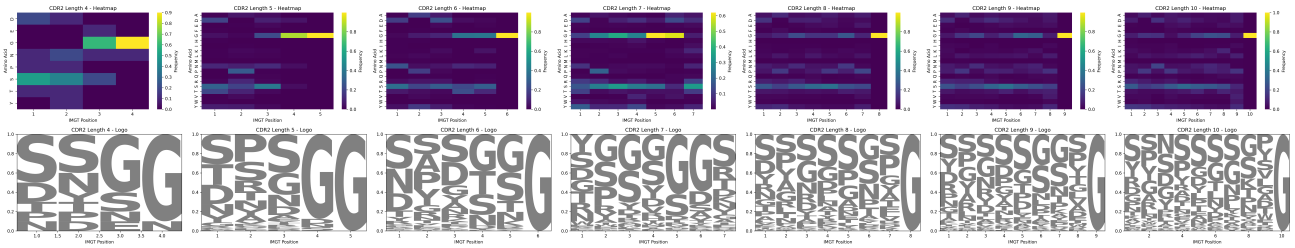

Figure S5. Library sequence diversity analysis for CDR H2 region positional amino acid frequencies for different CDR lengths. Heatmaps and sequence logos visualize the same underlying frequencies for a given CDR length.

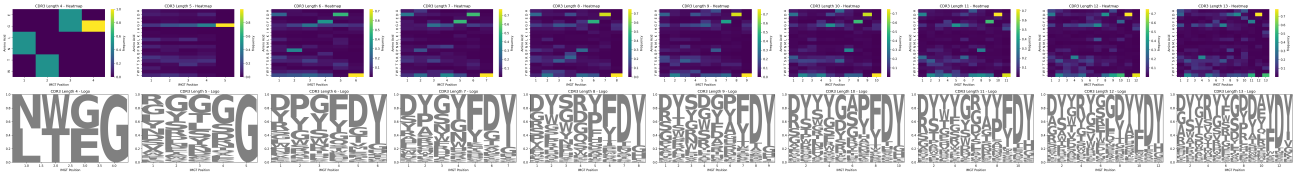

Figure S6. Library sequence diversity analysis for CDR H3 region positional amino acid frequencies for different CDR lengths. Heatmaps and sequence logos visualize the same underlying frequencies for a given CDR length.

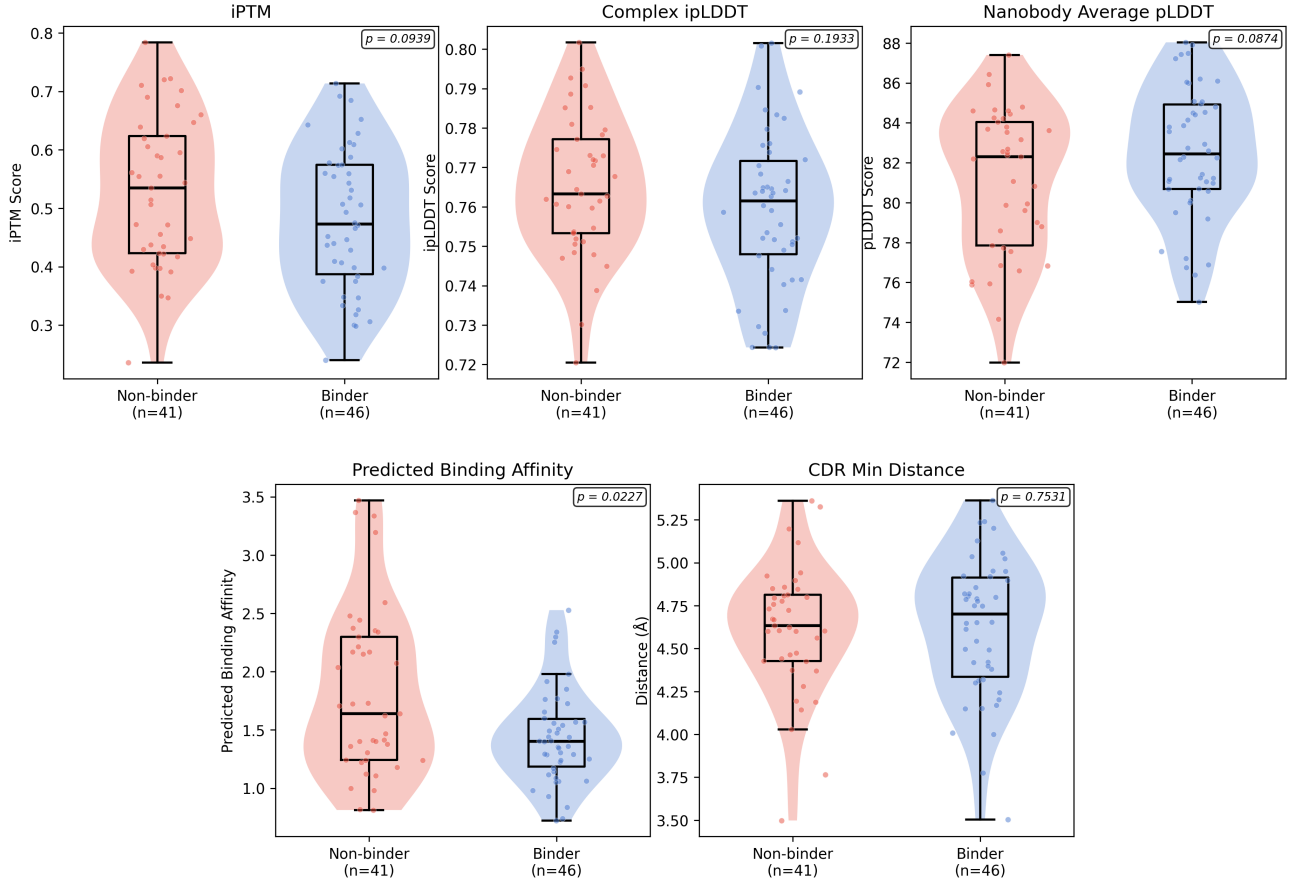

Figure S7. Comparison of *in silico* metrics for SPR-confirmed binders and non-binders. Violin plots comparing five computational metrics between nanobodies classified as binders ($R_{\max} \geq 30$ RU, n=46) and non-binders (n=41) from the 116 SPR-characterized candidates. Features include Boltz-2 interface predicted TM score (iPTM), complex interface predicted LDDT (ipLDDT), NanoBodyBuilder2 average pLDDT, MochiBind-predicted binding affinity, and minimum CDR-antigen distance. Box plots show median and interquartile range, p-values from two-sided Mann-Whitney U tests. MochiBind-predicted binding affinity shows the strongest discrimination between groups (p = 0.0227).

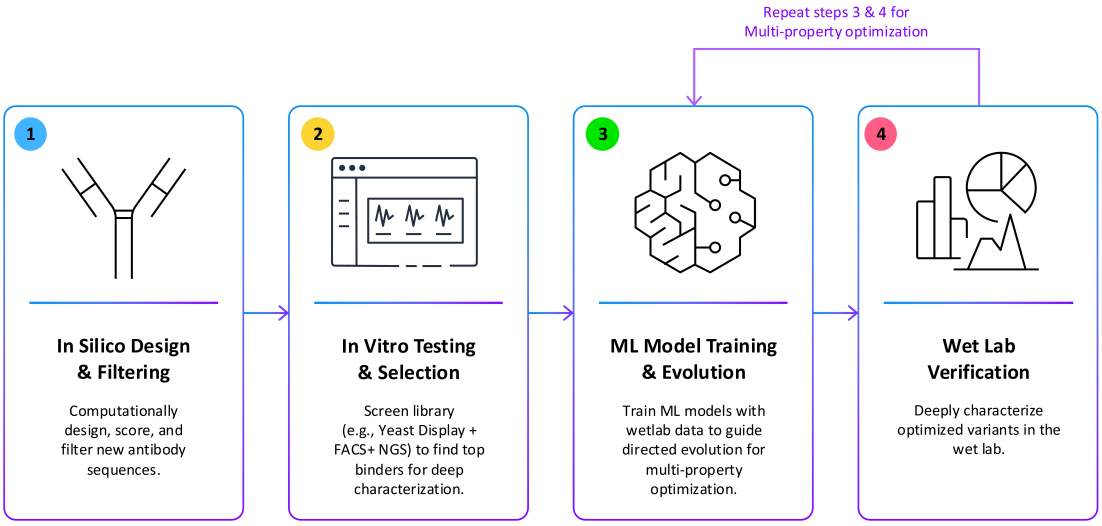

Figure S8. Design-Build-Test-Learn (DBTL) cycle enabling iterative multi-property optimization through Lab-in-the-Loop integration. The workflow comprises four interconnected stages: (1) *In Silico* Design & Filtering (blue) computationally generates, scores, and filters new antibody sequences using de novo design methods and multi-metric evaluation. (2) *In Vitro* Testing & Selection (yellow) screens libraries using high-throughput experimental methods (yeast display, FACS, NGS) to identify top binders for deep characterization. (3) ML Model Training & Evolution (green) trains target-specific machine learning models on experimental data to predict binding affinity, stability, and developability, guiding directed evolution campaigns. (4) Wet-lab Verification (red) performs comprehensive biophysical characterization of optimized variants including binding kinetics, thermal stability, and other therapeutic properties. The cycle iterates through stages 3 and 4, progressively refining candidates through ML-guided evolution and experimental validation. This closed-loop framework couples computational predictions with experimental feedback to systematically improve antibody properties while maintaining a mechanistic understanding of design-function relationships.