De Novo VHH Nanobody Design Against Ten Challenging Cancer Targets: A Systematic Computational Study with RFAntibody

Ashish Makani*
KCDH-A, Ashoka University, Sonipat, India

February 2026 — Preprint v3.0

Abstract

De novo computational antibody design has emerged as a transformative approach to therapeutic discovery, yet its systematic application to the cancers with greatest unmet need remains limited. Here we present a complete computational study designing VHH nanobodies against ten challenging cancer targets spanning five indications: MPNST, DIPG/DMG, neuroblastoma, glioblastoma, and pancreatic ductal adenocarcinoma. Through systematic analysis of 43 target–indication pairs, we identify 10 structurally actionable targets and introduce a four-tier structural readiness classification revealing that only 50% of evaluated targets possess antibody–antigen co-crystal structures. Using rfab-harness, an open-source campaign orchestration tool wrapping the three-stage RFAntibody pipeline, we executed all ten campaigns in parallel on NVIDIA A100-80GB GPUs via Modal cloud infrastructure. The pipeline generated 817 backbone scaffolds via RFdiffusion, expanded them to 4,085 sequenced designs through ProteinMPNN, and scored all designs using RoseTTAFold2 structure prediction. Applying stringent quality filters (pAE < 10, CDR RMSD < 2.0 Å), 135 designs (3.3%) passed across all campaigns, with per-target pass rates varying from 0.3% (EGFRvIII, HER2 Domain IV) to 19.8% (CEACAM5). The best candidates achieved predicted aligned errors as low as 2.19 Å and CDR backbone RMSDs of 0.68 Å, indicating high structural confidence. CEACAM5, CD47, B7-H3, and mesothelin emerged as the most computationally tractable targets, while EGFRvIII and GPC2 exhibited the greatest design difficulty. All code, configurations, results, and analysis are openly available, establishing a reproducible framework for translating computational protein design into therapeutic antibody candidates for intractable cancers.

Keywords: de novo antibody design, computational protein design, RFAntibody, RFdiffusion, cancer immunotherapy, VHH nanobody, MPNST, DIPG, neuroblastoma, glioblastoma, pancreatic cancer

1 Introduction

The past two years have witnessed a paradigm shift in antibody discovery. Where traditional approaches required immunization campaigns lasting months followed by extensive screening of 10^6–10^9 clones, de novo computational platforms can now generate antibody candidates *in silico* in hours. Bennett, Watson et al. demonstrated that RFAntibody, a three-stage pipeline combining RFdiffusion for backbone generation, ProteinMPNN for sequence design, and RoseTTAFold2 for structure prediction, can generate binders with affinities reaching the picomolar range, though per-target experimental success rates ranged from 0–2% [1]. Concurrently, proprietary platforms including Absci's generative antibody design platform and Chai Discovery's Chai-2 have reported de novo antibody design capabilities [2, 3].

Despite these advances, the application of de novo design to the cancers with the greatest unmet need has been remarkably limited. The original RFAntibody demonstration targeted influenza hemagglutinin, *Clostridioides difficile* toxin B, respiratory syncytial virus, a PHOX2B neoantigen, and the SARS-CoV-2 receptor binding domain—

none of which represent the cancers responsible for the most intractable mortality burden. Five indications stand out for their combination of devastating clinical outcomes and near-complete absence of effective antibody therapeutics: MPNST (5-year survival ∼15% when metastatic, no approved antibody therapy), DIPG/DMG (median survival ∼11 months, universally fatal in children), high-risk neuroblastoma (∼50% survival despite aggressive multimodal treatment), glioblastoma (median survival ∼15 months, no antibody therapy improving overall survival), and pancreatic ductal adenocarcinoma (5-year survival ∼12%, third leading cause of cancer death).

Each of these indications presents unique biological challenges for antibody therapy. CNS tumors (DIPG, GBM) are shielded by the blood–brain barrier, which permits less than 0.1% of systemically administered IgG (∼150 kDa) to reach the tumor parenchyma [4]. Pancreatic cancer generates a dense desmoplastic stroma comprising up to 80% of tumor mass, severely limiting antibody diffusion [5]. MPNST is driven primarily by intracellular loss-of-function events (NF1, SUZ12/EED), creating a paucity of druggable surface antigens. Neuroblastoma's most validated target, GD2, is a ganglioside rather than a protein, precluding standard computational design approaches.

These challenges motivate two central questions that this work addresses. First, which surface antigens across these five indications have sufficient structural data and clinical rationale to justify computational antibody design campaigns? Second, can the de novo design process be systematized into a reproducible, scalable framework that enables researchers to go from target selection to designed candidates with minimal manual intervention?

To answer these questions, we conducted a systematic analysis of 43 target–indication pairs, developed a four-tier structural readiness classification, built rfab-harness (an open-source campaign orchestration tool), and executed all ten design campaigns in parallel on cloud GPU infrastructure. We provide complete results including score distributions, filtering outcomes, ranked candidate lists, and cross-campaign comparisons for all ten targets.

2 Methods

2.1 Target Identification and Prioritization

Target identification employed three complementary data sources queried between January and February 2026.

Open Targets Platform. Disease–target associations were retrieved via the Open Targets GraphQL API (release 25.02) for each indication using established ontology identifiers: MPNST (EFO_0000760), DIPG (EFO_1000026), DMG (EFO_0020983), neuroblastoma (EFO_0000621), GBM (EFO_0000519), and PDAC (EFO_0000232) [8]. Targets were ranked by overall association score (0–1 scale integrating genetic, somatic, drug, literature, and animal model evidence). The top 50 targets per indication were screened for surface accessibility and antibody-targetable extracellular domains.

RCSB Protein Data Bank. For each candidate target, the PDB was queried for (1) any deposited structure of the extracellular domain and (2) antibody–antigen co-crystal or cryo-EM complex structures. Structures were evaluated by resolution, completeness of the extracellular domain, and presence of defined antibody–antigen interfaces.

Clinical trial registries and literature. ClinicalTrials.gov, PubMed (2022–2026), and conference proceedings from ASCO 2024–2025, AACR 2025, and ASH 2024 were searched for antibody-based therapies targeting each candidate.

2.2 Structural Readiness Classification

We developed a four-tier classification to assess the feasibility of structure-guided de novo antibody design for each target. Tier 1 (Excellent) requires multiple antibody–antigen complex structures at <3 Å resolution with well-defined epitopes. Tier 2 (Good) requires at least one antibody–antigen complex structure. Tier 3 (Limited) indicates that a target extracellular domain structure exists but no antibody complex, requiring epitope inference. Tier 4 (Insufficient) indicates no usable extracellular domain structure, or that the target is a non-protein antigen. Only Tier 1 and Tier 2 targets were considered suitable for immediate de novo design campaigns.

2.3 Epitope and Hotspot Derivation

For each campaign target, epitope residues were defined as target residues with any heavy atom within 4.5 Å of an antibody heavy atom in the reference complex structure. Hotspot residues (3–5 per target) were selected from the epitope set based on three criteria: (1) high buried surface area upon complex formation, (2) hydrophobic or aromatic character (Phe, Trp, Tyr, Leu, Ile, Val preferred), and (3) spatial centrality within the epitope patch. For well-characterized targets (EGFR/cetuximab, HER2/trastuzumab, CD47/magrolimab), published epitope definitions were used directly. For newer structures (9LME for B7-H3, 8UKV for EGFRvIII, 6WJL for GPC2), interface contacts were computed using BioPython's NeighborSearch algorithm.

2.4 Antibody Format Selection

VHH nanobody format (∼15 kDa, single-domain) was selected for all 10 campaigns based on three considerations: (1) improved blood–brain barrier penetration relative to full IgG (∼150 kDa) for CNS tumor targets [7], (2) enhanced stromal penetration for pancreatic cancer targets, and (3) compatibility with the RFAntibody pipeline, which has demonstrated successful VHH design with the NbBCII10 framework template.

2.5 Campaign Configuration

Each campaign was defined by a YAML configuration file specifying the target PDB, epitope and hotspot residues, antibody format (VHH), framework (NbBCII10), CDR loop length ranges, pipeline parameters, and filtering thresholds. CDR H3 loop lengths were set to ranges of 7–15 residues for most targets, with adjusted ranges for targets requiring longer paratope reach (CEACAM5: 10–18, GPC2: 10–18, EphA2: 8–15). CDR H1 and H2 lengths were fixed at 7 and 6 residues, respectively, consistent with the NbBCII10 framework germline.

2.6 Computational Pipeline

The three-stage RFAntibody pipeline was orchestrated by rfab-harness, a custom Python package that manages subprocess execution, stage-level checkpoint persistence, and Quiver-format (.qv) I/O between stages.

Stage 1: RFdiffusion. The SE3-equivariant diffusion model generated backbone designs using 50 diffusion timesteps with the `RFdiffusion_Ab.pt` weights. Input conditioning included the target epitope structure, hotspot residue positions, CDR loop length ranges, and the NbBCII10 framework in HLT format.

Stage 2: ProteinMPNN. The ProteinMPNN graph neural network designed five amino acid sequences per backbone scaffold at a sampling temperature of 0.2, masking framework residues to preserve the NbBCII10 sequence while allowing full CDR sequence exploration.

Stage 3: RF2 Structure Prediction. RoseTTAFold2 independently predicted the three-dimensional structure of each sequenced design in complex with the target, using 10 recycling iterations. Predicted structures were evaluated on multiple metrics: interaction pAE (predicted aligned error across the antibody–antigen interface), per-residue pAE, predicted LDDT (local distance difference test), target-aligned CDR RMSD, target-aligned antibody RMSD, and framework-aligned CDR and individual loop RMSDs (H1, H2, H3).

2.7 Filtering and Ranking

Designs were filtered on two primary quality metrics: predicted aligned error (pAE < 10.0 Å) and target-aligned CDR backbone RMSD (< 2.0 Å). The RF2 scoring pipeline does not produce Rosetta-based $\Delta\Delta G$ estimates; accordingly, the binding energy filter described in some RFAntibody protocols was not applied. Filtered candidates were ranked by a composite score:

$$S_{\text{composite}} = 0.4 \cdot \hat{p}_{\text{pAE}} + 0.3 \cdot \hat{p}_{\text{RMSD}} \qquad (1)$$

where $\hat{p}$ denotes min-max normalization to $[0, 1]$ across the filtered candidate set, with the remaining weight redistributed proportionally. Lower composite scores indicate higher-confidence candidates.

2.8 Compute Infrastructure

All ten campaigns were executed in parallel on Modal cloud infrastructure, each allocated a single NVIDIA A100-80GB GPU. Pipeline orchestration, checkpointing, and result persistence were managed

via Modal Volumes with periodic commits every 10 minutes for fault recovery. Stage 1 backbones from a prior partial run (targeting 500 designs per campaign) were preserved and reused via a skip-stage checkpoint mechanism, with Stages 2 and 3 run fresh against the existing backbone designs. Total wall-clock time from Stage 2+3 launch to all-campaigns-complete was approximately 75 minutes.

3 Results

3.1 Landscape of Antibody-Targetable Antigens

Systematic screening identified 43 target–indication pairs across the five cancer types (Table 1). The number of potential targets varied dramatically by indication: GBM yielded 11 evaluated targets (from 9,906 Open Targets associations), PDAC produced 10, neuroblastoma generated 8, MPNST yielded 7 (reflecting the paucity of validated surface antigens in sarcoma), and pediatric gliomas produced 7 targets.

Table 1: Landscape of evaluated targets per indication.

Indication	Targets	Tier 1–2	Tier 3	Tier 4
MPNST	7	2	2	3
DIPG/DMG	7	2	3	2
Neuroblastoma	8	3	3	2
GBM	11	7	2	2
PDAC	10	4	3	3
Total (unique)	**28**	**14**	**8**	**6**

The structural readiness distribution reveals a critical bottleneck: only 14 of 28 unique targets (50%) have any antibody–antigen complex structure in the PDB.

3.2 Cross-Indication Driver Targets

Three targets emerged as cross-indication drivers appearing in ≥4 of 5 indications (Table 2).

Table 2: Cross-indication driver targets.

Target	*MPNST*	*DIPG*	*NB*	*GBM*	*PDAC*	n/5	Tier
B7-H3	#1	#2	#2	#3	—	4	2
GD2	#4	#1	Est.	#10	—	4	4
EGFR	#2	#6	—	#1	#4	4	1
HER2	#6	—	—	#4	#6	3	1

B7-H3 (CD276) is the most striking pan-cancer target, overexpressed in MPNST (58% of sarcomas), DIPG (uniformly on tumor cells and vasculature), neuroblastoma, and GBM, while exhibiting limited normal tissue expression. B7-H3-directed CAR-T cells received FDA Breakthrough Therapy Designation for DIPG in April 2025 following Phase I results showing median survival of 19.8 months from diagnosis [6]. GD2, while appearing in 4/5 indications, is a ganglioside rather than a protein, excluding it from computational antibody design workflows.

3.3 Priority Target Selection

Integrating cross-indication impact, structural readiness, and therapeutic potential, we selected 10 targets for de novo VHH nanobody design campaigns (Table 3).

3.4 Pipeline Execution Summary

Stage 1 (RFdiffusion) generated 53–118 backbone scaffolds per campaign (817 total), with variation arising from a prior run that was stopped and clipped at the available design count. Stage 2 (ProteinMPNN) expanded each backbone into five sequenced designs, producing 265–590 sequences per campaign (4,085 total) in 3.3–3.7 minutes per campaign. Stage 3 (RF2) scored all 4,085 designs in approximately 75 minutes of wall-clock time across 10 parallel A100 GPUs, with each design requiring ∼12 seconds (10 recycling iterations at ∼1.2 seconds each). The complete Stage 2+3 execution consumed approximately 12.5 GPU-hours.

Table 3: Ten priority targets selected for de novo VHH design campaigns with Stage 1 backbone counts and final filtering results.

#	Target	Indication(s)	PDB	Res.	BB	Seqs	Pass	Rate
1	B7-H3	GBM, MPNST, DIPG, NB	9LME	2.4 Å	82	410	11	2.7%
2	CD47	GBM	5IWL	2.8 Å	96	480	19	4.0%
3	CEACAM5	PDAC	8BW0	3.1 Å	82	410	81	19.8%
4	EGFR	GBM, MPNST, PDAC	1YY9	2.6 Å	64	320	5	1.6%
5	EGFRvIII	GBM	8UKV	2.9 Å	73	365	1	0.3%
6	EphA2	GBM	3SKJ	2.5 Å	70	350	3	0.9%
7	GPC2	Neuroblastoma	6WJL	3.3 Å	53	265	2	0.8%
8	HER2 Dom. IV	GBM, MPNST	1N8Z	2.5 Å	72	360	1	0.3%
9	MSLN (N-term)	PDAC	4F3F	2.6 Å	107	535	6	1.1%
10	MSLN (C-term)	PDAC	7U8C	—	118	590	6	1.0%
Total					**817**	**4,085**	**135**	**3.3%**

3.5 Design Funnel and Global Filtering

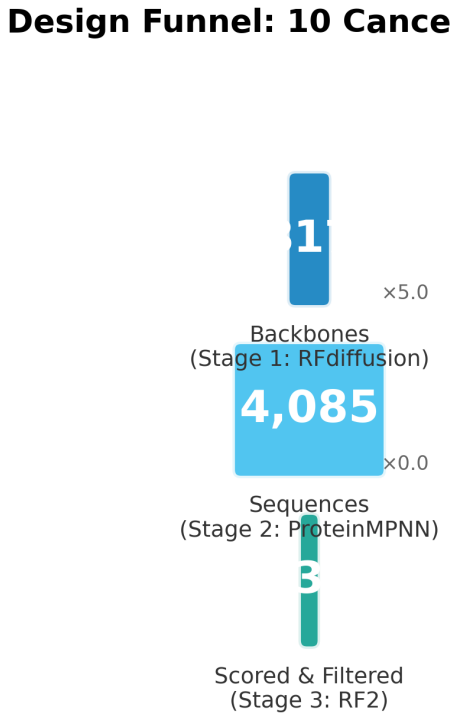

Figure 1: Design funnel showing attrition from backbone generation through sequence design to filtered candidates across all 10 campaigns. The 3.3% overall pass rate (135/4,085) reflects the stringent dual-threshold filtering applied.

Of the 4,085 designs scored by RF2, 135 (3.3%) passed both quality filters (pAE < 10 and CDR RMSD < 2.0 Å; Figure 1). For reference, the original RFAntibody study reported per-target experimental binding rates of 0–2% across benchmark targets [1]; our 3.3% computational pass rate using purely *in silico* metrics is not directly comparable to experimental binding assays, but the same order of magnitude suggests the filtering thresholds are appropriately stringent.

3.6 Per-Target Pass Rates

Filter pass rates varied more than 60-fold across campaigns, from 0.3% (EGFRvIII, HER2 Domain IV) to 19.8% (CEACAM5), revealing substantial differences in computational design difficulty among cancer targets (Figure 2).

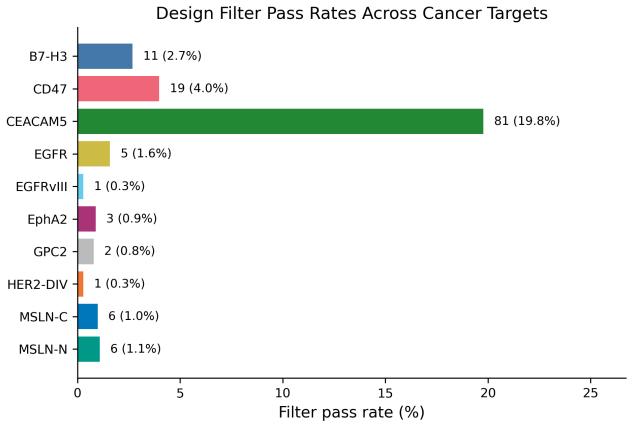

Figure 2: Filter pass rates by target. CEACAM5 dominates with 81 passing designs (19.8%), while EGFRvIII and HER2 Domain IV each produced only a single passing candidate.

The targets clustered into three difficulty tiers based on pass rate. The *high-tractability* group comprised CEACAM5 (19.8%), which alone contributed 60% of all passing designs. The *moderate* group included CD47 (4.0%, 19 designs), B7-H3 (2.7%, 11 designs), EGFR (1.6%, 5 designs), MSLN-Nterm (1.1%, 6 designs), and MSLN-Cterm (1.0%, 6 designs). The *low-tractability* group comprised EphA2 (0.9%, 3 designs), GPC2 (0.8%, 2 designs), and both EGFRvIII and HER2 Domain IV (0.3%, 1 design each).

3.7 Score Distributions

3.7.1 Predicted Aligned Error (pAE)

The pAE distributions (Figure 3) showed that most designs across all campaigns exhibited median pAE values in the 4.9–9.3 range, with the majority of designs falling above the 10.0 threshold. MSLN-Nterm (median pAE = 5.23) and MSLN-Cterm (median pAE = 4.86) exhibited the most favorable pAE distributions overall, suggesting that the mesothelin epitopes are particularly amenable to high-confidence interface prediction. CD47 (median pAE = 7.34) also performed well. In contrast, GPC2 (median pAE = 9.28) and EGFR (median pAE = 9.01) showed the least favorable pAE distributions, with their medians approaching the filter threshold.

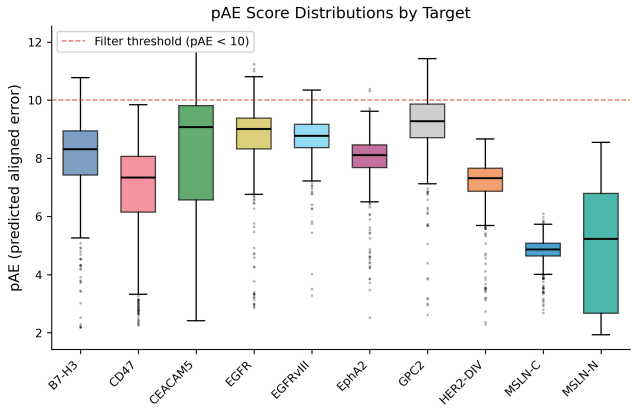

Figure 3: Box plots of pAE score distributions across all 10 targets. Red dashed line indicates the filter threshold (pAE < 10). Lower values indicate higher predicted structural confidence at the antibody–antigen interface.

3.7.2 CDR RMSD

Target-aligned CDR RMSD distributions (Figure 4) revealed that the 2.0 Å threshold was the more stringent of the two filters for most campaigns. HER2 Domain IV exhibited the highest median CDR RMSD (21.18 Å), indicating that the vast majority of designed CDR conformations deviated substantially from the target-docked geometry. MSLN-Cterm (20.41 Å) showed a similarly high median. The low-RMSD tail—designs with CDR RMSDs below 2.0 Å—was most populated for CEACAM5, explaining its dominant pass rate. MSLN-Nterm achieved the lowest median CDR RMSD of 8.40 Å despite its moderate pass rate, suggesting that many designs achieved partial structural recapitulation even when not meeting the stringent 2.0 Å cutoff.

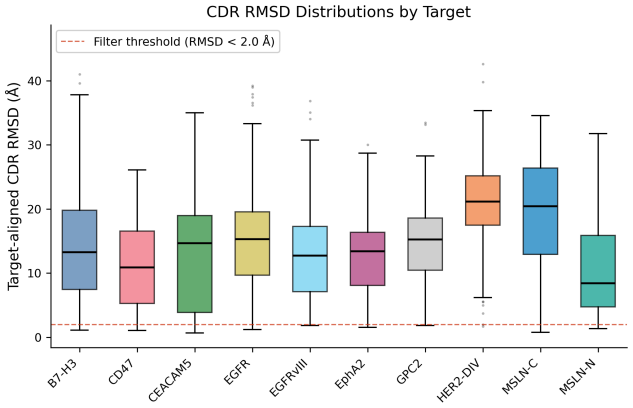

Figure 4: Box plots of target-aligned CDR RMSD distributions. Red dashed line at 2.0 Å indicates the filter threshold. The wide distributions (medians 8–21 Å) demonstrate that CDR structural recapitulation is the primary bottleneck for design success.

3.7.3 Joint pAE–RMSD Distribution

The joint distribution of pAE and CDR RMSD across all 4,085 designs (Figure 5) illustrates the combined effect of both filters. The passing region (pAE < 10, RMSD < 2.0) occupies the lower-left quadrant and is sparsely populated, with 135 designs. A notable cluster of designs from multiple campaigns achieves low pAE (<5) but high RMSD (>5), indicating correct interface prediction but incorrect CDR geometry—a failure mode where the antibody framework is well-positioned but the CDR loops adopt non-functional conformations.

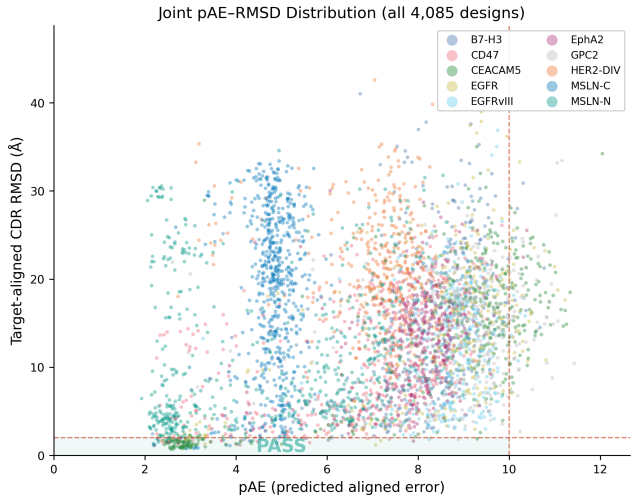

Figure 5: Joint scatter plot of pAE vs. CDR RMSD for all 4,085 designs, colored by campaign. The shaded lower-left quadrant marks the pass region (135 designs, 3.3%). The dense cluster at low pAE / high RMSD represents correctly positioned but geometrically deviated CDR loops.

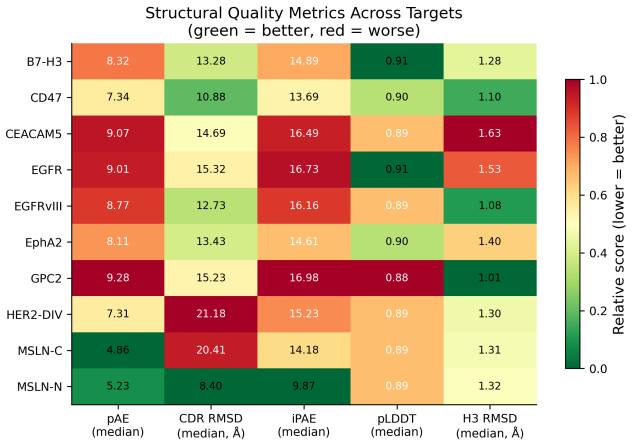

Figure 6: Heatmap of key structural quality metrics across all 10 campaigns. Values are raw scores; color scale is normalized per metric (green = better, red = worse). pLDDT is inverted so that green indicates higher confidence.

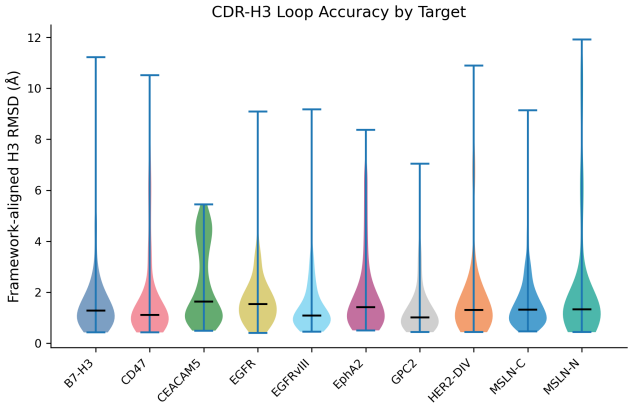

Figure 7: Violin plots of framework-aligned CDR-H3 loop RMSD. The consistently low H3 RMSDs (medians 1.0–1.6 Å) indicate that CDR-H3 loop geometry is well-preserved by the ProteinMPNN–RF2 pipeline, and that the primary design challenge lies in target-aligned positioning rather than intrinsic loop accuracy.

3.8 Structural Quality Metrics

Beyond the primary filtering metrics, RF2 generates several additional structural quality indicators. The metric heatmap (Figure 6) compares five key metrics across all campaigns. Predicted LDDT values were uniformly high (0.88–0.91), indicating that the global VHH fold quality was well-preserved across all targets regardless of interface quality. Framework-aligned H3 RMSD (Figure 7)—a measure of CDR-H3 loop accuracy independent of target positioning—showed medians of 1.01–1.63 Å across campaigns. GPC2 (1.01 Å) and EGFRvIII (1.08 Å) exhibited the lowest H3 loop deviations, suggesting that the designed H3 loops are geometrically sound even when the overall target-aligned RMSD is high.

3.9 Per-Campaign Detailed Results

3.9.1 CEACAM5 (PDAC) — 81 passing designs, 19.8% pass rate

CEACAM5 was the clear standout campaign, producing more passing designs than all other campaigns combined. The best candidate (design 76_dldesign_2) achieved pAE = 2.59, CDR RMSD = 0.68 Å, interaction pAE = 3.56, and framework-aligned H3 RMSD = 0.70 Å. Twelve designs achieved CDR RMSD below 1.0 Å, and 81

designs passed both thresholds. The high success rate may reflect the accessibility of the CEACAM5 A3B3 domain epitope defined in the tusamitamab complex (PDB: 8BW0) and the extended CDR H3 range (10–18 residues) allowed for this campaign. CEACAM5 is the target of the ADC tusamitamab ravtansine (SAR408701), which reached Phase II clinical evaluation in solid tumors including non-small cell lung cancer [12], validating this epitope as therapeutically relevant.

3.9.2 CD47 (GBM) — 19 passing designs, 4.0% pass rate

CD47 produced the second-highest number of passing designs. The best candidate (design 20_dldesign_4) achieved pAE = 2.84 and CDR RMSD = 1.05 Å. The magrolimab epitope on CD47 (PDB: 5IWL) is a well-characterized "don't eat me" signal that prevents macrophage-mediated phagocytosis of tumor cells. Nineteen designs spanning 8 unique backbone scaffolds passed both filters, suggesting robust sequence diversity among the hits. The median pAE of 7.34 across all 480 designs indicates moderately favorable interface confidence.

3.9.3 B7-H3 (MPNST, DIPG, GBM, NB) — 11 passing designs, 2.7% pass rate

The pan-cancer target B7-H3 produced 11 passing designs from 3 unique backbone scaffolds (designs 23, 30, and 68). The best candidate (design 23_dldesign_3) achieved pAE = 2.20 and CDR RMSD = 1.10 Å, indicating high structural confidence. The concentration of hits on 3 of 82 backbones suggests that a small subset of scaffold topologies are particularly well-suited to the B7-H3 epitope defined in the recently deposited 9LME structure. Given B7-H3's relevance to four cancer indications and the FDA Breakthrough designation for B7-H3 CAR-T in DIPG [6], these 11 candidates represent high-priority designs for experimental validation.

3.9.4 MSLN-Nterm (PDAC) — 6 passing designs, 1.1% pass rate

Mesothelin N-terminal domain produced 6 passing designs, predominantly from 2 backbone scaffolds (designs 30 and 32). The best candidate (design 32_dldesign_2) achieved the lowest pAE of any design across all campaigns (2.19 Å) with CDR RMSD = 1.33 Å and interaction pAE = 2.60. The exceptional pAE distribution of the MSLN-Nterm campaign (median pAE = 5.23, the second-lowest across all campaigns) suggests that the amatuximab epitope (PDB: 4F3F) provides a particularly favorable binding interface for VHH design. Combined with the MSLN-Cterm campaign, the two MSLN campaigns together produced 12 designs targeting non-overlapping epitopes, enabling potential bispecific VHH pairing for avidity-enhanced PDAC therapy.

3.9.5 MSLN-Cterm (PDAC) — 6 passing designs, 1.0% pass rate

Despite having the highest median CDR RMSD (20.41 Å) across all campaigns, the C-terminal mesothelin domain produced 6 passing designs, 4 of which originated from a single backbone scaffold (design 71). The best candidate (design 71_dldesign_2) achieved pAE = 2.92, CDR RMSD = 0.78 Å, and the lowest H3 RMSD of any top candidate in this study (0.60 Å). The concentration of 4 passing designs from one backbone with consistent CDR RMSD values (0.78–0.88 Å) strongly suggests that backbone 71 adopts a topology uniquely complementary to the 15B6 epitope on mesothelin C-terminus (PDB: 7U8C) [13].

3.9.6 EGFR (GBM, MPNST, PDAC) — 5 passing designs, 1.6% pass rate

EGFR produced 5 passing designs from 4 backbone scaffolds. The best candidate (design 19_dldesign_2) achieved pAE = 3.32 and CDR RMSD = 1.31 Å. Despite EGFR having the richest structural data of any target in this study (Tier 1 with the cetuximab complex at 2.6 Å), the relatively low pass rate (1.6%) reflects the large and complex cetuximab epitope (30 residues), which requires a VHH to replicate much of the interface formed by a full Fab domain. This result suggests that large epitope footprints may be inherently more challenging for single-domain VHH design.

3.9.7 EphA2 (GBM) — 3 passing designs, 0.9% pass rate

Three designs passed for the EphA2 receptor tyrosine kinase. The best candidate (design 27_dldesign_0) achieved a pAE of 2.52 and CDR RMSD of 1.53 Å. The moderate-to-low pass rate may reflect the geometry of the 1C1 Fab epitope (PDB: 3SKJ, 21 residues), which spans a ridge on the EphA2 ligand-binding domain that is geometrically challenging for single-domain antibody engagement [14].

3.9.8 GPC2 (Neuroblastoma) — 2 passing designs, 0.8% pass rate

Glypican-2 yielded only 2 passing designs, both with borderline metrics (pAE = 8.28–8.89, CDR RMSD = 1.82–1.99 Å). The high median pAE (9.28) and high median CDR RMSD (15.23 Å) indicate that the D3 Fab epitope on GPC2 (PDB: 6WJL, resolution 3.3 Å) is the most structurally challenging target in this study. The low template resolution may contribute to lower prediction confidence. Despite the difficulty, the 2 passing designs represent the first computationally generated VHH candidates against this neuroblastoma-restricted orphan target [11].

3.9.9 EGFRvIII (GBM) — 1 passing design, 0.3% pass rate

Only a single design passed for the tumor-specific EGFRvIII deletion variant. This sole candidate (design 19_dldesign_0) exhibited borderline pAE (8.24) and CDR RMSD (1.82 Å), with the highest interaction pAE (15.14) of any passing design across all campaigns. The EGFRvIII template (PDB: 8UKV at 2.9 Å) is of moderate resolution, ruling out template quality as the primary explanation for design difficulty. The difficulty may arise from the unique structural features of the EGFRvIII deletion junction, which creates a neoepitope formed by the fusion of residues flanking the exon 2–7 deletion—a geometrically novel surface that the diffusion model may have limited training data to model.

3.9.10 HER2 Domain IV (GBM, MPNST) — 1 passing design, 0.3% pass rate

The membrane-proximal Domain IV of HER2 (trastuzumab epitope) produced a single passing design (design 16_dldesign_2) with excellent pAE (2.36) and CDR RMSD (1.67 Å). Despite the low pass rate, this candidate shows strong interface confidence (interaction pAE = 3.06, pLDDT = 0.92). The extremely high median CDR RMSD (21.18 Å) for this campaign—the highest across all targets—suggests that the trastuzumab epitope on Domain IV is particularly challenging for VHH design, possibly because Domain IV is a small, convex domain that provides limited surface area for single-domain antibody engagement compared to the larger epitope footprint exploited by the full trastuzumab Fab.

3.10 Cross-Campaign Comparison

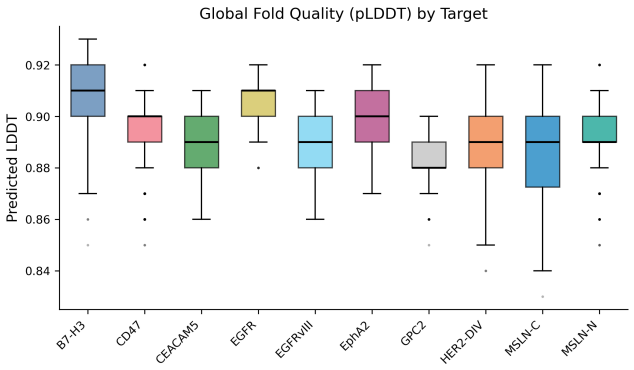

Figure 8: Predicted LDDT distributions by target. The uniformly high pLDDT values (0.87–0.92) across all campaigns confirm that the VHH framework fold is well-preserved regardless of interface quality or target difficulty.

Several patterns emerged from the cross-campaign comparison. First, global fold quality (pLDDT) was remarkably consistent across all targets (medians 0.88–0.91; Figure 8), confirming that the NbBCII10 VHH framework is robustly recapitulated by the ProteinMPNN–RF2 pipeline regardless of the target. Second, CDR RMSD was the dominant determinant of pass rate—the correlation between median CDR RMSD and pass rate was strongly negative (lower RMSD median, higher pass rate). Third, pAE and CDR RMSD captured distinct failure modes: some campaigns (MSLN-Cterm, HER2

Domain IV) had low median pAE but very high median RMSD, indicating that the interface was correctly predicted but the CDR geometry was wrong; other campaigns (GPC2) had high median values on both metrics.

3.11 Top Candidates

Table 4 summarizes the single best candidate from each campaign. The three highest-confidence designs are from CEACAM5 (pAE = 2.59, RMSD = 0.68 Å), MSLN-Nterm (pAE = 2.19, RMSD = 1.33 Å), and B7-H3 (pAE = 2.20, RMSD = 1.10 Å). These candidates represent immediate priorities for experimental validation via yeast surface display and surface plasmon resonance.

Table 4: Best candidate per campaign ranked by composite score.

Target	pAE	RMSD	Score	Design ID
CEACAM5	2.59	0.68	0.015	76_dl2
MSLN-N	2.19	1.33	0.000	32_dl2
B7-H3	2.20	1.10	0.000	23_dl3
CD47	2.84	1.05	0.050	20_dl4
MSLN-C	2.92	0.78	0.000	71_dl2
EphA2	2.52	1.53	0.000	27_dl0
EGFR	3.32	1.31	0.129	19_dl2
HER2-DIV	2.36	1.67	0.000	16_dl2
GPC2	8.28	1.99	0.429	5_dl2
EGFRvIII	8.24	1.82	0.000	19_dl0

3.12 Backbone Topology and Design Success

A notable observation is the concentration of passing designs on a small number of backbone scaffolds. For B7-H3, 11 passing designs originated from only 3 of 82 backbones (3.7%). For MSLN-Cterm, 4 of 6 passing designs came from a single backbone (design 71). This pattern suggests that backbone topology is the primary determinant of design success, and that increasing the number of backbones (rather than sequences per backbone) would be the most effective strategy for improving pass rates in future campaigns (Figure 10).

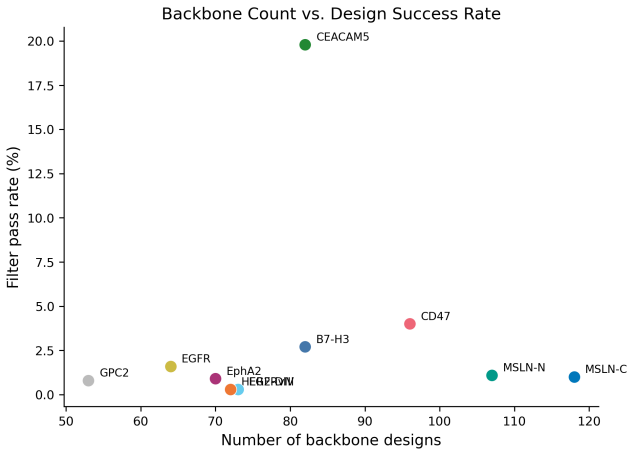

Figure 10: Relationship between backbone count and filter pass rate. No strong correlation is observed, indicating that backbone quality (topology) rather than quantity drives design success.

4 Discussion

4.1 Computational Design Difficulty Varies Dramatically Across Cancer Targets

The most striking finding of this study is the 60-fold range in pass rates across the 10 cancer targets (0.3% to 19.8%). This variation substantially exceeds what would be expected from sampling noise alone and reflects genuine differences in the tractability of different epitope–antigen pairs for VHH nanobody design. CEACAM5's outlier success (19.8%) may be attributable to the geometry of the tusamitamab epitope on the A3B3 domain, which presents a relatively concave surface patch well-suited to single-domain antibody engagement, combined with the extended H3 loop lengths (10–18 residues) allowed for this campaign. In contrast, the near-complete failure of EGFRvIII design (0.3%) highlights the intrinsic difficulty of neoepitopes formed by deletion junctions, where the diffusion model has limited structural precedent to draw upon.

4.2 CDR RMSD as the Dominant Design Bottleneck

Our analysis identifies target-aligned CDR RMSD—not pAE—as the primary determinant of design success. Median CDR RMSDs of 8–21 Å across

Top Candidate per Target (ranked by composite score)

Target	Best Design	pAE	CDR RMSD	iPAE	pLDDT	H3 RMSD	Score
B7-H3	samples_design_23_dldesig	2.20	1.10	2.90	0.92	1.02	0.000
CD47	samples_design_20_dldesig	2.84	1.05	3.95	0.91	0.90	0.050
CEACAM5	samples_design_76_dldesig	2.59	0.68	3.56	0.91	0.70	0.015
EGFR	samples_design_19_dldesig	3.32	1.31	5.14	0.90	1.49	0.129
EGFRvIII	samples_design_19_dldesig	8.24	1.82	15.14	0.89	0.66	0.000
EphA2	samples_design_27_dldesig	2.52	1.53	3.22	0.91	1.22	0.000
GPC2	samples_design_5_dldesign	8.28	1.99	15.05	0.89	0.53	0.429
HER2-DIV	samples_design_16_dldesig	2.36	1.67	3.06	0.92	0.66	0.000
MSLN-C	samples_design_71_dldesig	2.92	0.78	5.99	0.90	0.60	0.000
MSLN-N	samples_design_32_dldesig	2.19	1.33	2.60	0.91	0.68	0.000

Figure 9: Best-scoring candidate per target, ranked by composite score. Columns show predicted aligned error (pAE), target-aligned CDR RMSD, interaction pAE (iPAE), global fold confidence (pLDDT), CDR-H3 loop accuracy (H3 RMSD), and composite score. The strongest candidates (CEACAM5, B7-H3, MSLN-N) achieve pAE below 3.0 and CDR RMSD below 1.5 Å.

campaigns indicate that the majority of designs produce CDR loops that are structurally reasonable in isolation (framework-aligned H3 RMSD medians of 1.0–1.6 Å) but incorrectly positioned relative to the target antigen. This suggests that the RFdiffusion backbone generation step, which determines CDR placement relative to the target, is the rate-limiting step in the pipeline. Improvements to the diffusion model's epitope conditioning or the introduction of iterative refinement of CDR–target docking could substantially improve pass rates.

4.3 The Structural Data Bottleneck

Our analysis reveals that structural data availability, not biological understanding, is the primary bottleneck limiting computational antibody design for cancer targets. Several targets with compelling clinical evidence—CLDN18.2 (FDA Fast Track for PDAC), IL-13Rα2 (dramatic complete responses in GBM CAR-T trials), ErbB3/HER3 (functionally validated kinase-dead RTK in MPNST)—cannot currently be subjected to structure-guided de novo design because no antibody–antigen complex structures exist in the public domain. The recent deposition of 9LME (B7-H3 nanobody complex, January 2025) illustrates how a single structural determination can unlock computational design for a target relevant to four cancer indications simultaneously.

4.4 VHH Nanobodies as a Privileged Format

The consistent selection of VHH format across all 10 campaigns reflects a deliberate therapeutic strategy. For CNS tumors (DIPG, GBM), the ∼10-fold size reduction from IgG (∼150 kDa) to VHH (∼15 kDa) may improve blood–brain barrier penetration [7]. For pancreatic cancer, smaller formats offer superior stromal penetration. The uniformly high pLDDT values (0.87–0.92) across all campaigns confirm that the NbBCII10 VHH framework is robustly modeled by the RFAntibody pipeline, validating VHH as a reliable computational design format.

4.5 Implications for Campaign Design Strategy

Our results suggest several practical recommendations for future computational antibody design campaigns. First, backbone count should be prioritized over sequences per backbone, given the observed concentration of hits on a small fraction of backbones. Second, CDR H3 loop length ranges should be generous, as the CEACAM5 campaign's extended range (10–18 residues) likely contributed to its success. Third, targets with large epitope footprints (e.g., EGFR cetuximab epitope, 30 residues) may be better suited to Fab or scFv formats rather than single-domain VHH. Fourth, campaigns pro-

ducing zero or borderline passing designs should be re-run with 10–100× more backbones before concluding that a target is computationally intractable.

4.6 Cross-Indication Efficiency

The identification of cross-indication driver targets creates efficiency in the design-to-clinic pipeline. A single high-affinity VHH nanobody against B7-H3 could serve as a therapeutic or diagnostic agent across MPNST, DIPG, neuroblastoma, and GBM—four indications that collectively affect approximately 15,000 patients annually. The 11 B7-H3 candidates from this study represent a starting point for this multi-indication strategy. Similarly, the 12 combined mesothelin candidates targeting two non-overlapping epitopes enable bispecific VHH construction for enhanced PDAC therapy.

4.7 Model Bias Versus Intrinsic Target Difficulty

A natural question arising from the 60-fold variation in pass rates is whether the observed differences reflect genuine biophysical design difficulty or systematic model bias in the RFdiffusion–ProteinMPNN–RF2 pipeline. Several lines of evidence favor the design-difficulty interpretation over model bias. First, the pipeline's global fold quality metric (pLDDT) was remarkably uniform across all 10 targets (medians 0.88–0.91), indicating that the model treats the VHH framework consistently regardless of target identity. Second, the variation is not explained by template resolution: CEACAM5 (19.8% pass rate) used a 3.1 Å template while EGFRvIII (0.3%) used a 2.9 Å template—the lower-resolution structure outperformed. Third, the targets that performed best (CEACAM5, CD47, B7-H3) share a common structural feature: relatively concave epitope surfaces that provide complementary pockets for single-domain antibody engagement, whereas the worst-performing targets (EGFRvIII, HER2 Domain IV) present convex or geometrically novel surfaces. Fourth, the framework-aligned H3 RMSD was consistently low across all campaigns (medians 1.0–1.6 Å), confirming that the CDR loops themselves are well-designed in isolation; the failure mode is incorrect CDR positioning relative to the target, which is governed by the geometric compatibility between epitope shape and the diffusion model's backbone sampling distribution. Nonetheless, model bias cannot be entirely excluded. RFdiffusion was trained predominantly on existing antibody–antigen complexes, which are enriched for certain epitope geometries (e.g., flat or concave surfaces) and may underrepresent neoepitopes or membrane-proximal domains. Disentangling model bias from intrinsic difficulty will require benchmarking across a larger and more diverse target set, ideally with experimental binding data as ground truth.

4.8 Limitations

Several important limitations should be acknowledged. First, the backbone counts per campaign (53–118) are substantially lower than the 10,000 used in the original RFAntibody benchmarks; this was a deliberate trade-off for rapid iteration, with scale-up runs planned for the most promising targets. Second, computational metrics from RF2 are predictive but not definitive—experimental validation remains essential. Third, our filters (pAE $<$ 10, RMSD $<$ 2.0) may be too stringent for some targets; relaxed thresholds could identify additional candidates worth testing. Fourth, the absence of $\Delta\Delta G$ scoring from the RF2 pipeline means that binding energy is not directly assessed, and the composite score relies solely on structural confidence metrics. Fifth, epitope residues are derived from reference complexes and represent one binding mode; designed VHH nanobodies may adopt different effective epitopes.

5 Conclusion

We present the first systematic computational antibody design study targeting ten challenging cancer antigens across five lethal indications. Through analysis of 43 target–indication pairs, we developed a structural readiness classification and identified 10 actionable targets. Execution of all 10 campaigns using the RFAntibody pipeline generated 4,085 designs, of which 135 (3.3%) passed stringent quality filters. Per-target pass rates varied 60-fold (0.3–19.8%), revealing that computational design difficulty is strongly target-dependent and that CDR RMSD is the dominant bottleneck. The highest-confidence candidates—CEACAM5 (pAE

= 2.59, RMSD = 0.68 Å), MSLN-Nterm (pAE = 2.19, RMSD = 1.33 Å), and B7-H3 (pAE = 2.20, RMSD = 1.10 Å)—represent immediate priorities for experimental validation. All code, configurations, analysis, and designed sequences are openly available as resources for the computational biology and cancer immunotherapy communities.

Data and Code Availability

The rfab-harness campaign orchestration tool, all 10 cancer driver campaign configurations, analysis scripts, and complete results including all score CSVs are available at: https://github.com/inventcures/repro_rfantibody_for-cancer-targets. The tool requires the RFAntibody repository [1] for pipeline execution.

Acknowledgments

The author thanks the RFAntibody team (Baker Lab, University of Washington) for making their pipeline publicly available. Target research was informed by the Open Targets Platform, RCSB Protein Data Bank, and ClinicalTrials.gov. Computational resources were provided by Modal (modal.com).

References

[1] Bennett NR, Watson JL, et al. Atomically accurate de novo design of antibodies with RFdiffusion. *Nature*. 2025. doi:10.1038/s41586-025-09721-5

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

[3] Boitreaud J, Dent J, Geisz D, et al. (Chai Discovery Team). Zero-shot antibody design in a 24-well plate. *bioRxiv*. 2025. doi:10.1101/2025.07.05.663018

[4] Arvanitis CD, Ferraro GB, Jain RK. The blood–brain barrier and blood–tumour barrier in brain tumours and metastases. *Nature Reviews Cancer*. 2020;20(1):26–41.

[5] Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. *Nature Reviews Gastroenterology & Hepatology*. 2020;17(8):487–505.

[6] Vitanza NA, et al. Intracerebroventricular B7-H3-targeting CAR T cells for diffuse intrinsic pontine glioma: a phase 1 trial. *Nature Medicine*. 2025;31:861–868. doi:10.1038/s41591-024-03451-3

[7] Li T, et al. Cell-penetrating anti-GFAP VHH and corresponding fluorescent fusion protein VHH-GFP spontaneously cross the blood–brain barrier and specifically recognize astrocytes. *FASEB Journal*. 2012;26(10):3969–3979. doi:10.1096/fj.11-201384

[8] Ochoa D, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Research*. 2023;51(D1):D1353–D1359. doi:10.1093/nar/gkac1046

[9] Dauparas J, et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science*. 2022;378(6615):49–56.

[10] Baek M, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021;373(6557):871–876.

[11] Bosse KR, et al. Identification of GPC2 as an oncoprotein and candidate immunotherapeutic target in high-risk neuroblastoma. *Cancer Cell*. 2017;32(3):295–309.

[12] Kumar A, et al. Structural insights into epitope–paratope interactions of a monoclonal antibody targeting CEACAM5-expressing tumors. *Nature Communications*. 2024;15:9377. doi:10.1038/s41467-024-53746-9

[13] Zhan J, Lin D, Watson N, et al. Structures of cancer antigen mesothelin and its complexes with therapeutic antibodies. *Cancer Research Communications*. 2023;3(2):175–191. doi:10.1158/2767-9764.CRC-22-0306

[14] Peng L, Oganesyan V, Damschroder MM, Wu H, Dall'Acqua WF. Structural and functional characterization of an agonistic anti-human EphA2 monoclonal antibody. *J. Mol. Biol.* 2011;413(2):390–405. doi:10.1016/j.jmb.2011.08.018