MANIFOLD BINDER ENGINEERING AND REFINEMENT

mBER-2

Broadening epitopic coverage for therapeutic targets with *de novo* design of VHH, minibinders, and structured peptides

Manifold Bio Team

September 2026

Abstract

We report on mBER-2, the most recent iteration of our scaled protein binder design system. We evaluate its design capabilities by designing binders in multiple formats to a panel of 56 therapeutically-relevant targets, including historically challenging target classes such as GPCRs. Across formats, we observe successful designed binders to 98% (55/56) of these targets, with binders to multiple predicted epitope clusters on 91% of them. The pooled VHH hit rate is 8.6%, rising to 17% after retrospective top-100 selection per target. mBER-2 also advances design capabilities beyond VHH—in the same design campaign, we introduced 50 structured peptide scaffolds ranging from 10 to 45 amino acids, which we term "microbinders," drawn from natural disulfide-constrained folds and engineered miniaturized domains. Using mBER-2, we designed binding interfaces into these scaffolds, with 31 of the 50 microbinder scaffolds yielding specific on-design hits, and 14 of the 50 microbinder scaffolds yielding at least 1% success in the overall or top-100 filtered designs. Lastly, we demonstrate structure-driven refinement of existing binders with mBER-2, resulting in improved affinity and specificity characteristics, while diversifying by as many as 20 mutations from the parent sequence while maintaining function. Design of binders at this scale and fidelity enables the exploration of all epitopes in the human proteome, a project we term the EpiTome.

Introduction

The field of computational protein design has developed at a dramatic pace over the course of the past decade. The release of foundational tools including AlphaFold2 [1], RFdiffusion [2], and ProteinMPNN [3] have been broadly enabling in both academia and industry for the design of protein-protein interactions. Several updates, derivations, and new modeling approaches have followed, with recent developments expanding the binder design space to antibody formats. These include mBER-open [4], Germinal [5], Chai-2 [6], JAM-2 [7], and BoltzGen [8].

As binder design capabilities have improved, progress has been tracked by increases in per-binder hit rates, hits raised to unseen targets, epitopic coverage, bespoke design problems, and developability [7], [9], [10].

In this work, we introduce mBER-2, the next generation of mBER (Manifold Binder Engineering and Refinement), which we released as mBER-open last fall [4]. mBER-2 follows a similar generative approach to mBER-open, but with substantial optimizations intended to improve inference speed and efficiency, epitope-targeting capabilities, binder properties, and overall hit rates. These optimizations were enabled by multiple rounds of high throughput experimental characterization, in which we

repeatedly evaluated various algorithmic changes to mBER with respect to experimental target coverage and hit rate. We report on a recent design campaign we ran against 56 targets which were selected to advance therapeutic tissue-targeting programs. In this campaign, we successfully raised specific, on-design binders to 55/56 (98%) of those targets (Supplementary Methods S1.4); 51/56 (91%) targets were bound on at least 2 predicted epitope clusters, and for some targets we hit as many as 21 unique clusters. Selected VHH binders generated by mBER-2 exhibit nanomolar affinities, with the best BLI-measured affinity at 8.2 nM.

In addition to frontier antibody design capabilities, we showcase two additional features of mBER-2. Using mBER-2, we design novel binding interfaces into a collection of 50 small "microbinder" scaffolds—structured peptides ranging from 10–45 amino acids in length, obtaining specific hits for 31 scaffolds. We also follow up on previous experimental mBER hits and show that a second round of optimization with mBER-2 efficiently explores functional sequence space under high mutational loads, compared with deep mutational scanning (DMS) and ESM2-guided designs [11].

Finally, we discuss how we are using mBER-2 to generate a library of binders against the EpiTome, the space of all addressable epitopes in the human surface proteome. The purpose of the EpiTome library is to take advantage of our proprietary high throughput *in vivo* assays to build a comprehensive map of biodistribution for binders against every accessible protein in the human body. To this end, we have scaled our *in silico* methods to efficiently sample millions of designs against tens of thousands of putative epitopes.

Results

We selected 56 target proteins for design, including GPCRs, receptor kinases, cadherins, adhesion proteins, ectoenzymes, membrane transport proteins, and other membrane-anchored proteins. These targets were selected to advance therapeutic tissue-targeting programs. For each target, we used mBER-2 to select a set of accessible epitopes and generated binder designs across targets and binder formats. 1,298,334 sequences were prioritized and ordered in oligopool format from Twist for expression in a set of phage display libraries. The phage libraries were screened against all 56 targets as previously described [4] (Supplementary Methods S2.1–S2.4). The primary analysis includes 1,250,720 *de novo* on-panel sequences, of which 584,450 have high-confidence sequencing signal (hereafter, high-signal designs; Supplementary Methods S1.4).

Following the phage binding experiment, hits were called based on NGS counts relative to all other target wells (Supplementary Methods S1.4). Combining all formats, 26,026 binder designs were identified as specific, on-design hits (Table S1). 55 of the 56 targets received at least one specific, on-design hit (Figure 1). Furthermore, 51 of 56 targets (91%) received at least one hit on at least two unique predicted epitope clusters (Figure 1b; Supplementary Methods S1.1). A subset of high-quality binders were later characterized for binding affinity on BLI.

De novo VHH design raises binders to 54 of 56 targets with double digit hit rates

Since the release of mBER-open, we have performed several large scale binder design campaigns with experimental validation to drive greater target and epitope coverage with higher hit rates. Like mBER-open, mBER-2 combines gradient-based sequence optimization through a folding model with a collection of other design and evaluation models to generate binder sequences and their complex structures with target epitopes.

In our VHH design protocol, mBER-2 takes in a target and a partial VHH sequence, in which framework residues are mostly defined, and CDR residues are entirely masked. The target is partitioned into a set of putative epitopes, against which potential VHH binders are repeatedly docked, using gradients from protein folding and auxiliary models to iteratively design and optimize the sequence. At the end of this

iterative process, the final design is redocked and expanded into a larger set of potential sequence designs by resampling from the design trajectory and incorporating additional inverse folding models (Supplementary Methods S1.2; Figure S1).

Against 54 of the 56 total design targets, we raised at least one successful, designed VHH hit (Figure 1). Our VHH designs access a large space of unique epitopes, with an average of 5 unique predicted epitope clusters bound per target (Figure 1b). Using one-sided Fisher exact tests to compare on-design hits and off-design hits for each target, 50/56 targets were significant at nominal $P < 0.05$ and 48/56 (86%) after Bonferroni correction across 56 targets (Supplementary Methods S1.6; Table S9). Selected VHH hits were later assessed on BLI, with affinities ranging from 8.2 to 362.1 nM (29 binders; Figures 2b and S4).

Hit rates per design vary across targets and epitopes, an effect that has been observed broadly in the binder design literature. The overall hit rate for VHH in the library was 8.6% among high-signal designs. A retrospective analysis reveals that ranking with mBER-2's design evaluation model and filtering to the top 100 binders among high-signal designs per target elevates hit rates to 16.8% (941/5,600). Separately, selecting the most productive design-hotspot group for each target from the full high-signal cohort yielded a mean target-level VHH hit rate of 37.6%; ranking and selecting up to ten designs within each selected group increased the mean to 45.9%, with several design hotspots reaching 100% binding among the top ten ranked designs (Figure S2). Under the primary hit calls, selecting the top ten VHH designs per target across all hotspots yielded 96/560 hits (17.1%) and reached 23/56 targets; top-100 selection reached 40/56 targets (Table S11). While this analysis suggests that we may substantially improve hit rates by applying our ranking and filtering process more strictly in advance, we intentionally opt to saturate our testing funnel and maximize the number of candidates that we move forward to experimental testing in order to generate useful data for training and evaluation.

mBER-2 designs novel formats beyond VHH

Designing minibinders with mBER-2

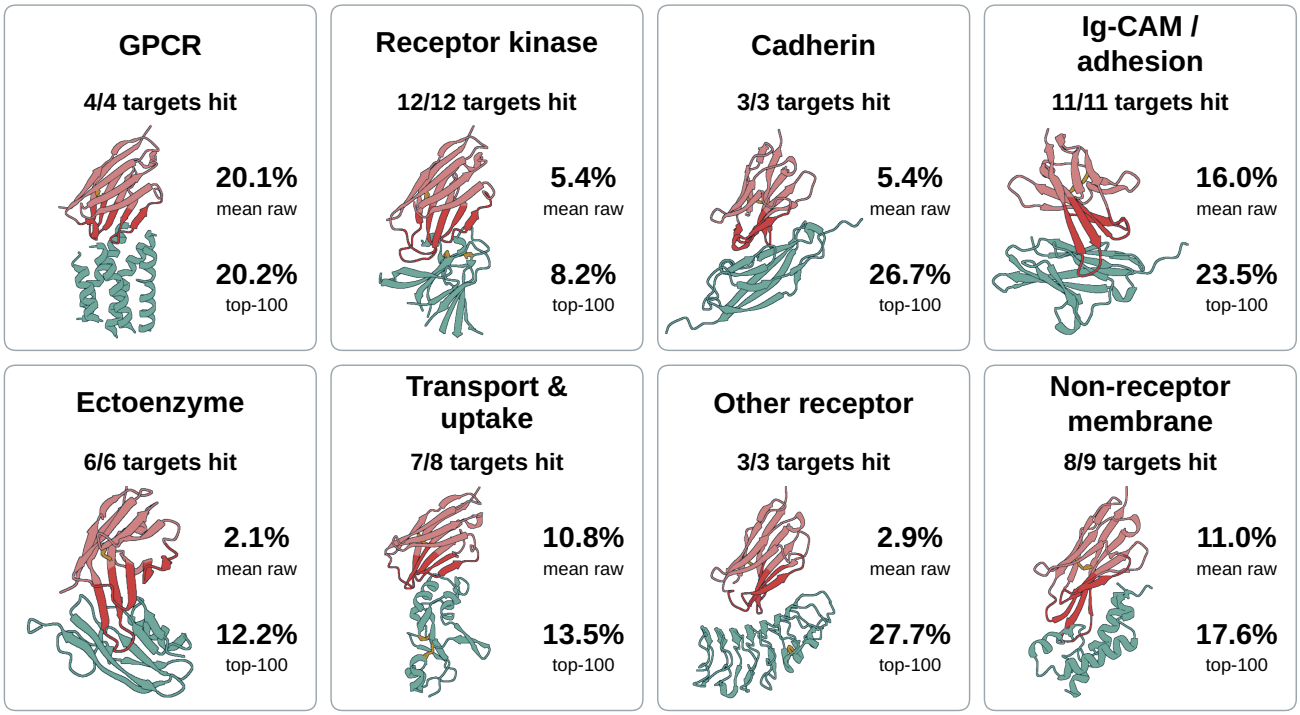
We also applied mBER-2 to design minibinders of 80–88 residues without providing a sequence or structural template for the binder. All binder positions were designable, with no amino acid restrictions or imposed secondary-structure biases. Binder sequences were optimized against selected target epitopes through the same generative process as our VHH protocol, allowing the binder fold and binding interface to emerge during optimization. This follows the general AlphaFold-based binder design approach used by methods such as BindCraft [12]. Candidate designs were ranked and selected in the same manner as the VHH library. 462,470 minibinder sequences were included in the minibinder phage display library.

Minibinder designs yielded specific, on-design hits against 46 of 56 targets (82%), with an overall hit rate of 4.64% among high-signal designs (Figure 2). Retrospective selection of the top 100 designs per target by mBER-2's design evaluation model increased the hit rate to 8.62%. Selected minibinders were subsequently characterized by BLI, with K_D values ranging from 8.4 to 347.0 nM (Figures 2b and S4). Minibinder hits also covered predicted epitopes that were not represented among the VHH hits in this campaign, expanding the combined epitopic coverage, with an average of 1.77 unique predicted epitope clusters per target hit by minibinders. Minibinder hit rates were markedly lower at β-strand-dominated epitopes than at helix-dominated epitopes, whereas VHH designs retained substantial success at both (Figure S3).

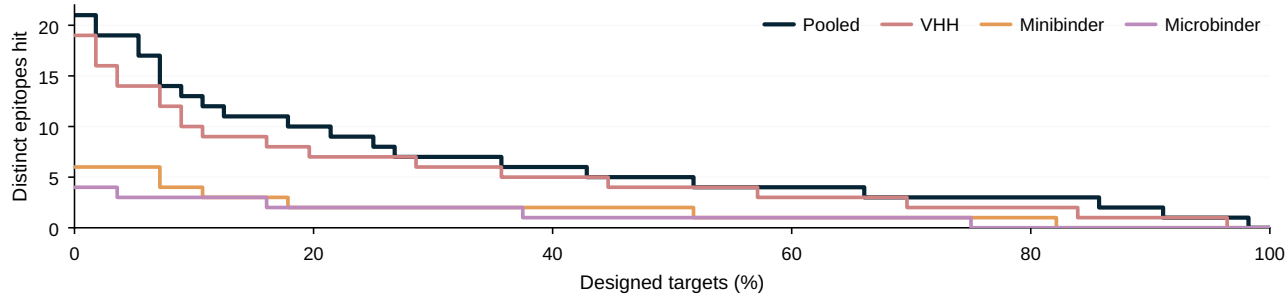
Designing microbinders with mBER-2

Design of binders smaller than 40 residues has been challenging to date, though recent methods have shown some success in peptide binder design [8], [13]. Binders of this size are attractive formats for drugs, as their small size may give them distinct pharmacokinetic and developability properties, with

a **VHH binders across eight target classes**



b **Distinct epitopes reached across the designed target panel**



c **From the EpiTome to three distinct TFRC binding sites**

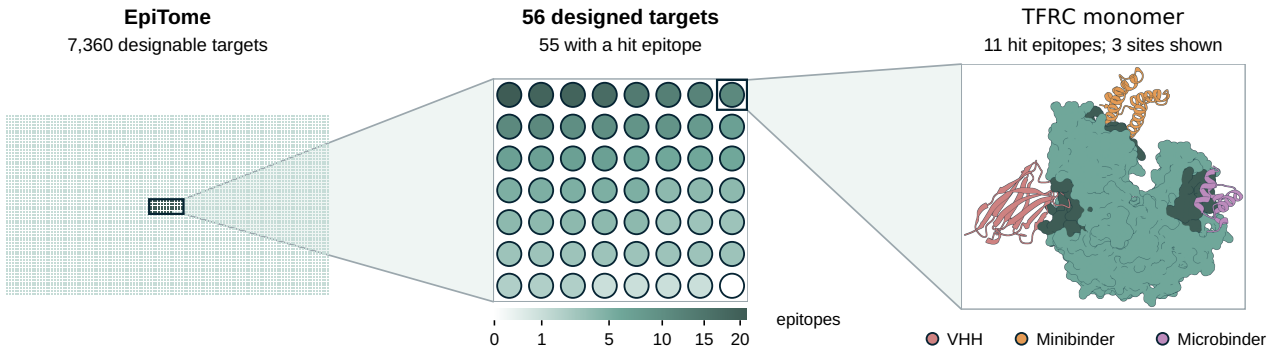


Figure 1. Target and predicted epitope coverage. a. VHH binding results for 56 targets across eight classes. Raw rates are unweighted means of target-level hit/high-signal rates; top-100 rates use ranked high-signal designs. Targets are green, VHHs pink, and designed CDR positions red. **b.** Epitope coverage over all design targets across 3 designed binder formats. The normalized areas under the curves represent the average number of unique predicted epitope clusters hit per target: 6.30 pooled (black), 5.07 VHH, 1.77 minibinder, and 1.32 microbinder. **c.** Contextual atlas of 7,360 designable protein targets across the proteome, the 56 campaign targets colored by pooled cluster count, and three illustrative sites on a TFRC ectodomain monomer. TFRC was bound on 11 unique epitopes; three exclusive sites are illustrated, each demonstrating binder hits from different formats. A VHH, minibinder, and microbinder hit are shown at their predicted epitopes, colored in pink, orange, and purple, respectively.

potential advantages in tissue penetration, chemical synthesis, and subcutaneous delivery [14]. Expanding design capability into this range is expected to further diversify the epitope and *in vivo* behavior space we can explore. We reasoned that mBER's ability to condition binder design on a defined format could be extended beyond VHH to other formats that are otherwise challenging to design.

We selected 50 compact, structured peptide scaffolds up to 45 residues in size, spanning natural disulfide-constrained peptides, engineered miniaturized domains, and designed miniproteins. Screening results are reported in Figure 2 and Table S8. Format conditioning in mBER-2 combines a partially masked sequence with a structural template to preserve scaffold features while redesigning the binding interface. For each format, positions important for structural stability were held fixed, leaving exposed positions available for design.

Microbinder designs yielded specific, on-design hits against 42 of 56 targets (75%), with an overall hit rate of 0.93% among high-signal designs (Figure 2). Retrospective selection of the top 100 designs per target by mBER-2's design evaluation model increased the hit rate to 3.55%. Of the 50 scaffolds, 31 yielded at least one hit, and 14 reached at least 1% in the overall or top-100 analysis. Selected microbinders were characterized on BLI, with affinities ranging from 113.9 to 336.8 nM (Figures 2b and S4).

mBER-2 refinement protocol performs structure-guided variant exploration

In addition to expanding the accessible binder format design space, mBER-2 also enables structure-aware mutation scanning of previously identified mBER hits, using their predicted binding interfaces to guide combinations of mutations. Upon identifying target-specific hits, we need robust methods to reliably explore the set of viable binders and sample a maximally diverse phenotypic space across stability, specificity, relative association/dissociation rate, etc. We conduct this sequence exploration within our high-throughput affinity maturation workflow, where we routinely optimize upwards of 400 binders, testing thousands of variants of each for improved specific affinity across ~30 targets in parallel.

We've recently expanded our explorative protein sequence design approaches by using mBER refinement, our method for binder optimization leveraging the structural prior informed by discovery of target-specific mBER hits (Figure 3). When we confirm an mBER variant selectively binds its intended design target, it significantly increases our confidence in the predicted structural pose. This provides a constrained optimization problem where we can sample combinations of mutations that maintain a functional binding interface while optimizing across other *in silico* metrics (Figure 3a). With identification of on-design mBER hits, we're able to apply this constraint much more broadly, sampling high combinatorial mutations from a known starting point. This high mutational sampling is intended to refine an interface with compensating mutations to balance affinity and specificity, or co-optimize a protein with stabilizing core mutations and substitutions at surface residues that might improve solubility or reduce aggregation propensity, for example.

Across two of our recent affinity maturation campaigns, we tested mBER refinement design on 98 mBER discovery hits, in parallel to DMS and ESM2-guided strategies. Variants were assessed for binding to their cognate targets and a set of various off-target soluble antigens via a phage panning assay, as well as assessed for binding to the polyspecificity reagent ovalbumin. Amongst these designs, mBER refinement samples a distinct sequence space from traditional maturation approaches (Figure 3b–c). While we deliberately only sampled 2- and 3-mutation variants based on ESM2 scoring, mBER refinement generates variants with a mean Hamming distance of 10.50 from the WT seed across 97 parents represented in all three design groups. mBER refined variants are also very different from one another, with a mean pairwise Hamming distance of 12.41 across these matched parents.

mBER refinement explores more distinct substitutions at low sampling depth, while reusing those substitutions in different variant combinations. In random samples of 1,000 designs across the 98 parents, mean counts of distinct substitutions were 4,416 for mBER, 2,203 for ESM2, and 964 for DMS. At

a **Campaign scale**

56	15.1 M	1.30 M	26,026	430
cell-surface targets	binder designs	designs ordered	binders called	tested by BLI

b **Three binder formats and 50 microbinder scaffolds**

VHH · 115–123aa
4 frameworks × 7 CDR3 lengths

Overall	8.6%
Top-100	16.8%
Targets hit	54/56
Best K_D	8.2 nM

Minibinder · 80–88aa
Unscaffolded

Overall	4.6%
Top-100	8.6%
Targets hit	46/56
Best K_D	8.4 nM

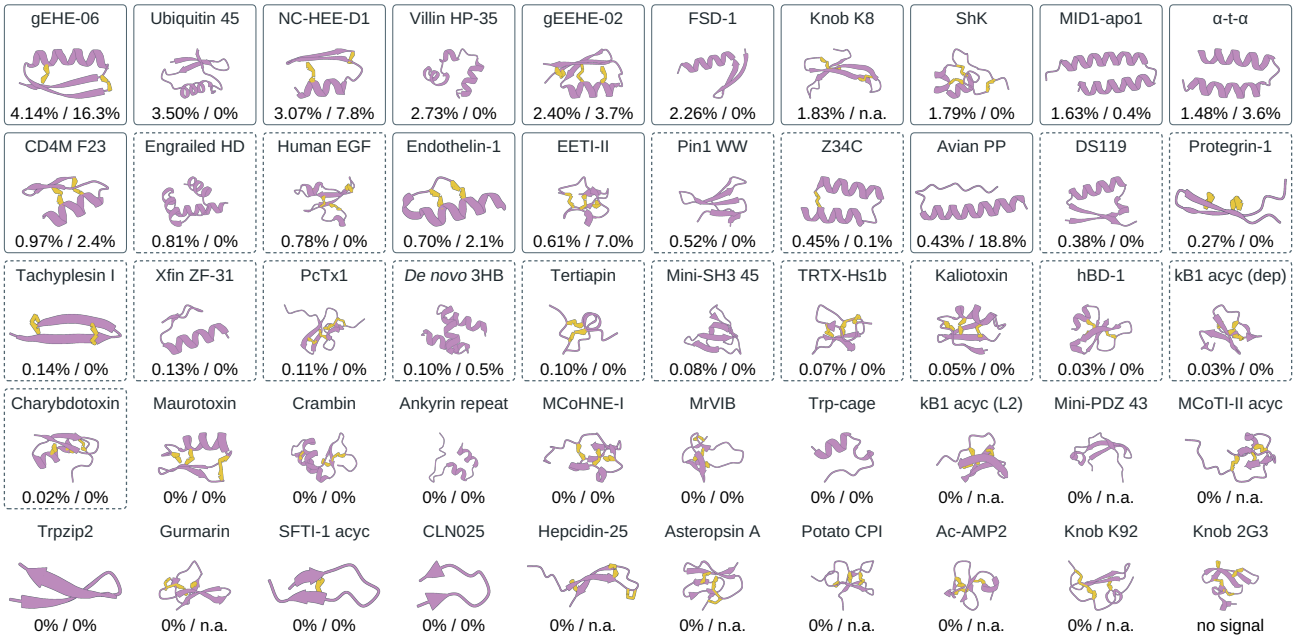
Microbinder · 10–45aa
50 scaffolds

31/50
scaffolds with successful binders

Overall	0.9%
Top-100	3.6%
Targets hit	42/56
Best K_D	113.9 nM

Microbinder scaffolds · ordered by overall hit rate

Solid: ≥1% Dashed: >0 to <1%



n.a.: undefined rate · no signal: no eligible designs

c **All-against-all binder design hit rates to 56 targets**

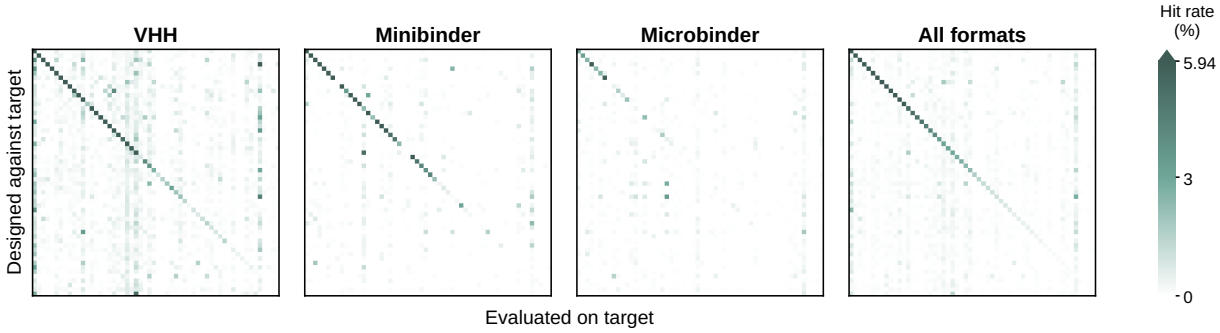


Figure 2. Campaign scale, formats, and all-against-all screening. a. Library design and measurement process. 56 cell-surface targets were selected and designed against using mBER-2, generating 15.1M binder designs. 1.3M binders were ordered in pooled libraries which were screened on phage, leading to 26,026 specific on-design binder hits. The BLI stage shows 430 tested binder–antigen pairs in the selected cohort presented in Figure S4. **b.** All binder formats in mBER-2, with their overall and top-100 ranked binder hit rates, target coverage, and best measured affinities. All 50 microbinder scaffolds are shown with descriptive scaffold names matching Table S8. Minibinders were designed without a scaffold template. 14 microbinder scaffolds achieved hit rates of at least 1% in the overall or top-100 analysis and are shown with solid outlines; 17 additional scaffolds with nonzero rates below 1% are shown with dashed outlines. Stabilizing disulfide bonds, present in many of the microbinder scaffolds, are shown in yellow. **c.** Four 56×56 matrices of hits/high-signal designs, showing the all-against-all phage binding data. Rows are intended targets; columns are evaluated antigens. All share the pooled diagonal ranking. The color scale caps at the maximum off-diagonal rate (5.94%); higher on-target rates saturate.

higher sampling depths, ESM2 and DMS covered more distinct substitutions: at 25,000 designs the corresponding counts were 12,659, 20,753, and 23,099 (Figure 3c; Supplementary Methods S1.3).

All three design groups yielded binders with improved on-target binding; however, mBER-refined variants appear to uniquely balance improvements in on-target binding with much lower polyspecificity signal (Figure 3d). mBER refinement returns higher on-target binding rates when compared with our DMS and ESM2-guided designs and a lower observed OVA-control binding rate. All three approaches showed low off-target binding rates to other soluble antigens. On-target rates were 1,371/18,013 (7.61%) for mBER, 65/1,219 (5.33%) for ESM2, and 265/4,734 (5.60%) for DMS. OVA binding rates were 3/6,449 (0.047%), 3/220 (1.36%), and 6/584 (1.03%).

We advanced many promising molecules across all three design methods for clonal affinity and developability measurements. For one representative target, all 23 nominated variants yielded clonal BLI affinities; the best mBER variant had a K_D of 21.2 nM with 11 substitutions, compared with 38.4 nM for the parent. All six mBER variants (100%) satisfied all three developability classification rules (Figure S5; Table S2). The corresponding clonal BLI curves are shown in Figure S6.

Our results here demonstrate a new approach for leveraging computational structural priors to broadly explore sequence space while seeking to co-optimize protein function. Here, mBER refinement identified sets of mutations compatible with retained binding. By incorporating structural information, we were able to sample high-order, functional mutational space. Thus, mBER refinement allows us to explore broad functional sequence-structure space to optimize binders across multiple phenotypes for downstream use.

Discussion

In this work, we present mBER-2, our latest protein binder design system. mBER-2 was designed specifically to enable efficient binder generation for novel targets and epitopes at proteome scale. Using mBER-2 we designed and screened over a million binders to a set of 56 therapeutically relevant targets, achieving successful hits to 55 of them. In addition to 98% target coverage, we demonstrate mBER-2's capability to design binders to many diverse epitopes on a single target. On average, we raised binders to 6.3 unique epitope clusters per target, with as many as 21 unique clusters bound on a single target.

The all-against-all nature of our phage-based *in vitro* binding assay ensures that only binders with high specificity to their target are called as hits. Polyspecific binders are filtered out by order statistics-based hit calling, in which the 55 other targets are considered as background signal.

Retrospective ranking shows that filtering to the top 100 designs per target roughly doubles hit rates, and the most successful targeted epitopes reach 100% success among their top ten. Given the scale of our binding experiment, we deliberately do not filter this aggressively. While successful binders to some targets can be achieved with few designs, coverage across a large set of targets and epitopes remains challenging. For this reason we choose to saturate the experimental funnel with diverse designs, maximizing coverage and useful positive and negative binding signal.

mBER-2 expands the binder format design space beyond VHH, to include unscaffolded minibinders and a set of 50 scaffolded peptide formats of 10-45 amino acids in length, which we term "microbinders." Design of binders in this size range is challenging for unscaffolded methods due to the paucity of stable structures at this length. The additional formats enable broader epitope coverage, and also offer potential advantages in clinical development, including chemical synthesis, tissue permeability, and reduced immunogenicity.

mBER-2 implements a novel refinement protocol that efficiently samples the sequence space of a known binder structure. This structure-aware redesign can consistently generate variants with as many as 20 mutations that improve both affinity and specificity. Such large leaps in the sequence-fitness landscape are beyond the ability of structure-unaware predictions from protein language models. This method allows for maturation and diversification of binders to broaden the sampled phenotypic space.

The ability to reliably design novel binders against a wide range of epitopes unlocks a broad therapeutic target space. We are particularly interested in the development of tissue-targeted medicines, in the

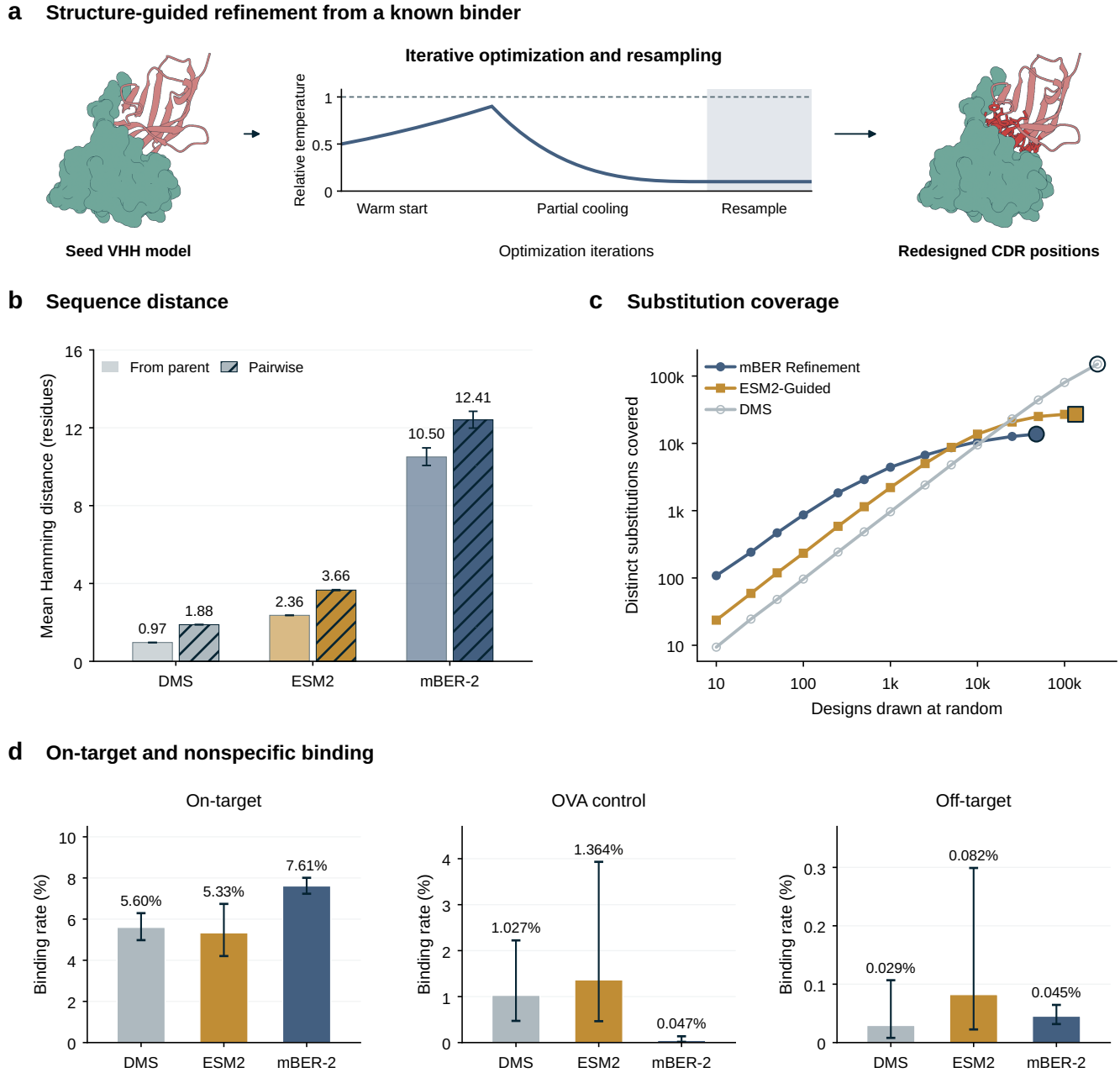


Figure 3. Sequence exploration and binding in refinement. a. Schematic warm start, partial cooling, and resampling used in mBER refinement protocol. In refinement, mBER-2 is seeded with a designed bound complex that has already been validated as an experimental binder. An optimization process is initialized from that binder sequence and constrained to resampling mutations under a fixed structural dock with the target. **b.** Mean parent-wise and pairwise sequence distances across 97 matched parents for each method (291 variant sets), with 95% intervals from 10,000 parent-level bootstrap resamples. **c.** Mean number of distinct position–residue substitutions across random samples of designs from 98 parents, with substitutions counted separately for each parent (Supplementary Methods S1.3). Larger final markers show all designs generated by each method. **d.** Phage-panning binding rates per detected protein–antigen pair, with Wilson's 95% intervals [15].

form of binders to specific epitopes that have the potential to enable delivery of therapeutic payloads to specific tissues, increasing efficacy and reducing toxicity. The success of mBER-2 against the targets in this work motivates a much larger design campaign, targeting the space of all exposed epitopes in the human proteome—a project we refer to as the EpiTome.

mBER-2's state of the art epitope coverage and hit rates suggest that an EpiTome-scale design campaign is likely to generate successful, specific binders to a majority of targets; however, screening binders to such a target set in an *in vitro* assay like the one presented in this work is infeasible. There are over 1,000 unique cell-surface targets of interest, many of which cannot be easily or faithfully displayed *in vitro* for binding assessment.

Manifold's proprietary molecular barcoding technologies enable us to move past this barrier by testing enormous libraries of designed binders directly *in vivo*. From this data, we gather multiplexed measurements of pharmacokinetics and biodistribution. We aim to test an EpiTome-scale library *in vivo*, linking design intent to biodistribution profiles, thus generating a map of binding-induced tissue targeting phenotypes at epitope-level resolution.

Combined with massively multiplexed *in vivo* measurement, mBER-2 allows us to generate a collection of potent tissue-targeting medicines whose behavior *in vivo* reveals unprecedented insights into the biology of target-mediated delivery.

Contributors

Listed alphabetically: John Avera, Marshall Case, Brent Dorr, Gleb Kuznetsov, Roasa Mehmood, Michael Nichols, Pierce Ogden, Sarah Post, Maggie Prescott, Supriya Ravichandran, Robert Scheidegger, Ernst Schmid, Ioana Stoica, Erik Swanson, Daniel Tresnak

References

[1] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.

[2] J. L. Watson *et al.*, "*De novo* design of protein structure and function with RFdiffusion," *Nature*, vol. 620, pp. 1089–1100, 2023, doi: 10.1038/s41586-023-06415-8.

[3] J. Dauparas *et al.*, "Robust deep learning–based protein sequence design using ProteinMPNN," *Science*, vol. 378, pp. 49–56, 2022, doi: 10.1126/science.add2187.

[4] E. Swanson, M. Nichols, S. Ravichandran, and P. Ogden, "mBER: Controllable *de novo* antibody design with million-scale experimental screening," *bioRxiv (preprint)*, 2025, doi: 10.1101/2025.09.26.678877.

[5] L. S. Mille-Fragoso *et al.*, "Efficient generation of epitope-targeted *de novo* antibodies with Germinal," *bioRxiv (preprint)*, 2025, doi: 10.1101/2025.09.19.677421.

[6] Chai Discovery Team *et al.*, "Zero-shot antibody design in a 24-well plate," *bioRxiv (preprint)*, 2025, doi: 10.1101/2025.07.05.663018.

[7] Nabla Bio, "JAM-2: Fully computational design of drug-like antibodies with high success rates," Nabla Bio, Technical report, 2025. Available: `https://nabla-public.s3.us-east-1.amazonaws.com/2025_Nabla_JAM2.pdf`

[8] H. Stark *et al.*, "BoltzGen: Toward Universal Binder Design," *bioRxiv (preprint)*, 2025, doi: 10.1101/2025.11.20.689494.

[9] Chai Discovery Team *et al.*, "Drug-like antibody design against challenging targets with atomic precision," *bioRxiv (preprint)*, 2025, doi: 10.1101/2025.11.29.691346.

[10] T. Uçar *et al.*, "BoltzProt-1: Towards Efficient *De Novo* Binder Design with Good Developability," Boltz, Technical report, 2026. Available: `https://boltz.bio/boltzprot1-technical-report.pdf`

[11] Z. Lin *et al.*, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, pp. 1123–1130, 2023, doi: 10.1126/science.ade2574.

[12] M. Pacesa *et al.*, "One-shot design of functional protein binders with BindCraft," *Nature*, vol. 646, pp. 483–492, 2025, doi: 10.1038/s41586-025-09429-6.

[13] L. T. Chen *et al.*, "Target sequence-conditioned design of peptide binders using masked language modeling," *Nature Biotechnology*, vol. 44, no. 6, pp. 1002–1010, 2026, doi: 10.1038/s41587-025-02761-2.

[14] M. Muttenthaler, G. F. King, D. J. Adams, and P. F. Alewood, "Trends in peptide drug discovery," *Nature Reviews Drug Discovery*, vol. 20, no. 4, pp. 309–325, 2021, doi: 10.1038/s41573-020-00135-8.

[15] E. B. Wilson, "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, vol. 22, pp. 209–212, 1927, doi: 10.1080/01621459.1927.10502953.

[16] G. Labesse, N. Colloc'h, J. Pothier, and J.-P. Mornon, "P-SEA: a new efficient assignment of secondary structure from Cα trace of proteins," *Computer Applications in the Biosciences*, vol. 13, no. 3, pp. 291–295, 1997, doi: 10.1093/bioinformatics/13.3.291.

SUPPLEMENTARY INFORMATION

S1. Computational Methods

S1.1 Epitope Selection and Clustering

Target surface sites were represented by local structural crops. For hit-producing trajectories with a resolved structure mapping, contact residues were identified as target heavy atoms within 5.0 Å of any binder heavy atom.

Within a target, contact sets A and B are connected when their Jaccard similarity $|A \cap B|/|A \cup B|$ is at least 0.5. Connected components define clusters. This rule can join neighboring footprints through intermediate members. Pooled clustering of designed structures for validated binders gives 353 clusters, including at least two on 51/56 targets. Independently clustered format totals are 284, 99, and 74, distinct from the pooled partition's format memberships of 276, 92, and 72.

Pooled rates divide hits by high-signal designs. Figure 1a instead uses unweighted means of target-level VHH rates. Top-100 selection ranks high-signal designs by mBER-2's design evaluation model within each format–target group.

S1.2 Scaled *de novo* binder design with mBER-2

mBER-2 follows a similar design approach to mBER-open, in which a core component of the generative process involves repeatedly folding a partially-designed binder sequence against a target epitope and backpropagating a set of confidence, sequence, and structure-based losses back to the sequence inputs. mBER-2 introduces changes throughout this generative process, including upstream template preparation, folding trajectories, and downstream sequence redesign and evaluation. We are intentionally light on details on the mBER-2 generative process and unlike mBER-open, we do not currently have plans to open-source it. A high-level schematic illustration of the mBER-2 generative process is shown in Figure S1.

S1.3 mBER-2 refinement

mBER refinement initializes with the complex structure of an mBER hit with its cognate epitope and an initial PSSM built across all CDR positions. Multiple stages of sequence sampling, refolding, loss calculation, and backpropagation to update the PSSM are conducted to optimize sampled sequences, followed by greedy sampling from the PSSM to generate final candidate sequences. mBER refinement skews the input PSSM with the parent sequence and forces resampling of interface contacts and CDR residues in the context of the input binding interface, such that loss terms account for shifts away from the predetermined interface.

For Figure 3c, each substitution is defined by its parent, sequence position, and substituted amino acid. The same position–residue change in several variants of one parent counts once; the analogous change in a different parent is counted separately. Random draws pool designs from 98 parents. The curve reports the mean number of distinct substitutions across repeated draws of the same size.

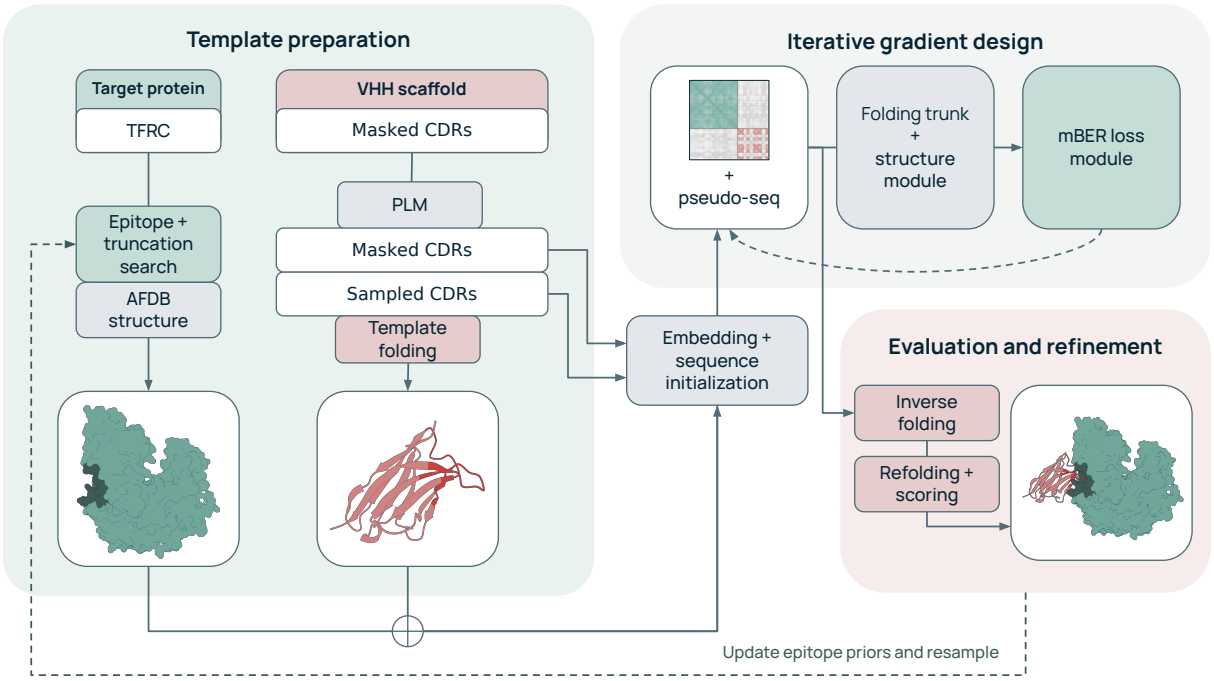


Figure S1. High-level mBER-2 design workflow. Template preparation and epitope selection provide target and binder inputs for iterative gradient design. Sequence and structure modules optimize candidate binders, followed by inverse folding, refolding, scoring, and refinement of epitope priors. The TFRC example illustrates an epitope patch definition; the workflow applies across the target panel.

S1.4 Calling Hits from Phage Data

We call hits on a per-sequence basis using order statistics over the signal measured to that sequence in every well of our assay. In order to ensure that signal across different wells is comparable, we normalize the counts in a well by the total counts in that well. Specifically, for a given sequence i and well j, we compute a normalized signal $\tilde{n}(i,j)$ in counts per million from the raw counts $n(i,j)$ as

$$\tilde{n}(i,j) = 10^6\, n(i,j) / \sum_{i'} n(i',j).$$

For a target t we then compute two order statistics over the wells of sequence i, a signal term taken inside the set K of replicate wells for target t,

$$S(i,t) = k\text{-th largest } \tilde{n}(i,j), \quad j \in K,$$

and a background term taken outside it,

$$B(i,t) = q\text{-th largest } \tilde{n}(i,j), \quad j \notin K,$$

from which the hit-calling ratio is

$$r(i,t) = \frac{S(i,t)}{\max(B(i,t), b_0)}.$$

The floor b_0 is the signal that a single read produces in a well of median depth, $b_0 = 10^6/(median\ well\ total) \approx 0.11$, and it prevents a sequence with no reads outside K from taking an arbitrarily large ratio. We call sequence i a hit to target t if $r(i,t) \geq \tau$.

In this experiment, targets were arrayed in an 8 x 8 format on a 96-well plate, with 56 of 64 positions corresponding to true design targets, 1 of 64 positions corresponding to a null, untargeted antigen, and 7 of 64 positions left antigen-free. This layout was replicated across 4 plates. All 256 positions across the 4 plates underwent the same experimental protocol and sequencing.

To establish an empirical false positive rate for our hit calling method, we score every sequence against a set of decoy groups, including seven groups of four antigen-free wells, and a set of 40 groups composed of unique target wells sampled from the 4 different replicate plates. This second set of groups establishes a randomly-shuffled target null, while the first establishes an antigen-free plate-binding background. For a target t, the expected number of false calls to that target is then the direct count

$$FP(t) = \frac{\text{calls that t's own designs make on the decoy groups}}{\text{number of decoy groups}}.$$

which we evaluate with whichever of the two decoy sets gives the larger value.

We fix k = 3 and q = 20, to allow for hit calling in cases where one of four replicates failed, while reducing sensitivity to signal in a small number of other target wells. We then choose τ for each library to be the threshold that admits the most calls while keeping the mean of FP(t) over the panel below one. Details on per-library hit calling metrics can be found in Table S3.

We quote hit rates over the high-signal population, meaning *de novo* designs whose intended target is on the panel and which have at least one read in at least three of the four wells of that target, the minimum that k = 3 requires. Of the 474,440, 420,575 and 355,705 *de novo* on-panel designs in the three libraries, 190,767 (40%), 161,263 (38%) and 232,420 (65%) are high-signal on this criterion, giving overall hit rates of 8.6%, 4.64% and 0.93%. An on-design hit is a sequence called binding to its intended design target; an off-design hit is a call to another antigen in the panel. In the all-against-all heatmaps and target-level enrichment tests, each design remains in the high-signal cohort defined at its intended target. For an evaluated antigen, the on-design rate uses designs intended for that antigen and the off-design rate uses designs intended for the other panel antigens.

Primary thresholds are 2.5, 3.0, and 2.0 for VHHs, minibinders, and microbinders. Estimated false-discovery fractions, computed as summed expected false calls divided by observed hits, are 0.26%, 0.50%, and 1.43%. A stricter rule limiting every target's expected false calls below one gives thresholds 4.0, 5.0, and 2.5, yielding 15,100, 6,465, and 2,065 hits.

S1.5 Epitope targeting

High-signal designs were grouped by target, binder format, and their complete design-epitope specification. Within each target and format, the group with the highest observed hit fraction was selected. The selected group's top ten designs were ranked by mBER-2's evaluation model; all designs were retained when fewer than ten were available. Means include all 56 targets, including targets with zero hits. Group selection starts from the full high-signal cohort, independently of the global top-100 analysis.

These retrospective groups were selected using the observed hits, and six contained fewer than ten high-signal designs; these rates therefore describe in-sample enrichment rather than prospective performance. Nine selected groups yielded 10/10 hits; three additional groups had perfect rates with only one or three designs.

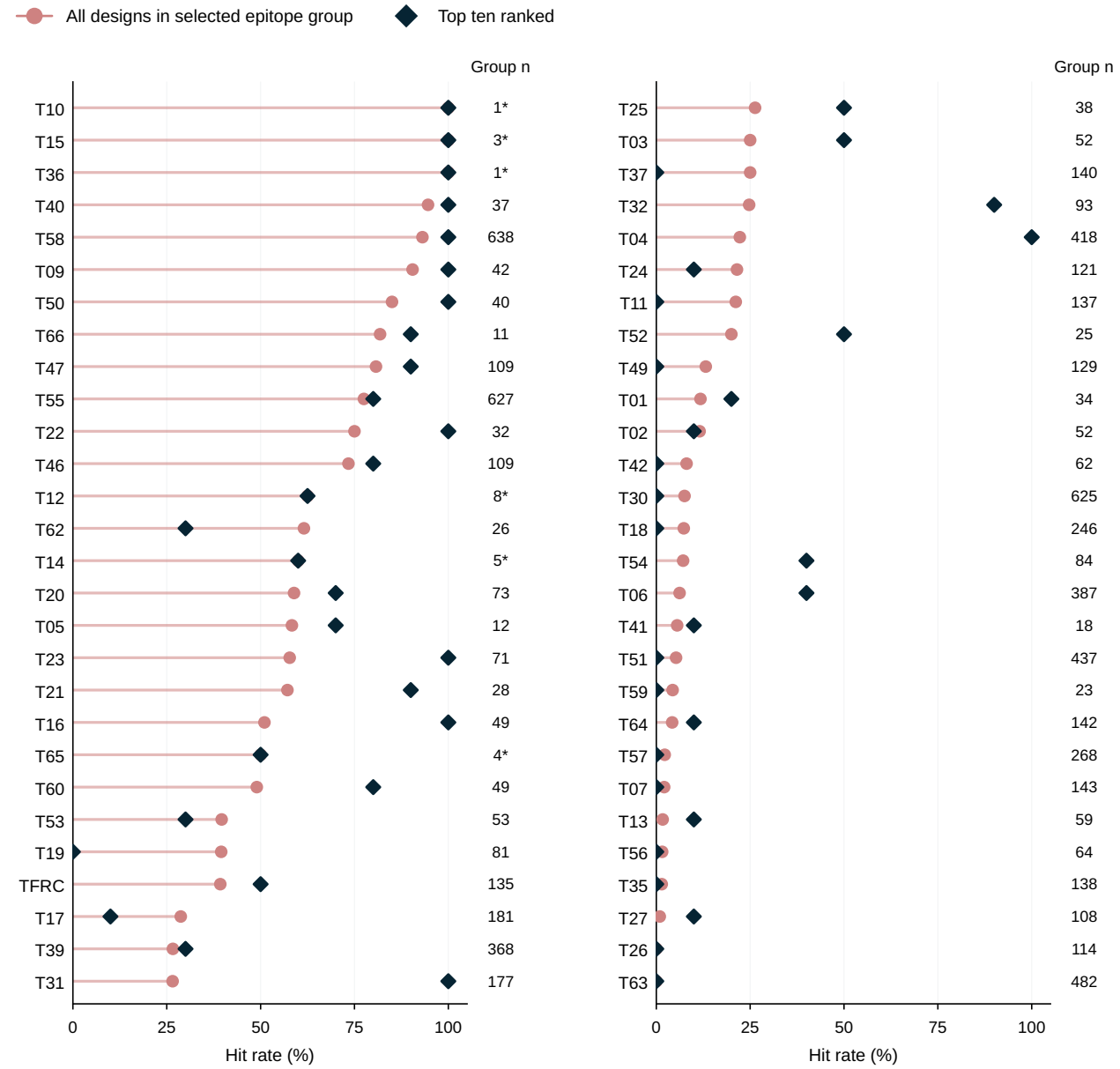


Figure S2. Retrospective enrichment within VHH design-epitope groups. Each row is one target, ordered by its selected-group hit rate. Points show the observed fraction of hits across the entire selected group and among its top ten ranked designs. Labels give selected-group size; an asterisk marks one of six groups with fewer than ten designs. Means across 56 targets are 37.6% and 45.9%, respectively. Both ranges are 0–100%. Selection and evaluation use the same hits, with no minimum group size, so the plot shows retrospective enrichment rather than held-out performance.

S1.6 Target-level binding enrichment

For each evaluated antigen and binder format, a 2×2 table compared hit and non-hit outcomes for high-signal designs intended for that antigen with outcomes for high-signal designs intended for all other panel antigens. A one-sided Fisher exact test assessed enrichment of on-design hits. Nominal significance used $P < 0.05$; Bonferroni significance used $P < 0.05/56$, separately within each format and the pooled analysis. Table S9 summarizes the numbers of significant targets. Target coverage defined by at least one hit is a different quantity from statistically significant enrichment.

S1.7 Hit rates by epitope secondary structure

Secondary structure was assigned from the full target-chain Cα trace using the P-SEA implementation in Biotite [16]. Target contact residues were defined by any heavy atom within 5 Å of the binder in each design trajectory's predicted complex. Epitopes were classified as helix, β-strand, or loop when at least 60% of contact residues had that assignment, and mixed otherwise. These trajectory-level annotations covered all 584,450 high-signal *de novo* designs in the primary analysis. Hit rates divide the number of hits by all high-signal designs in each format–epitope class, using the same hit calls as the main figures.

Minibinder hit rates were 5.74% at helix-dominated epitopes (1,671/29,110) and 0.25% at β-strand-dominated epitopes (19/7,731); corresponding VHH rates were 8.13% (1,801/22,149) and 5.85% (685/11,718). Target-stratified Mantel–Haenszel odds ratios compared each class with loop-dominated epitopes. For β-strand epitopes, odds ratios were 0.039 for minibinders (95% interval 0.025–0.061), 1.29 for VHHs (1.14–1.45), and 0.32 for microbinders (0.26–0.41). The lower minibinder success at β-strand epitopes persisted after target adjustment.

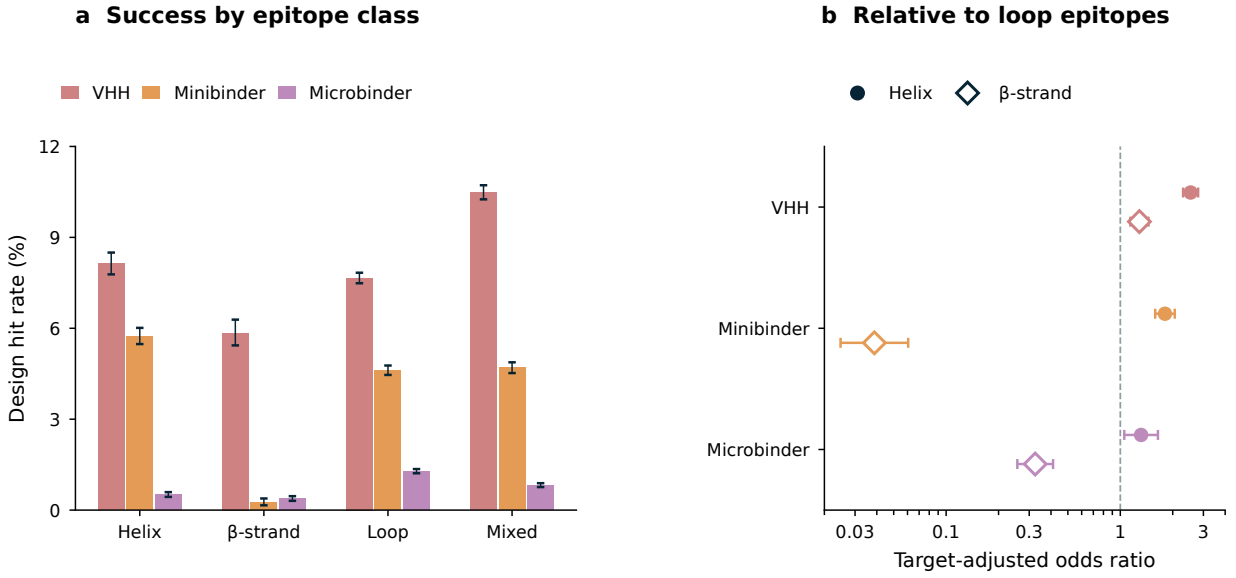


Figure S3. Design hit rates by predicted epitope secondary structure. a. Hit rates for the 584,450 high-signal designs, grouped by binder format and contact-based epitope class. Error bars are Wilson 95% intervals. **b.** Target-stratified Mantel–Haenszel odds ratios for helix- (filled circles) and β-strand-dominated epitopes (open diamonds) relative to loop-dominated epitopes, with 95% intervals. The dashed line marks equal odds. Classes are assigned per design trajectory; all designed epitopes contribute, including those without hits.

S1.8 Comparison with mBER-open

We compared VHH designs from the mBER-1 campaign underlying mBER-open [4] with the mBER-2 campaign. Reconstructing the original mBER-open paper cohorts reproduced their 331,578 tested

designs and 2,207 published hits. We then reapplied the hit-calling framework in S1.4 to their raw counts: CPM normalization, the third-highest intended-target signal, the 20th-highest signal outside those wells, and a one-read background floor. Thresholds were calibrated separately for each campaign to keep the mean expected false-positive count below one per target, yielding thresholds of 2.0 and 1.0 for the two mBER-1 campaign datasets; the mBER-2 campaign retains its primary VHH threshold of 2.5.

The intersection of the experimentally represented intended-target panels contains 15 human targets; all are retained. Among designs with positive raw counts in at least three of four intended-target wells, mBER-open yielded 119/9,265 hits (1.28%), and mBER-2 yielded 2,032/57,656 (3.52%). Retrospectively selecting the 100 highest-ranked designs per target using each campaign's design evaluation scores yielded 65/1,500 (4.33%) and 291/1,500 (19.40%), respectively (Figure S7; Table S10).

Under the primary hit calls used in this paper, retrospective selection of the top ten VHH designs per target yields 96/560 hits (17.1%) and covers 23/56 targets; top-100 selection yields 941/5,600 hits (16.8%) and covers 40/56 targets (Table S11).

S2. Experimental Methods

S2.1 Protein Sourcing

All proteins were sourced from Sino biologicals or Kactus Bio.

S2.2 Library Cloning

Libraries were cloned into VHH, or minibinder frameworks using standard Gibson Assembly protocols. Library inserts and the vector were combined at optimized ratios with assembly master mix and incubated under standard conditions. The assembled fragments were electroporated into bacterial cells and transformed colonies were analyzed for library coverage. Once adequate coverage was confirmed, libraries were prepared and stored at -80°C until further use.

S2.3 Phage Production

Bacterial cells for phage production were grown in supplemented media at 37°C with shaking until reaching optimal density. Cells were infected with helper phage and incubated for 1 hour. The infected cells were then resuspended in media containing antibiotics and incubated overnight with gentle shaking. Phage were PEG precipitated and purified. Once purified, phage were titered in bacterial cells, and stored at 4°C until use. Phage libraries were sequenced using next-generation sequencing (NGS) to ensure presence of all sequences.

S2.4 Phage Display for Binding Evaluation

Proteins were coated onto appropriate capture plates. Plates were washed and incubated with a blocking solution for 1.5 hours at room temperature. Phage libraries at appropriate titers were added and incubated for 2 hours. After washing to remove unbound phage, bound phage were eluted enzymatically and used to infect bacterial cells. Infected cells were selected overnight in antibiotic-containing media with shaking. Samples were prepared for next-generation sequencing using appropriate PCR primers to amplify binder sequences, then uniquely indexed and sequenced.

S2.5 Kinetic Binding Measurement by Bio-Layer Interferometry (BLI)

Binding was determined using an Octet RH96 (Sartorius). Fc-tagged VHHs, minibinders, and microbinders were expressed in Expi293 pro cells from transiently transfected plasmids (Twist Bioscience). Molecules were immobilized on 2nd generation Anti-Human-IgG Capture (AHC2) tips directly from the expression supernatant using a 1:50 dilution of supernatant into Octet buffer (1X PBS 0.1% Casein 0.05% Tween20 pH 7.4). This buffer was used for all subsequent steps.

Biosensors were first hydrated for 10 minutes, followed by a 300 second baseline step. Molecules from diluted supernatant were immobilized on the tips for 120 seconds. The sensors were baselined again for 60 seconds, then challenged with the respective target for a 120 second association. The antigens were then dissociated for 120 seconds, followed by a 120 second tip regeneration cycle consisting of 6 cycles of 10 seconds in regeneration buffer (10mM glycine pH 1.5) followed by 10 seconds in neutralization buffer (Octet buffer). Tips were regenerated up to 12 times. All steps were performed at 25°C with 1000rpm shaking, and sampling rate was set to 0.6Hz. Target concentrations were 500nM, 100nM, and 20nM. References included both blank association well (Octet buffer only) for each molecule and a no-ligand control to assess whether targets presented background association with the tips.

The 430-measurement overview shows a selected cohort of high-confidence phage hits on 25 antigens with clear binding signal represented (Figure S4). Reported affinities required clear signal, at least two fitted concentrations, $0 < K_D < 500$ nM, and $SD(K_D)/K_D < 0.3$. The selected refinement cohort was matched to its clonal BLI measurements, with the parent sequence measured separately.

S2.6 Developability analysis

Variants were matched across concentration, SEC, BVP-ELISA, and clonal BLI measurements. Concentration is the median of three Stunner readings. SEC reports peak-area fractions. BVP values are means of two technical replicates. Mutation counts are Hamming distances between each measured variant and its parent sequence.

Expression classes use a cohort-relative robust z score based on the median and $1.4826\times$ the median absolute deviation: in range above -3, marginal between -5 and -3, and out of range below -5. SEC HMW classes use less than 5%, 5–10%, and at least 10%. BVP is in range at no more than $1.25\times$ the Fc-NB2 reference, marginal between that level and the parent, and out of range above the parent. The reference means are 0.124 and 0.291 OD units for Fc-NB2 and the parent. The absence of parent controls for expression and SEC limits before/after claims.

S3. Library populations

Table S1. Ordered, analyzed, and high-signal populations. Raw and top-100 rates use high-signal designs. Top-100 selection ranks designs by mBER-2's design evaluation model within each format–target group; fewer than 100 are retained when fewer are high-signal. Ordered counts include arms excluded from the primary analysis.

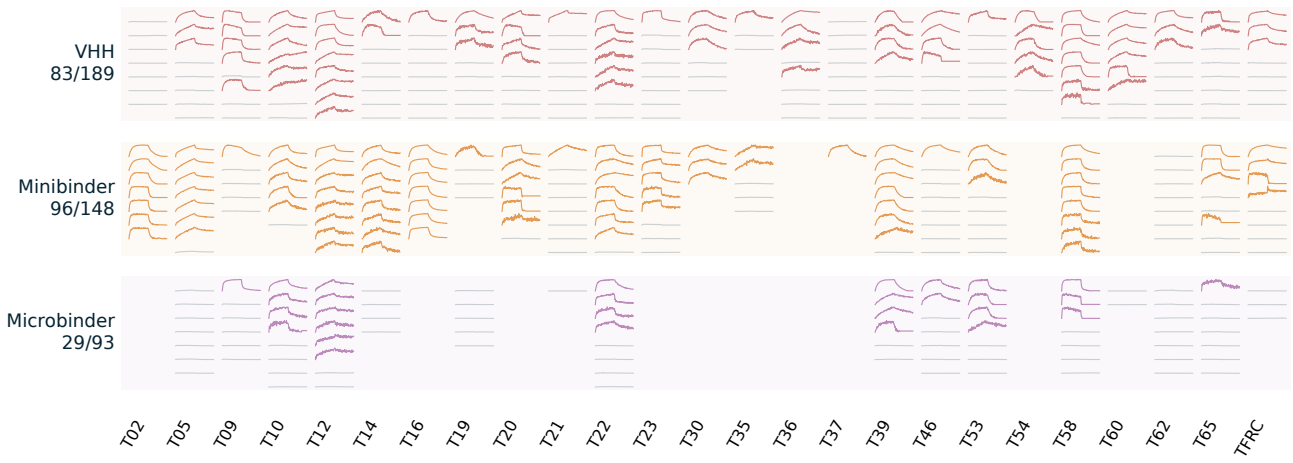
Format	Ordered	Primary	High-signal	Hits	Raw (%)	Top-100 (%)	Targets
VHH	476,622	474,440	190,767	16,387	8.59	16.80	54/56
Minibinder	462,470	420,575	161,263	7,475	4.64	8.62	46/56
Microbinder	359,242	355,705	232,420	2,164	0.93	3.55	42/56

S4. Supplementary validation

Following the phage binding experiment, we selected a subset of mBER-2 binders for validation in a high-throughput BLI assay (Supplementary Methods S2.5). 430 high-confidence binders against 25 antigens, spanning the 3 format classes (VHH, minibinders, microbinders) were assessed for binding affinity. This selected cohort contained 208 clear-signal measurements and 74 affinity fits meeting the criteria in Supplementary Methods S2.5: 29 VHHs (8.2–362.1 nM), 41 minibinders (8.4–347.0 nM), and 4 microbinders (113.9–336.8 nM). Representative binding curves for each class are shown in Figure S4.

Developability measurements cover 23 variants of one refinement parent. Nineteen met all three expression, SEC, and BVP-ELISA classification rules. The six mBER variants have a median of 10.5 substitutions, no reported SEC high-molecular-weight peak, and satisfy all three rules (Figures S5 and S6; Table S2). All 23 variants also yielded clonal BLI affinities: 5.9–32.9 nM for DMS (five variants), 8.6–50.8 nM for ESM2-guided designs (12 variants), and 21.2–48.4 nM for mBER (six variants). The best mBER variant has 11 substitutions and a K_D of 21.2 nM; the sequence-matched parent was measured separately at 38.4 nM. Parent controls were absent for expression and SEC, and the parent affinity was measured in a separate acquisition.

a BLI traces for high-confidence phage hits



b Representative binding curves

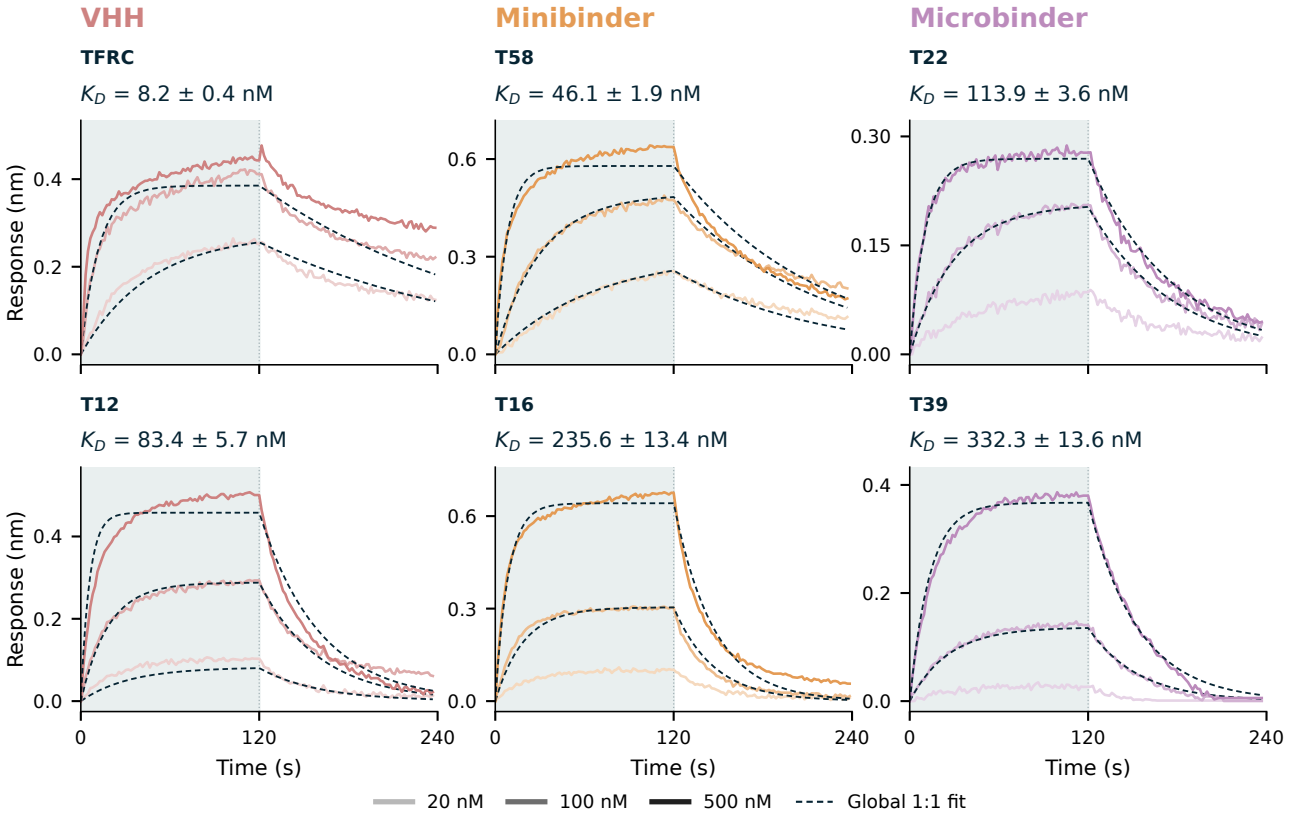


Figure S4. BLI sensorgrams and selected affinity fits. a. BLI measurements for 430 high-confidence phage hits on 25 antigens with at least one clear-signal measurement. Target identities are masked except TFRC. Each mini-trace is one binder–antigen pair at 500 nM analyte, shown over 0–240 s. Within each antigen and format, traces are ordered by maximum response. Colored traces have clear binding signal: 83/189 VHHs, 96/148 minibinders, and 29/93 microbinders (208/430 combined). Clear traces are divided by their own positive maximum response and clipped to 0–1 for display; gray, unclear traces share a fixed 0.8 nm scale. **b.** Six selected binding curves across three distinct binder formats, with their fitted K_D values and fit standard deviations. Points are unnormalized responses in nm; lines are stored global fits to the included concentrations. Each panel's vertical limits follow its maximum observed or fitted response, with 12% headroom.

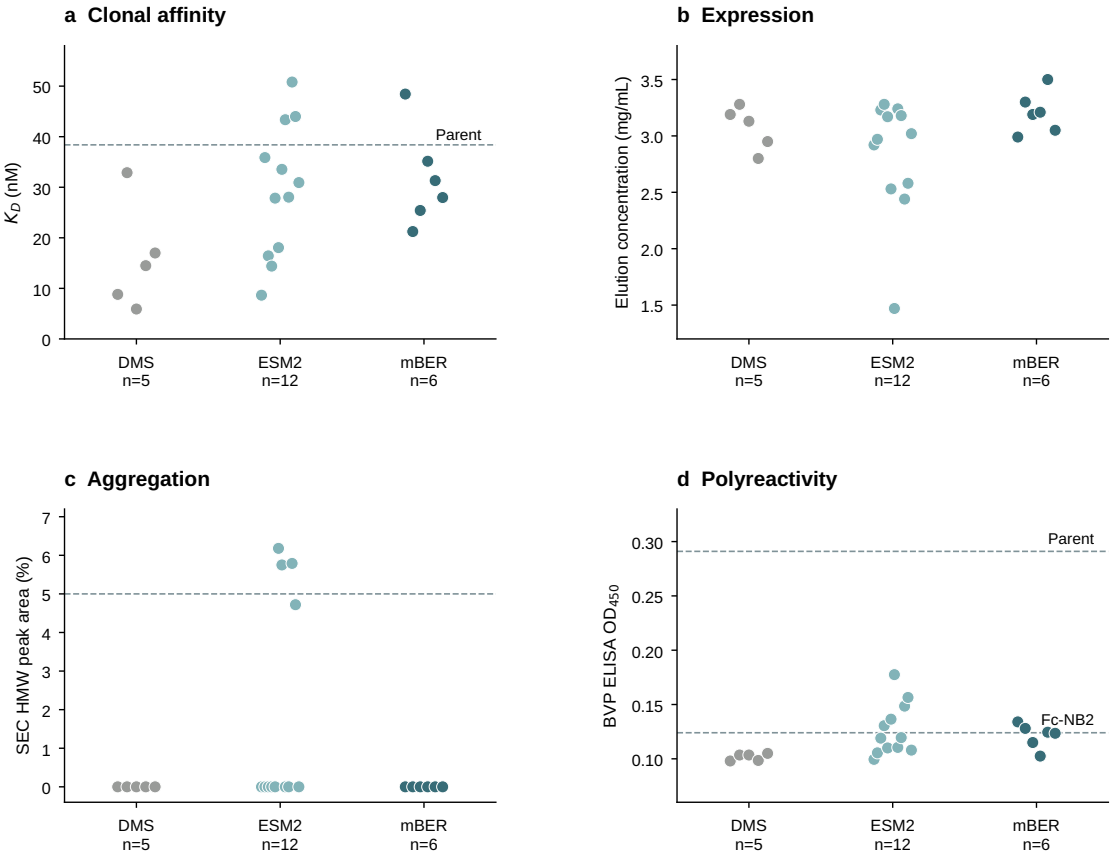


Figure S5. Clonal binding affinity and developability for selected variants. Each point is one of 23 variants of a parental mBER binder sequence to one target. **a.** Clonal BLI affinities. Corresponding sensorgrams are shown in Figure S6. The dashed line marks the separately measured parent affinity (38.4 nM). **b.** Concentration, the median of three instrument readings. **c.** SEC HMW peak-area percentage; the dashed line marks 5% HMW. **d.** BVP OD, the mean of two technical replicates. Dashed lines mark the Fc-NB2 and parent BVP means.

Table S2. Summary of the selected refinement cohort. Concentration is in mg/mL. "All in range" applies the three descriptive rules in Methods.

Source	Variants	Median edits	Conc.	HMW > 0	BVP OD	All in range
DMS	5	1.0	3.13	0	0.103	5
ESM2	12	2.0	3.00	4	0.119	8
mBER	6	10.5	3.20	0	0.124	6

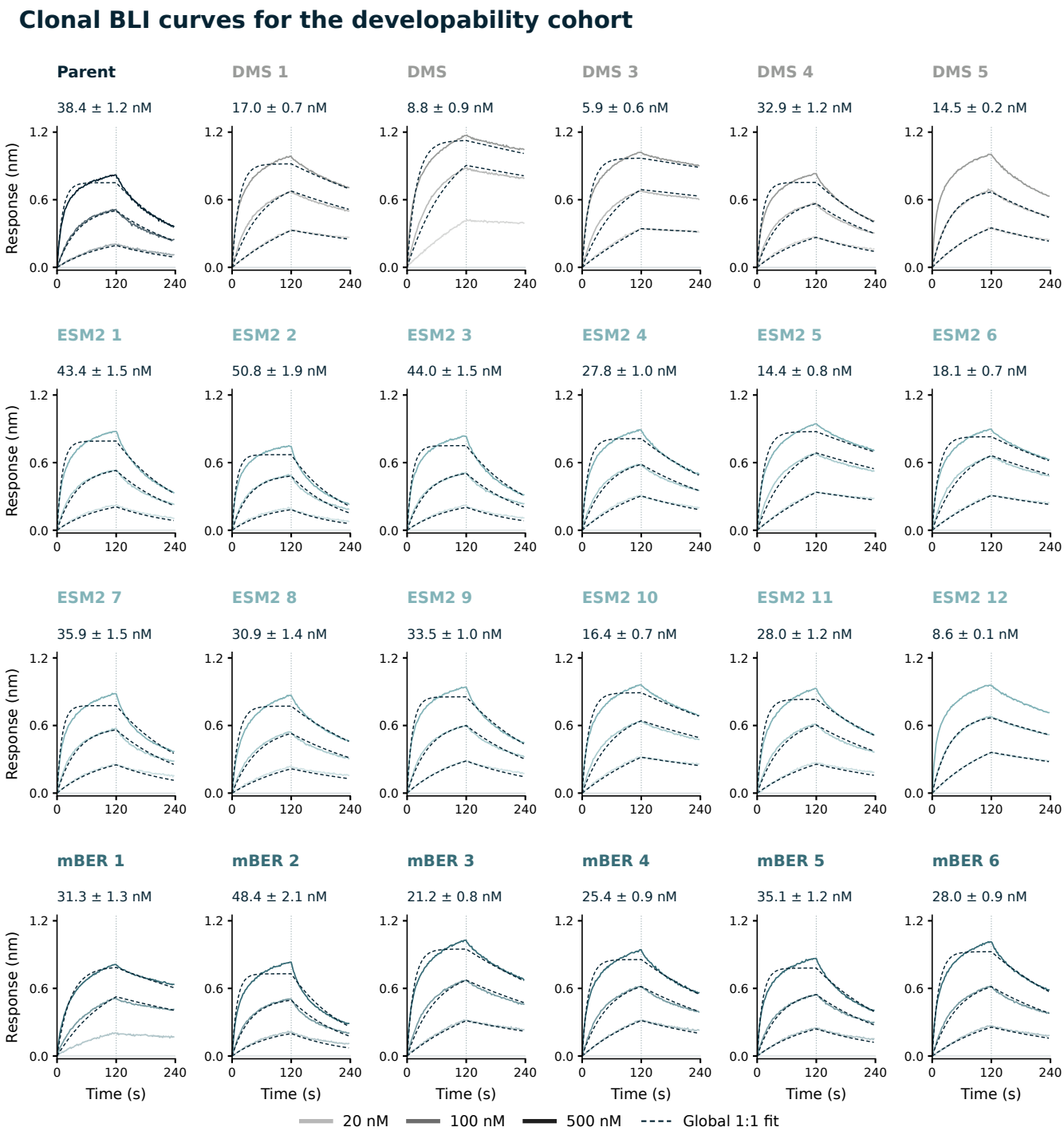


Figure S6. Clonal BLI sensorgrams underlying Figure S5a. Curves are shown for all 23 variants in the developability cohort and their parent: five DMS variants, 12 ESM2-guided variants, and six mBER variants. Color intensity indicates analyte concentration (20, 100, or 500 nM); dashed lines show global 1:1 fits. Labels report K_D ± fit standard deviation. The parent was measured separately. All panels use the same response scale, and variant labels are anonymized.

S5. Hit thresholds and coverage definitions

Table S3. Primary and strict operating points. Primary calibration constrains the mean estimated false positive calls (FP) across targets; strict calibration constrains the maximum. Estimated FDR is the sum of per-target FP estimates divided by the number of calls.

Format	Rule	τ	Hits	Targets	Mean FP	Max FP	Est. FDR (%)
VHH	primary	2.5	16,387	54/56	0.749	2.143	0.256
VHH	strict	4.0	15,100	54/56	0.196	0.857	0.073
Minibinder	primary	3.0	7,475	46/56	0.661	2.543	0.495
Minibinder	strict	5.0	6,465	42/56	0.216	0.857	0.187
Microbinder	primary	2.0	2,164	42/56	0.552	1.571	1.428
Microbinder	strict	2.5	2,065	37/56	0.231	0.714	0.625

Table S4. Epitope cluster membership across 3 binding formats.

Format	Pooled clusters containing format	Exclusive pooled clusters
VHH	276	217
Minibinder	92	40
Microbinder	72	34

S6. Refinement denominators

Table S5. Sequence-distance summaries from Figure 3b.

Route	Matched parents	Mean distance from parent	Mean pairwise distance
mBER refinement	97	10.50	12.41
ESM2-guided	97	2.36	3.66
DMS	97	0.97	1.88

Table S6. Counts underlying Figure 3d. Denominators are detected protein–antigen pairs. Confidence intervals are Wilson 95% intervals without within-seed dependence correction.

Route	Antigen class	Seeds	Fits	Pairs	Rate (%)	95% interval
mBER refinement	on-target	96	1,371	18,013	7.611	7.233–8.008
mBER refinement	off-target	96	30	66,399	0.045	0.032–0.065
mBER refinement	OVA control	96	3	6,449	0.047	0.016–0.137
ESM2-guided	on-target	92	65	1,219	5.332	4.205–6.740
ESM2-guided	off-target	92	2	2,435	0.082	0.022–0.299
ESM2-guided	OVA control	92	3	220	1.364	0.465–3.932
DMS	on-target	93	265	4,734	5.598	4.978–6.289
DMS	off-target	93	2	6,827	0.029	0.008–0.107
DMS	OVA control	93	6	584	1.027	0.472–2.223

S7. Anonymized target list

Table S7. On-design hits and pooled predicted contact-set clusters for all 56 targets. Class labels match the main figure. Target codes are consistent throughout the report; TFRC is the only named target.

Target	Class	VHH hits	Mini hits	Micro hits	Clusters
T22	GPCR	658	868	410	3
T55	GPCR	2873	504	161	3

Target	Class	VHH hits	Mini hits	Micro hits	Clusters
T49	GPCR	41	2	1	5
T53	GPCR	72	135	460	7
T57	Receptor kinase	7	38	18	3
T21	Receptor kinase	56	143	1	4
T36	Receptor kinase	275	2	2	7
T65	Receptor kinase	27	574	33	3
T64	Receptor kinase	6	11	0	2
T47	Receptor kinase	445	337	21	11
T62	Receptor kinase	137	76	48	11
T13	Receptor kinase	4	0	4	4
T59	Receptor kinase	2	5	1	3
T25	Receptor kinase	28	3	0	5
T19	Receptor kinase	152	58	19	10
T37	Receptor kinase	42	2	0	4
T52	Cadherin	31	480	2	3
T66	Cadherin	255	27	7	14
T24	Cadherin	95	10	9	8
T09	Ig-CAM / adhesion	1595	194	176	4
T46	Ig-CAM / adhesion	622	320	113	7
T02	Ig-CAM / adhesion	40	216	0	9
T50	Ig-CAM / adhesion	142	186	61	21
T39	Ig-CAM / adhesion	371	70	140	1
T41	Ig-CAM / adhesion	2	0	1	3
T31	Ig-CAM / adhesion	112	112	5	13
T20	Ig-CAM / adhesion	173	174	0	6
T35	Ig-CAM / adhesion	5	18	1	6
T12	Ig-CAM / adhesion	707	343	65	6
T60	Ig-CAM / adhesion	321	0	6	7
T17	Ectoenzyme	133	16	2	7
T56	Ectoenzyme	1	0	0	1
T54	Ectoenzyme	26	1	1	3
T51	Ectoenzyme	35	1	2	4
T18	Ectoenzyme	19	0	1	4
T16	Ectoenzyme	511	203	2	6
T58	Transport & uptake	4177	1087	209	10
T42	Transport & uptake	41	0	0	3
T06	Transport & uptake	25	1	0	1
T07	Transport & uptake	11	1	7	2
T11	Transport & uptake	34	1	1	5
T30	Transport & uptake	114	5	0	5
T63	Transport & uptake	0	1	0	1
TFRC	Transport & uptake	129	234	4	11
T01	Other receptor	34	0	0	4
T32	Other receptor	102	293	4	3
T04	Other receptor	404	161	27	19
T27	Non-receptor membrane	1	1	0	2
T15	Non-receptor membrane	146	0	0	4
T03	Non-receptor membrane	66	0	29	9
T10	Non-receptor membrane	181	175	16	17
T40	Non-receptor membrane	497	117	72	19
T23	Non-receptor membrane	134	54	1	12
T05	Non-receptor membrane	91	99	13	5
T26	Non-receptor membrane	0	0	0	0
T14	Non-receptor membrane	179	116	8	3

S8. Microbinder scaffold results

Table S8. Results for the 50 assayed microbinder scaffolds, ordered by overall hit rate. Scaffold names match Figure 2b.

Scaffold	Length (aa)	High-signal	Hits	Raw (%)	Top-100 (%)
gEHE-06	35	9,642	399	4.14	16.35
Ubiquitin 45	45	1,087	38	3.50	0.00
NC-HEE-D1	27	10,379	319	3.07	7.75
Villin HP-35	35	2,854	78	2.73	0.00

Scaffold	Length (aa)	High-signal	Hits	Raw (%)	Top-100 (%)
gEEHE-02	37	11,579	278	2.40	3.74
FSD-1	28	2,348	53	2.26	0.00
Knob K8	44	437	8	1.83	–
ShK	35	2,456	44	1.79	0.00
MID1-apo1	44	15,359	250	1.63	0.36
α-t-α	38	15,847	235	1.48	3.63
CD4M F23	25	8,578	83	0.97	2.37
Engrailed HD	45	9,390	76	0.81	0.00
Human EGF	42	1,677	13	0.78	0.00
Endothelin-1	21	5,314	37	0.70	2.12
EETI-II	28	4,609	28	0.61	6.98
Pin1 WW	37	2,908	15	0.52	0.00
Z34C	34	26,049	116	0.45	0.09
Avian PP	36	8,087	35	0.43	18.75
DS119	36	3,693	14	0.38	0.00
Protegrin-1	18	376	1	0.27	0.00
Tachyplesin I	17	1,458	2	0.14	0.00
Xfin ZF-31	25	3,978	5	0.13	0.00
PcTx1	41	4,643	5	0.11	0.00
De novo 3HB	43	10,787	11	0.10	0.48
Tertiapin	21	2,006	2	0.10	0.00
Mini-SH3 45	45	2,552	2	0.08	0.00
TRTX-Hs1b	29	7,621	5	0.07	0.00
Kaliotoxin	38	12,447	6	0.05	0.00
hBD-1	36	9,332	3	0.03	0.00
kB1 acyc (dep)	29	3,929	1	0.03	0.00
Charybdotoxin	36	8,944	2	0.02	0.00
Maurotoxin	34	6,098	0	0.00	0.00
Crambin	45	3,170	0	0.00	0.00
Ankyrin repeat	33	3,576	0	0.00	0.00
MCoHNE-I	33	2,397	0	0.00	0.00
MrVIB	31	2,337	0	0.00	0.00
Trp-cage	20	1,311	0	0.00	0.00
kB1 acyc (L2)	29	1,217	0	0.00	–
Mini-PDZ 43	43	495	0	0.00	–
MCoTI-II acyc	34	421	0	0.00	–
Trpzip2	12	331	0	0.00	0.00
Gurmarin	33	258	0	0.00	–
SFTI-1 acyc	14	168	0	0.00	0.00
CLN025	10	105	0	0.00	0.00
Hepcidin-25	25	41	0	0.00	–
Asteropsin A	36	53	0	0.00	–
Potato CPI	36	57	0	0.00	–
Ac-AMP2	30	17	0	0.00	–
Knob K92	35	2	0	0.00	–
Knob 2G3	33	0	0	–	–

S9. Target-level enrichment

Table S9. Targets with significant on-design enrichment relative to off-design binding. One-sided Fisher exact tests were evaluated separately for 56 antigens within each row. Bonferroni correction uses $0.05/56$.

Format	Nominal $P < 0.05$	Bonferroni
Pooled	51/56	50/56
VHH	50/56	48/56
Minibinder	40/56	36/56
Microbinder	27/56	22/56

S10. Historical comparison and sampling depth

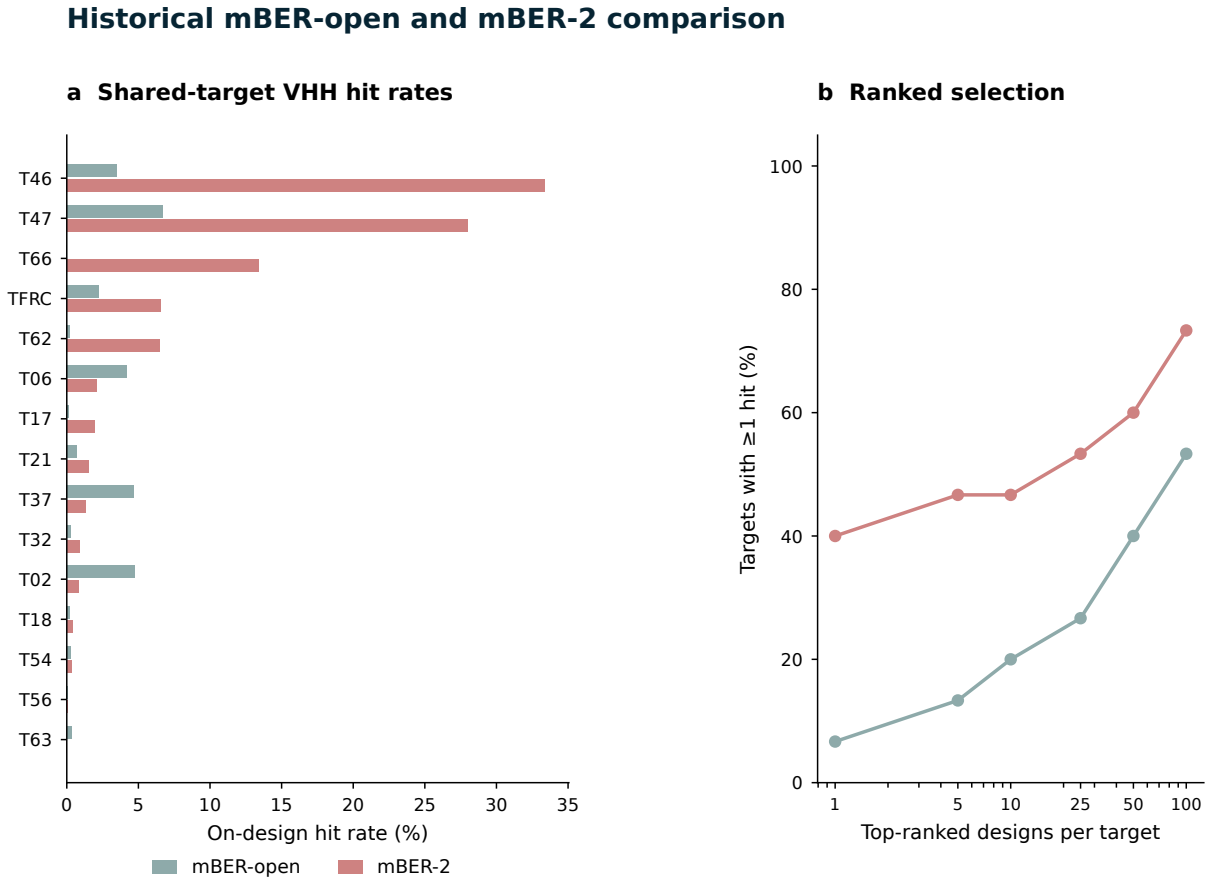


Figure S7. Historical VHH comparison with mBER-open. a. On-design hit fractions among high-signal designs for all 15 shared targets, using the S1.4 hit-calling framework with campaign-specific false-positive calibration for the mBER-1 and mBER-2 campaigns. **b.** The fraction of these targets with at least one hit when retrospectively retaining the top-ranked 1, 5, 10, 25, 50, or 100 designs per target. Ranking uses each campaign's design evaluation score. No target is excluded for having zero hits. The mBER-2 subset uses the same primary calls as the main figures. Counts are in Table S10.

Table S10. Shared-target VHH results under the same hit-calling framework with campaign-specific false-positive calibration (S1.8). Entries give hits/high-signal designs and the resulting percentage. Top-100 denominators are 100 for every target in each generation.

Target	mBER-open hits/n (%)	mBER-2 hits/n (%)	mBER-open top-100 hits	mBER-2 top-100 hits
T02	12/254 (4.72)	40/4,756 (0.84)	9	1
T06	7/166 (4.22)	25/1,189 (2.10)	5	6
T17	1/710 (0.14)	133/6,836 (1.95)	0	1
T18	1/455 (0.22)	19/4,874 (0.39)	0	0
T21	3/454 (0.66)	56/3,677 (1.52)	2	3
T32	3/1,140 (0.26)	102/11,344 (0.90)	2	32
T37	17/363 (4.68)	42/3,133 (1.34)	13	0
T46	24/689 (3.48)	622/1,863 (33.39)	17	74
T47	21/315 (6.67)	445/1,591 (27.97)	7	46
T54	1/351 (0.28)	26/7,082 (0.37)	0	5
T56	1/1,236 (0.08)	1/2,410 (0.04)	0	0
T62	1/439 (0.23)	137/2,104 (6.51)	0	27
T63	3/909 (0.33)	0/2,916 (0.00)	0	0
T66	0/710 (0.00)	255/1,904 (13.39)	0	74
TFRC	24/1,074 (2.23)	129/1,977 (6.53)	10	22

Table S11. Ranked VHH sampling depth under this paper's primary hit caller. All 56 targets contribute the indicated number of high-signal designs. Ranking is applied across all hotspots within each target.

Designs per target	Hits	Designs tested	Hit rate (%)	Targets with a hit
1	11	56	19.64	11/56
5	51	280	18.21	16/56
10	96	560	17.14	23/56
25	259	1,400	18.50	27/56
50	478	2,800	17.07	33/56
100	941	5,600	16.80	40/56

S11. Hit rates by target class and binder format

Table S12 gives the denominators underlying the class and format comparisons in Figures 1 and 2. The all-design denominator includes every primary *de novo* on-panel design, including designs without high signal; controls and other library arms are excluded. Top-100 selection is retrospective, using mBER-2's design evaluation model among high-signal designs for each target and format. Minibinders contribute 5,474 designs because two targets have fewer than 100 high-signal designs; VHHs and microbinders each contribute 5,600. Pooled rates divide total hits by total designs. Mean target rates weight each target equally, including targets with no hits, as in Figure 1a; Figure 2b uses pooled rates.

Table S12. Hit rates by target class and binder format. Each denominator column reports hits/designs (pooled percentage). The final column gives mean target-level percentages in high-signal / all-design / top-100 order. Parentheses after class names give the number of targets. All rows use the primary hit calls.

Target class (n)	Format	High-signal	All designs	Top-100	Target means (%)
GPCR (4)	VHH	3,644/14,316 (25.45%)	3,644/30,487 (11.95%)	81/400 (20.25%)	20.13 / 8.51 / 20.25
GPCR (4)	Minibinder	1,509/12,847 (11.75%)	1,509/28,897 (5.22%)	26/400 (6.50%)	9.91 / 4.30 / 6.50
GPCR (4)	Microbinder	1,032/17,580 (5.87%)	1,032/28,383 (3.64%)	58/400 (14.50%)	6.70 / 4.07 / 14.50
Receptor kinase (12)	VHH	1,181/32,774 (3.60%)	1,181/68,200 (1.73%)	99/1,200 (8.25%)	5.37 / 1.43 / 8.25
Receptor kinase (12)	Minibinder	1,249/36,322 (3.44%)	1,249/88,641 (1.41%)	118/1,200 (9.83%)	3.40 / 1.06 / 9.83
Receptor kinase (12)	Microbinder	147/41,647 (0.35%)	147/64,494 (0.23%)	13/1,200 (1.08%)	0.32 / 0.20 / 1.08
Cadherin (3)	VHH	381/13,889 (2.74%)	381/30,759 (1.24%)	80/300 (26.67%)	5.38 / 1.31 / 26.67
Cadherin (3)	Minibinder	517/7,302 (7.08%)	517/19,533 (2.65%)	9/300 (3.00%)	5.98 / 2.46 / 3.00
Cadherin (3)	Microbinder	18/8,467 (0.21%)	18/13,852 (0.13%)	11/300 (3.67%)	0.22 / 0.13 / 3.67
Ig-CAM / adhesion (11)	VHH	4,090/29,659 (13.79%)	4,090/101,150 (4.04%)	259/1,100 (23.55%)	16.00 / 3.20 / 23.55
Ig-CAM / adhesion (11)	Minibinder	1,633/27,684 (5.90%)	1,633/72,914 (2.24%)	144/1,054 (13.66%)	6.00 / 2.09 / 13.09
Ig-CAM / adhesion (11)	Microbinder	568/43,774 (1.30%)	568/65,156 (0.87%)	82/1,100 (7.45%)	1.08 / 0.69 / 7.45
Ectoenzyme (6)	VHH	725/33,287 (2.18%)	725/70,435 (1.03%)	73/600 (12.17%)	2.12 / 0.81 / 12.17
Ectoenzyme (6)	Minibinder	221/22,782 (0.97%)	221/54,151 (0.41%)	56/600 (9.33%)	1.34 / 0.54 / 9.33
Ectoenzyme (6)	Microbinder	8/40,138 (0.02%)	8/60,127 (0.01%)	0/600 (0.00%)	0.02 / 0.01 / 0.00
Transport & uptake (8)	VHH	4,531/26,707 (16.97%)	4,531/61,348 (7.39%)	108/800 (13.50%)	10.76 / 3.53 / 13.50
Transport & uptake (8)	Minibinder	1,330/18,732 (7.10%)	1,330/64,106 (2.07%)	43/800 (5.38%)	6.01 / 1.51 / 5.38
Transport & uptake (8)	Microbinder	221/34,502 (0.64%)	221/51,180 (0.43%)	18/800 (2.25%)	0.41 / 0.28 / 2.25
Other receptor (3)	VHH	540/21,483 (2.51%)	540/42,387 (1.27%)	83/300 (27.67%)	2.92 / 1.10 / 27.67
Other receptor (3)	Minibinder	454/16,451 (2.76%)	454/41,082 (1.11%)	33/300 (11.00%)	2.22 / 0.81 / 11.00
Other receptor (3)	Microbinder	31/15,047 (0.21%)	31/25,217 (0.12%)	7/300 (2.33%)	0.16 / 0.09 / 2.33
Non-receptor membrane (9)	VHH	1,295/18,652 (6.94%)	1,295/69,674 (1.86%)	158/900 (17.56%)	11.04 / 1.65 / 17.56
Non-receptor membrane (9)	Minibinder	562/19,143 (2.94%)	562/51,251 (1.10%)	43/820 (5.24%)	2.60 / 0.98 / 4.78
Non-receptor membrane (9)	Microbinder	139/31,265 (0.44%)	139/47,296 (0.29%)	10/900 (1.11%)	0.32 / 0.20 / 1.11
All targets (56)	VHH	16,387/190,767 (8.59%)	16,387/474,440 (3.45%)	941/5,600 (16.80%)	9.72 / 2.53 / 16.80
All targets (56)	Minibinder	7,475/161,263 (4.64%)	7,475/420,575 (1.78%)	472/5,474 (8.62%)	4.47 / 1.55 / 8.43
All targets (56)	Microbinder	2,164/232,420 (0.93%)	2,164/355,705 (0.61%)	199/5,600 (3.55%)	0.89 / 0.55 / 3.55