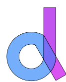


RamaX: Ultra-fast and accurate screening of large AI-designed and naive binder libraries

Diffuse Bio Team

Abstract

In this report, we share data from **RamaX** — our platform for extremely fast, sensitive, and selective screening of large binder libraries. RamaX outputs affinity data in **as little as 1 week**. We use RamaX to simultaneously screen a naive VHH library and a large AI-designed VHH binder library against several antigens in parallel. Validated against past yeast surface display (YSD) data collected for each library independently, RamaX is equally accurate, while being up to 8 times faster.

1 Background

Diffuse is collecting massive biological datasets to train our next generation of protein generative models. To this end, we have developed RamaX — our method for rapid in-house screening of large AI designed libraries and naive binder libraries. RamaX expands our ability to **rapidly discover binders for increasingly challenging targets** and to **collect large datasets quickly and cheaply to improve our generative AI protein design models**.

As generative AI systems for binder design continue to improve, we are increasingly able to—in some cases—screen tens to hundreds of designs and see success. Designing against more challenging targets, however, still requires larger-scale screening, necessitating platforms that can screen large AI designed binder libraries as quickly and cheaply as possible.

The most performant generative AI models for binder design today are trained on limited public datasets like the Protein Data Bank. AI model performance scales with training data; as a result, the rapid collection of large, accurate datasets is critical for unlocking future advances.

In this report, we share data on using RamaX to screen both a naive VHH library of $\sim 1 \times 10^9$ variants and an $\sim$ 80K AI-designed VHH library against 5 soluble target antigens in parallel. We also investigate how RamaX compares to our optimized yeast surface display (YSD) workflow. Compared to our established YSD workflow, which was previously used to screen each library independently, we found that RamaX (1) matched YSD in sensitivity and selectivity, (2) was up to 8x faster per round, and (3) could screen against all antigens simultaneously.

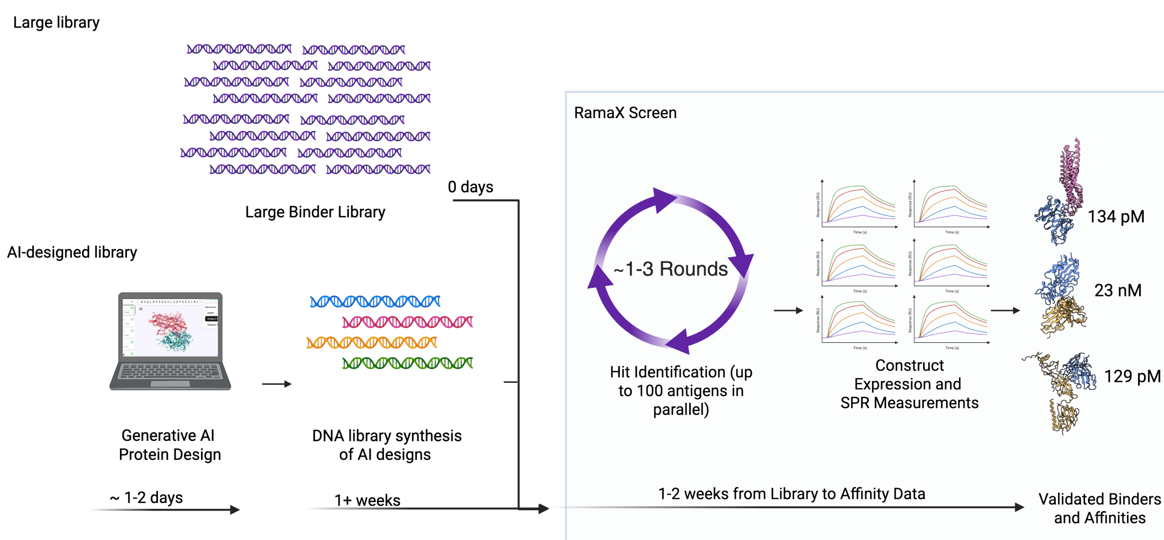


Figure 1 **Overview of the RamaX discovery workflow.** Starting from a large binder library (top) and/or an AI-designed library (bottom), binder libraries are screened against up to 100 antigens in parallel to identify hits, and affinities for hits of interest are measured via surface plasmon resonance (SPR).

2 RamaX is fast, accurate, and sensitive

In as little as 1 week, RamaX can screen binder libraries against several antigens in parallel and validate affinities for true binders (Figure 1).

With both YSD and RamaX, additional rounds of selection are necessary to identify weak binders or rare true binders in large libraries with low hit rates. One round of selection takes 1 day with RamaX, versus around 8 days with YSD. RamaX's speed advantage grows with the number of selection rounds needed.

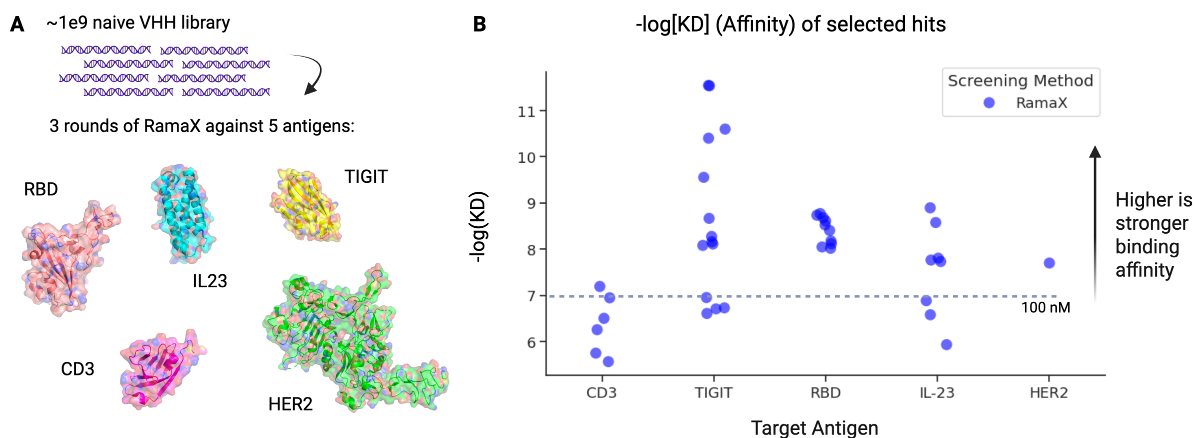


Figure 2 RamaX discovery of VHHs from a starting naive VHH library. RamaX finds binders to all 5 antigens within 3 rounds of screening, a workflow which takes < 1 week. $-\log(K_D)$ affinities measured via SPR (Surface Plasmon Resonance). Since TIGIT is a dimer, the measured K_D values represent avid binding rather than monovalent binding.

3 Screening naive libraries with RamaX

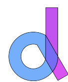
We report data on using RamaX to screen a naive VHH library of $\sim 1 \times 10^9$ variants against a diverse panel of soluble target antigens: CD3, TIGIT, SARS-CoV-2 RBD, IL-23, and HER2 (Figure 2). Across all targets, RamaX rapidly enriched antigen-specific binders within three rounds of screening, with screening done in parallel across all antigens. This workflow took less than a week.

Discovered VHH binders have a range of affinities across the targets, spanning μM to sub-nM range (Figure 2, Supplementary Figure 1). We only test a subset of putative binders here, and report the number of putative binders, number of binders screened, and number of true binders in Table 1.

RamaX enables rapid discovery of new VHHs in a ~ 1 week workflow against a range of therapeutic antigens.

	CD3	TIGIT	RBD	IL-23	HER2
Putative binders[a]	266	42	17	90	231
Screened by SPR[b]	21	25	14	21	15
Confirmed binders	5	15	12	11	1
Hit rate (%)[c]	23.81%	60.00%	85.71%	52.38%	6.67%
Best affinity (K_D)	64 nM	$<$ 100 pM[d]	1.7 nM	1.23 nM	20.0 nM

Table 1 Summary of naive library panning and SPR validation across five targets. [a] Number of putative binders reported after clustering CDR3 sequences at 75% sequence identity and after removing likely nonspecific or polyreactive clones. [b] Highest-enriching clones selected for SPR screening. [c] Percentage of confirmed binders among SPR-screened clones. [d] Sub-100 pM affinity observed against a dimeric antigen, consistent with avid binding.



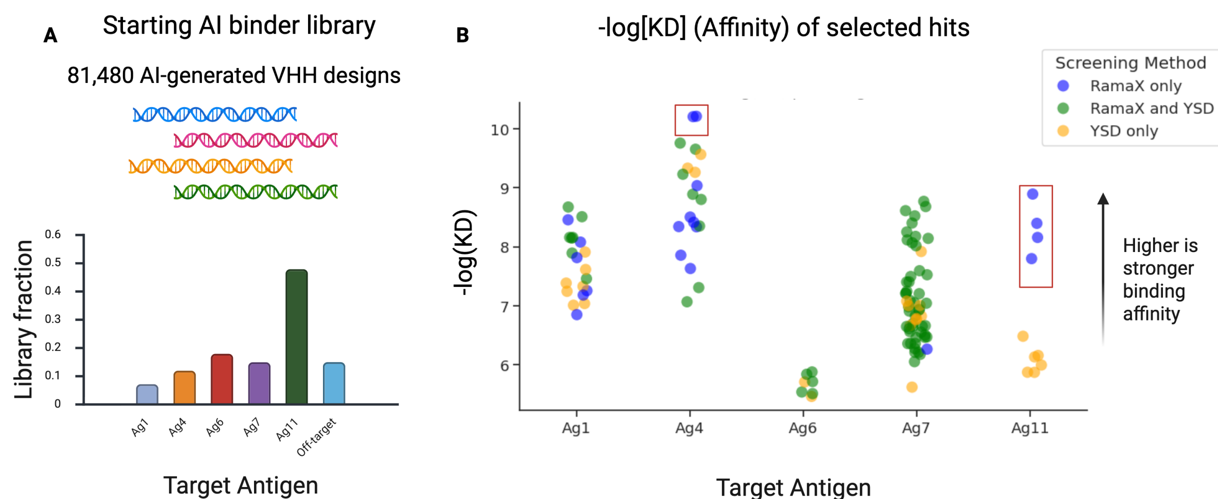


Figure 3 **Screening ML-designed VHH libraries with RamaX identifies antigen-specific binders across a broad affinity range from an ∼80K library and is comparable to screening on yeast surface display (YSD).** (A) Library set up and fraction of on-target designs per antigen. (B) $-\log(K_D)$ affinities measured via SPR (Surface Plasmon Resonance) per antigen for hits found by RamaX alone (blue), both RamaX and YSD (blue), or YSD alone (orange). RamaX can pull out both weak (e.g., for Ag6) and strong (e.g., for Ag4) AI-designed binders.

4 Screening AI-designed libraries with RamaX

We report data on using RamaX to screen an ∼80K AI-designed VHH library against 5 soluble target antigens (Figure 3A). The library is made up of many pools of designs designed against the 5 targets, as well as numerous other targets not included in this screen. These pools of designs were already screened individually via yeast surface display (YSD), and have a range of hit rates from less than 0.1% to over 25%.

RamaX successfully identifies antigen-specific binders across a broad affinity range (100pM–1μM) from the starting 80K library (Figure 3). RamaX can pull out binders on the first round of screening, even from pools with $< 1\%$ hit rates. When the underlying hit rate is low, subsequent selection rounds can find more hits. The total workflow takes less than a week [1].

RamaX demonstrates equal or better sensitivity to YSD, identifying largely the same binders as YSD (Figure 3, green), as well as new binders not found via YSD (Figure 3, blue). RamaX succeeds in finding 10–100× tighter binders than YSD for target antigen Ag11, while taking a fraction of the time (Figure 3, left red box).

We also find that RamaX has extremely good selectivity: In Table 2, we report the fraction of on-target designs that enrich per target. We see that enriching designs per target are nearly all on-target AI designs, despite starting with a large pool of AI designs against several different antigens.

We further find that RamaX correctly identifies the same binding populations as YSD (Figure 4), with minimal background off-target binding (Supplementary Figure 3).

Overall, RamaX is comparable to YSD in sensitivity and selectivity, while being nearly an order of magnitude faster.

[1]Note that this doesn't account for time to generate AI designs (1–2 days) or to synthesize the starting DNA libraries (1+ weeks, with timelines depending on oligo lengths and total library size).

Antigen	% on-target designs among enriching designs
Ag1	95.9%
Ag4	89.8%
Ag6	76.9%
Ag7	97.1%
Ag11	73.1%

Table 2 Percentage of on-target AI designs among all RamaX enriching designs per target. RamaX shows precise selectivity, enriching preferentially for on-target AI designs per target.

5 Future Work

We are continuing to improve upon the capabilities of the RamaX platform, including increasing the number of antigens that can be screened in parallel, further reducing the time and cost per iteration, and demonstrating applications beyond binder identification. AI models are not yet able to reliably predict developability-related characteristics of therapeutic candidates, for example, which is limited by the amount of data available for such properties. Diffuse is developing additional assays on the RamaX platform to enable screens for a variety of biophysical properties in addition to binding.

Pursuing increasingly challenging antigens and generating meaningfully large training datasets requires large-scale screening. While some groups have developed high-throughput platforms through automation, Diffuse focuses on developing assay systems that are intrinsically ultra high-throughput. This strategy reduces costs, supports reproducibility, and ultimately enables exponential scaling potential.

We are excited by the possibilities of rapid and cost-efficient methods for large-scale data generation, which we believe will be critical for powering the next generation of performant and useful generative AI protein design models.

Contributions and Acknowledgements

Contributors, ordered alphabetically by first name: Allison Cooke*, Anna Chase, Anne Ye, Brian Hoskins, Carla Perez*, Davian Ho, Jack Wang*, Jessica Pan, Jonas Lindquist, Michael Young, Namrata Anand, Tarun Prasad (*Lead contributors)

Figures created in Biorender (https://BioRender.com).

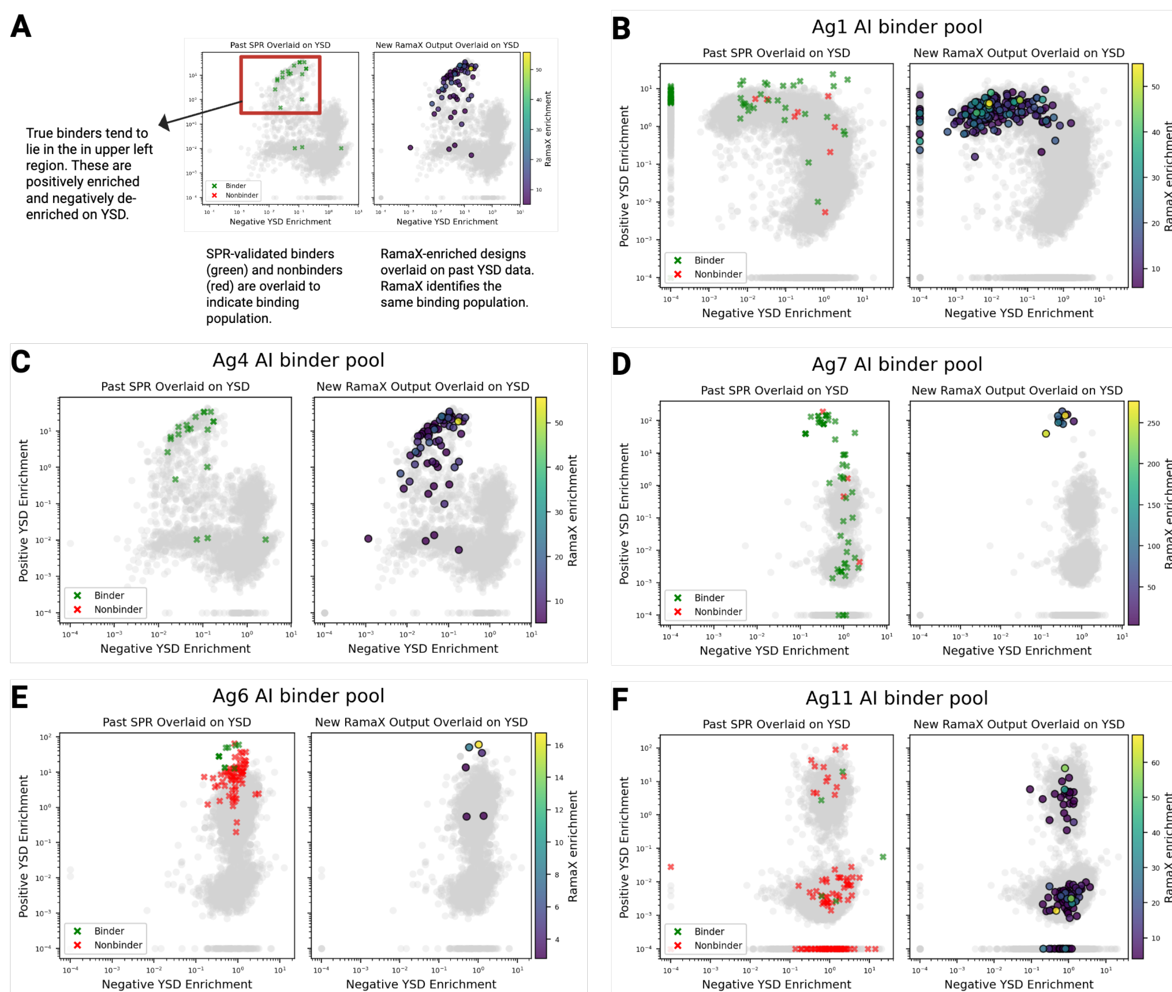
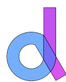
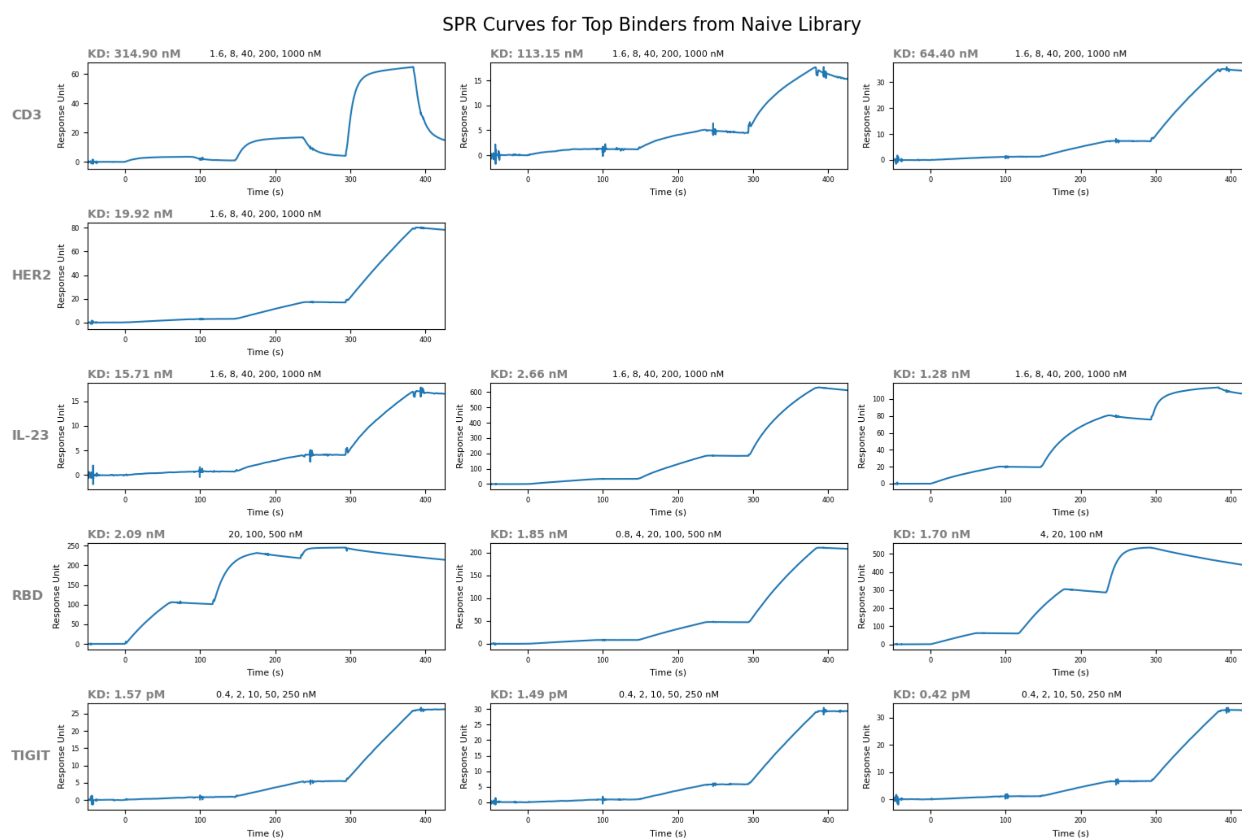
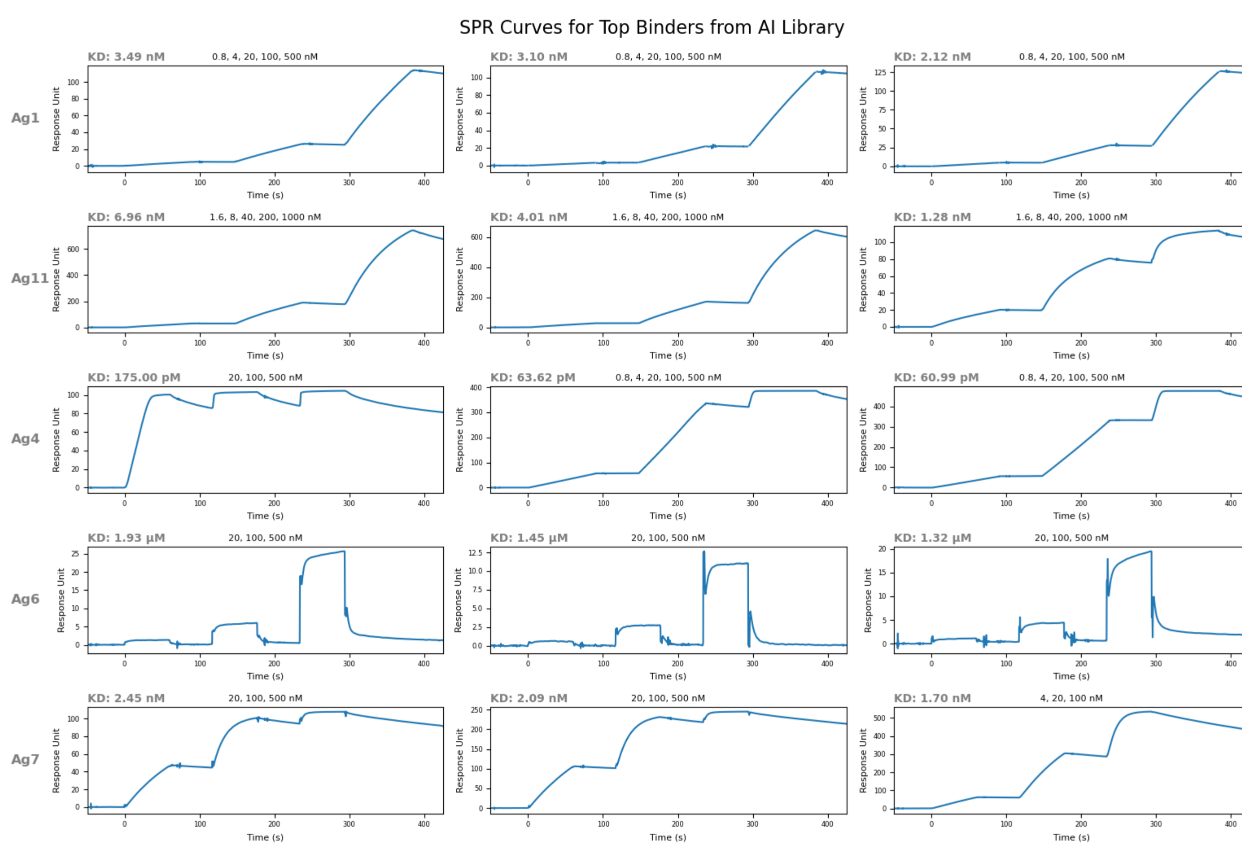


Figure 4 **RamaX recapitulates the yeast surface display (YSD) binding population.** Positive versus negative YSD enrichment is shown for each sequence in the AI-designed pool (gray). Positive enrichment is defined as double-positive (binding + expression) gate abundance divided by input abundance, while negative enrichment is defined as expression-only gate abundance divided by input abundance. True binders are expected to be enriched in the double-positive gate and de-enriched in the expression-only gate, and therefore cluster in the upper-left region of the plot. SPR-validated binders (green) and non-binders (red) are overlaid to indicate the experimentally confirmed binding population. RamaX-enriched designs, overlaid on historical YSD data, co-localize with this binding population, demonstrating that RamaX not only enriches validated on-target binders per antigen but also correctly selects the same binding population identified by YSD.

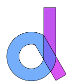


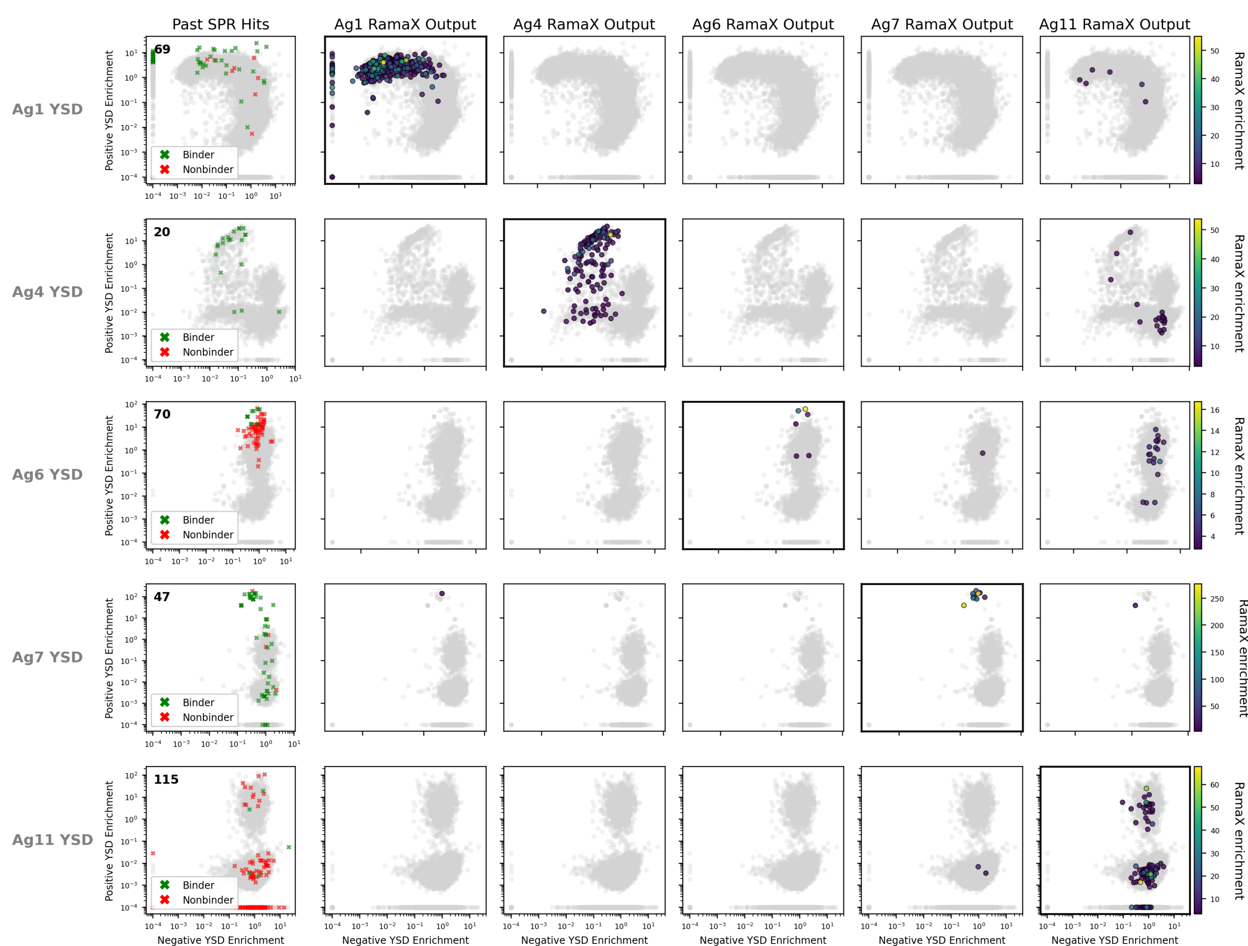


Supplementary Figure 1 **SPR validation of RamaX-selected naive library binders.** Surface plasmon resonance (SPR) binding curves for the top three RamaX-enriched designs per antigen, ranked by apparent affinity. Sensorgrams were fit to derive equilibrium dissociation constants (K_D). Antigen concentrations used in each titration are indicated above the corresponding plots. Note: In the naive panning, we also spiked in the large AI-designed pool. Ag7 is RBD, and we find that the top AI-designed clones get enriched in the naive panning experiment as well, alongside new naive library clones. The second and third top hits for RBD here are also top AI-designed pool hits.



Supplementary Figure 2 **SPR validation of RamaX-selected AI-designed binders.** Surface plasmon resonance (SPR) binding curves for the top three RamaX-enriched designs per antigen, ranked by apparent affinity. Sensorgrams were fit to derive equilibrium dissociation constants (K_D). Antigen concentrations used in each titration are indicated above the corresponding plots.





Supplementary Figure 3 RamaX recapitulates YSD-defined binding populations with high on-target selectivity and minimal off-target background. Yeast surface display (YSD) positive versus negative enrichment is shown for AI-designed sequences, where true binders cluster in the upper-left region (high positive enrichment, low negative enrichment). (Left) SPR-validated on-target binders (green) and non-binders/off-target designs (red) overlaid on YSD enrichment space, defining the binding population. (Right, columns) RamaX-enriched designs overlaid on the same YSD enrichment plots for each antigen. RamaX selectively enriches the YSD-defined binding population while showing little to no enrichment of off-target designs (columns), demonstrating fine specificity and low background binding.