Mapping the diverse topologies of protein-protein interaction fitness landscapes

Shannon S. Lu*[1], Matthew J. Styles*[1], Cheng Frank Gao*[1], Aditya Nandy[1], Christopher Basile[1], Joshua A. Pixley[1], Siyuan Tao[1], Aaron R. Dinner[1], Suriyanarayanan Vaikuntanathan[1], Bryan C. Dickinson**[1, 2]

[1]Department of Chemistry, University of Chicago, 5735 S. Ellis Ave., Chicago, IL 60637
[2]Chan Zuckerberg Biohub, Chicago, IL 60642
**These authors contributed equally to this work*
***Corresponding author*. Email: *dickinson@uchicago.edu*

Abstract:

De *novo* binder discovery is unpredictable and inefficient due to a lack of quantitative understanding of protein-protein interaction (PPI) sequence-function landscapes. Here, we use our PANCS-Binder technology to perform >1,300 independent selections of various library sizes and compositions of a randomized small protein to identify binders to a panel of 96 distinct target proteins. For successful selections, we discovered reproducible fitness landscapes that group into a few, target-specific, clusters. Each cluster defines a minimal binding motif whose frequency is inversely proportional to the number of specified amino acids (~2–8) and determines selection success, which is quantifiable by the density of binders to the target within a theoretical sequence space. We leverage these data to develop a supervised contrastive learning approach that discriminates binders from non-binders and demonstrates generalization beyond a threshold amount of data. Together, this framework renders PPI landscapes measurable and predictive, accelerating *de novo* binder discovery and optimization.

Introduction

Protein-protein interactions (PPIs) underlie virtually every cellular process – structure, localization, function, and regulation are governed in large part by PPIs (*1-3*). Beyond native biology, proteins engineered to bind a protein of interest, termed binders, are invaluable as research tools and therapeutics (*4-7*). In nature, new PPIs frequently evolve by gene duplication and co-evolution of the interacting pair (*8*) or are induced via post-translational modification (PTM) of one protein such that it can interact with general binders of that PTM (*9*). *De novo* interactions between two proteins can also arise naturally, usually via mutagenesis of one interacting partner, through somatic hypermutation (*10*) and diversity generating retroelements (*11*). Directed evolution techniques conducted in the laboratory aim to mimic these natural systems to evolve *de novo* binders by first building a library of diverse variants and then selecting for active variants. Although a number of selection methods (*12-16*), high-throughput screening (*17-19*), and computational design (*20-22*) techniques have been developed for synthetic *de novo* binder discovery, these are generally slow and inefficient, with unpredictable rates of success. Part of this failure likely stems from a misunderstanding of the determinants of successful experimental selections and the lack of a theoretical understanding of the factors driving the discrimination between binders and non-binders.

To identify binders, both natural and directed *de novo* binder evolution rely on sampling a subset of the PPI evolutionary sequence space, which conceptually encompasses all possible variants. This PPI evolutionary landscape perspective has been employed to determine which antigens are amenable to binder discovery by hypermutation by the adaptive immune system (*23-25*); how viruses can evolve new cell-recognition elements (*26-28*); and the intrinsic amenability of any given protein to *de novo* binder discovery (*29-32*). We sought to address two perennial questions concerning the emergence of *de novo* PPIs: 1) What is the a priori probability (denoted as binder density) that a random scaffold variant binds a given target? 2) To what extent is this binder density governed by the randomization strategy (denoted as design) in sequence space and to what extent does the density depend on the identity of the target (e.g. how “easy” or “difficult” a target is)? Methods for mapping fitness across a broad array of targets and protein variant libraries are key to answering these questions. Current methods, however, are hampered by tradeoffs in fidelity and throughput capacity (*13*). For example, the highest throughput selection methods – such as mRNA or ribosome display – can sample 10^{12+} variants simultaneously but are low fidelity. Conversely, high-fidelity screening methods, such as on chip ribosome display or two-hybrid systems, are low throughput, capable of sampling ~$10^{<6}$ variants at a time. We recently developed a high fidelity, high-throughput screening platform, Phage-Assisted Non-Continuous

Selection of Binders (PANCS-Binders) (*16*), which is capable of rapidly assessing billions of PPIs across dozens of targets in parallel, providing a new opportunity for large-scale high-resolution PPI landscape mapping.

Here, we leveraged the PANCS-Binder platform to map the PPI fitness landscapes between a protein binder scaffold and a suite of diverse target proteins. For the protein binder scaffold, we used affibody, a small (58 amino acid), highly stable three-helix bundle (a PPI motif found in nature) that is widely used in *de novo* binder discovery (*16, 33-35*). We lay out a pipeline for characterizing several critical aspects of the affibody-target PPI fitness landscape topology, including how common binders are within the sequence space (binder density), how many types of binders exist within the sequence space, whether high affinity binders have rare or common motifs, and whether the fitness landscape is reproducible and predictable. We determined the reproducibility of binder discovery from randomly sampled variants from library designs to assess binder density and characterize PPI fitness landscape topologies using a panel of 96 diverse target proteins and 10 distinct randomization strategies. Binder density varied widely by target, from more than 1 in 100,000 variant to less than 1 in 10,000,000,000 variants. Analysis of the ~1,300 selections disentangled the relative contributions of the design strategy vs. effects from each unique target, revealing that binder density is mainly controlled by target features. Finally, we used the thousands of novel binder sequences and hundreds of thousands of non-binder sequences obtained from the selections to train a neural network to discriminate between binders and nonbinders. In line with the statistical findings of the previous sections, we find that the neural network is able to generalize and correctly identify binders in previously unseen data sets. This analysis also reveals a highly heterogeneous topology across targets. This work elaborates a new conceptual framework for understanding *de novo* binder discovery through the simple lens of a minimal binding motif, delineates how to experimentally map the fitness landscape of a PPI, and provides a pipeline to produce extensive binding data sets for machine learning (ML) approaches.

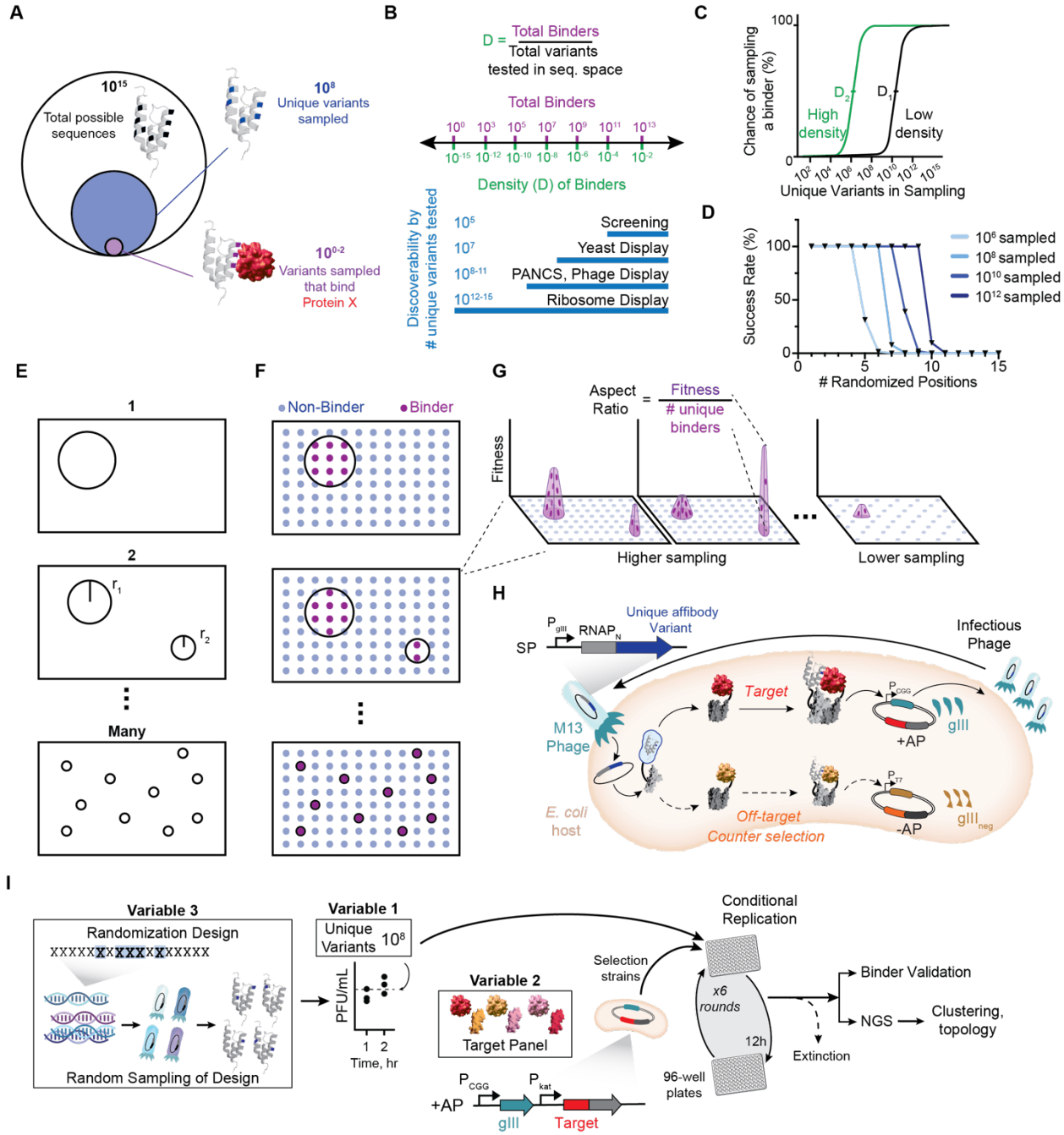

Figure 1: Defining the topology of a PPI landscape. **A**) The affibody sequence space includes all possible sequences based on the randomization strategy (white circle). A subset of this sequence space can be encoded into a physical random sampling of the overall sequence space (blue circle). Finally, using selections, the binding variants within the sampling can be identified for a given target protein (purple circle). **B**) We define the density of binders as the total number of binders divided by the number of possible sequences within a given theoretical sequence space. We show the affibody-target binder density (assuming 10^{15} possible sequences) and the maximum sampling size of different selection techniques. **C**) A transition in the chance of sampling a binder can be used to determine the binder density experimentally. **D**) The minimal number of amino acids required to specify a binding motif determines the frequency of the binding

motif (**Fig. S1a**), the number of binding motifs within a random sampling of variants (**Fig. S1b**), and the chance of success for a given sampling size. **E**) A representation of the possible underlying topologies of a PPI where clusters of binders with a common binding motif (from one to many) are represented by circles whose radii vary with the relative density of these binding motifs **F**) Random sampling can be used to determine the underlying topology. **G**) Fitness of variants within a cluster can be mapped onto a third dimension to show the fitness of different clusters. **H**) PANCS-Binder selection. Defined libraries of affibody fused to the N-terminal portion of a split-RNA polymerase are encoded in replication deficient phage (one unique variant encoded per phage genome) are passaged on *E. coli* selection strains that express a target protein fused to the C-terminal portion of a split RNA polymerase (RNAP). PPI-dependent reconstitution of the RNAP drives the expression of *gIII*, allowing phage replication. The replication rate of an individual phage is dependent on the affinity of the interaction of its encoded affibody for the target protein; therefore, endpoint phage populations represent a rank ordering of the binding affinity of affibody variants. **I**) General workflow for PANCS. Following generation of a a sampling of the randomization design, outgrowth, and titer estimation, the sampling is subject to PANCS. After six 12-hour rounds of selection, the titer of the endpoint population is measured, and populations with titers >1 x 10^7 PFU/mL are submitted for next-generation sequencing (NGS). Following validation of the presence of full-length scaffolds, the top variant is subcloned and tested for binding activity, and the NGS data are subject to clustering analysis to map the topology of the PPI landscape.

Results

Conceptualizing the topological characteristics of a PPI sequence landscape

The potential sequence space for a 58 amino acid protein interacting with a single, invariant target includes 10^{75} variants, akin to the number of atoms that exist in the universe, which is far too vast to search in its entirety. *De novo* binder discovery adopts the heuristic tenets that a well-folded protein scaffold is more likely to bind a target, and that randomization should be focused on a few select residues (*36-38*). This strategy is frequently adopted in natural generation of diversity systems such as antibodies (*39*) and phage coat proteins (*11, 40*). In practice, 10-15 positions are randomized, yielding 10^{13-19} theoretical variants (**Fig. 1A**). However, experimental methods can only sample ~10^6 to 10^{11} variants, far less than 1% of the theoretical variants. Therefore, if a hit is consistently identified from screening 10^9 unique variants, this implies that the density of binding variants within this sequence space is at least >10^{-9} (**Fig. 1B**). Binder density, a critical, target-dependent metric for discovery and randomization design, can therefore in principle be determined empirically by repeatedly sampling different numbers of unique variants, that are subsampled independently from the total theoretical set of variants, for any that bind to the target (*32*) (**Fig. 1C**). We hypothesize that binder density is primarily determined by the minimal number of amino acids that must be specified to create a minimal binding motif (**Fig. 1D**). The frequency of this minimal binding motif within the randomization design (**Fig. S1**) constrains how designs can be improved: increasing the diversity allows testing of more potential motifs but reduces the frequency of all motifs encoded in the randomization design. To further

illustrate these concepts, we set out to visualize PPI sequence spaces using the conceptual framework outlined below.

One can represent the relationship between fitness and sequence as a landscape with high and low fitness represented by peaks and valleys, respectively (*41, 42*). In the context of PPIs, this representation has been used to analyze deep mutational scanning and affinity maturation datasets that investigated a single peak or the potential paths to traverse between two peaks (*43-50*), and more rarely to analyze *de novo* binder discovery (*51, 52*). We aimed to define the topological characteristics of a "PPI landscapes" and use *de novo* binder selection to experimentally probe these characteristics. For a given target, we can sort binders into clusters based on the presence of distinct binding motifs – for a given target from the number of distinct binding motifs that binders may cluster into could range from one to many. Representing all possible sequences in a 2D plane, we can indicate clusters as circles where the geometric size of the cluster is expected to be proportional to the number of binders it contains (**Fig. 1E**). To experimentally map this landscape, we randomly sample variants (dots, **Fig. 1F**) and use sequence similarity among binders to sort them into binding motif clusters. The fitness of each binder sequence or cluster can then be visualized in a third dimension (**Fig. 1G**). We expect that the sampling sizes will dictate the possible clusters that can be observed: if a cluster has a frequency of 10^{-8}, then sampling 10^{6} variants is unlikely to observe this motif, but sampling 10^{10} variants will consistently observe this motif (**Fig. 1G**). Additionally, rarer clusters could have higher fitness than more common clusters; therefore, we define the height/radius as the "aspect ratio" of a cluster. For similarly fit variants/clusters, a low aspect ratio would indicate a relatively common motif while a high aspect ratio would indicate a relatively rare motif (e.g. a motif specified by four amino acids would have a low aspect ratio relative to a motif specified by six amino acids if the motifs confer similar fitness). We sought to apply this conceptual framework to experimental random samplings of a randomization design; using PANCS-Binder selections (**Fig. 1H**), we systematically manipulated three variables: 1) the number of unique variants sampled; 2) the targets tested; and 3) the randomization design (**Fig. 1I**).

Experimentally mapping the topology of affibody-target PPI landscapes

To determine the density of binding variants within a scaffold-target sequence space, we first generated distinct samplings of the affibody scaffold at different sizes of unique variants with a previously identified randomization design (*53*). In this work, a "design" refers to a scaffold and randomization strategy with a set of possible variants, and a "sampling" refers to a specific, physical phage population with a specific number of unique variants. For example, the

randomization used for design A consists of 2.4 x 10^{15} distinct possible amino acid sequences (**Fig. 2A, Table S1**). Using this design, we made 24 independent samplings (six independent replicates of 10^5, 10^6, 10^7, and 10^8 unique variants samplings) (**Fig. 2A**) (**Table S1**), which should have nearly no overlapping variants between one another. Based on our previous work, we selected nine target proteins (see **Table S2** for target details) and performed PANCS-Binder selections with these 24 samplings **(Fig. 2A**). At the end of each PANCS-Binder selection, phage titer was used to determine which selections enriched binder variants. Those with high titers were sent for next-generation sequencing (NGS) and the dominant variant was subcloned into a separate binding assay to validate binding activity of the most enriched variant (see Supplemental Text for further details of selection validation, **Figs. S2-4**, **Source Data 1**). We observed a wide range of affibody (A)-target binder densities from >10^{-5} to <10^{-8} for the targets IFNG and MAX, respectively (**Fig. 2B, Table S3).** Reproducibility between replicate samplings was high: greater than or equal to 5/6 samplings having identical results for all targets except CDKN1A (only 50% hits at 10^8 variants). This suggests that the assay outcomes are not dictated by random chance, consistent with prior studies (*54, 55*). Therefore, when a single replicate hits, it is likely that replicate unique samplings of the same size would also produce hits. Likewise, if a single replicate does not hit, replicate samplings would likely also fail to produce a hit, except when the density is near the sample size being tested (e.g., CDNK1A). Finally, we observed that the binding affinity, as observed in our *E. coli* RNAP complementation reporter, of the most fit variant (i.e. highest percentage of endpoint population) from each successful selection increases with larger sampling sizes (**Fig. 2C**).

Next, we sought to characterize the topology of a PPI landscape, using IFNG-affibody binding as a representative example. We pooled all variants that enriched to >0.1% of the population across all successful selections for all targets and performed hierarchical clustering as a surrogate visualization of the PPI landscape (Methods, **Source Data 2**). To visualize the clusters in 2D, we plotted sequence variation on a t-SNE embedding (*56*). This revealed discrete, well-separated sequence communities for IFNG across sampling sizes spanning 4 orders of magnitude (**Fig. 2C**). We observed high reproducibility, both in the number of prominent clusters (>1% of the post-selection population, **Fig. 2D**) and in the distribution of binders between these clusters (**Fig. 2E**) across selections with the same sampling size. As the sampling size increases, the binders converge primarily onto two clusters with distinct binding modes, clusters 7 and 22 (**Fig. 2F, S5**). Using both the measured fitness for individual dominant variants from each selection (**Fig. S4**) and the relative population of variants within each selection (NGS reads, **Fig. S3**, **Source Data 1 and 2**), we computed the fitness of each variant and cluster (see Methods for

details). As expected from the trend with the dominant variant (**Fig. 2C**), the fitness of variants and clusters increases along with sampling size (**Fig. 2F**). We again observed high reproducibility in fitness of the dominant variant (**Figs. S4, 6-10**) and of the prominent clusters (**Fig. 2F**) between selections with the same sampling size. We mapped fitness onto our topology (**Fig. 2G**) and were able to readily visualize the growth of the clusters across increasing sampling size. We calculated the aspect ratio of clusters for IFNG (**Fig. 2H**): at high sampling sizes ($\geq 10^7$), no aspect ratios were found to be larger than 3. This result is consistent with the observation that the two most fit clusters at 10^8 sampling size are observed at much lower sampling sizes; if a high aspect ratio cluster were observed, we would have expected to identify a cluster of high fitness that appeared only at the highest sampling size. Because most clusters have few sequences from each sampling at our lowest sampling size, we observed several clusters with high aspect ratios (**Fig. 2H**). Interestingly, high aspect ratios might be a general predictor of sampling at the boundary of what is needed to identify a new binding motif. Critically, these results demonstrate that binder density and PPI landscape topology can be experimentally determined and are highly reproducible.

Similar analyses for each of the other five targets that produced hits (**Fig. 2I**, **Figs. S6-10, Source Data 2**) illustrate that each PPI landscape has a unique topology consisting of varying numbers of prominent clusters (**Fig. 2J**), numbers of sequences within each prominent cluster (**Fig. 2K**), relative fitness between clusters (**Fig. 2L**), and presence (or lack thereof) of high aspect-ratio clusters (**Fig. 2M**). Again, we observe high reproducibility between selections at the same sampling size (**Figs. 2I-M, S6-10**). Although the fitness of each of the top two clusters in each target are not significantly greater at 10^8 compared to smaller sampling sizes (**Fig. 2F, S6-10**), we nonetheless noted an overall increase in fitness with increasing sampling size (**Fig. S4**). Four general topologies were observed: IFNG displayed two prominent, well-separated, clusters; XIAP displayed two prominent but highly similar clusters; KIX displayed a combination of both of these feature- prominent, well separated clusters and prominent, closely related clusters; and LC3B, PP2A-B'',and CDKN1A have much more sparse landscapes, consistent with the sampling size being nearer to the binder density for these two targets. The consistency between replicate selections, when the sampling size exceeds a critical amount corresponding to an accessible binder density, reflects both the fidelity of the PANCS-Binder selection platform and the underlying predictability of the PPI fitness landscape.

Next, we sought to investigate our model of how minimal binding motifs relate to the fitness landscape. We identified the minimal binding motifs for the dominant clusters for each target (excluding clusters with <10 unique variants identified, which we postulated is required for robust

identification of convergence) and calculated how frequent that motif is within the randomization design (**Fig. S11**). We mapped the convergence at each randomized position across the variants within each cluster (blue in the affibody structures in **Fig. 2I**, **Fig. S11**) to visualize each binding motif. While the exact binding motif varied between targets, similar positional motifs are observed across targets. Binders to KIX, LC3B, and PP2A-B'' have motifs that act as single helix binders while binding motifs for XIAP and IFNG utilize both helices on one face of the affibody. Next, we compared the frequency of a binding motif within the randomization design (**Fig. S11**) to the observed experimental binder density (**Fig. 2B**); we observed strong concordance. KIX and IFNG had binder densities $\geq 10^{-5}$, and have motif frequencies of 10^{-2} and ~5 x 10^{-6}, respectively. Additionally, the binder motif for PP2A-B'' has a frequency of 10^{-6} and observed binder densities of 10^{-6}. While initially LC3B appears to be an exception (observed binder density of $>10^{-5}$ but a 10^{-6} motif frequency), we see that at low sampling sizes, this motif (cluster 28) is not identified (**Fig. S10**), rather a more frequent motif (cluster 5) is identified. This finding is in line with the observed high aspect ratios for LC3B (**Fig. 2M**)—by increasing sampling size we accessed a more complex binding motif (more specified amino acids) than we were able to at lower sampling sizes. Note, we also observe a clear bifurcation in the affinity of the top variant from selections at the higher (10^{7-8}) and lower (10^{5-6}) sampling sizes (**Fig. S4**). Finally, XIAP has an observed binder density between 10^{-5} and 10^{-6} but a motif frequency as high as 2 x 10^{-4}; we suspect that within the non-converged positions (**Fig. S11**) that there are likely strong preferences for an amino acid within a pair of positions allowing for this amino acid to not appear as converged. For example, two adjacent residues have Arg 30-40% converged at each position (low enough to not be considered originally as part of the binding motif); however, if the underlying binding motif simply requires an Arg at one of these two positions, then the frequency of the motif would decrease to ~10^{-5} (approaching the observed binder density). These results are in strong agreement with our simple model of binding motif frequency as the determinant of observed binder density. We next sought to test whether the topological characteristics observed for this limited 9-target set generalize across many targets.

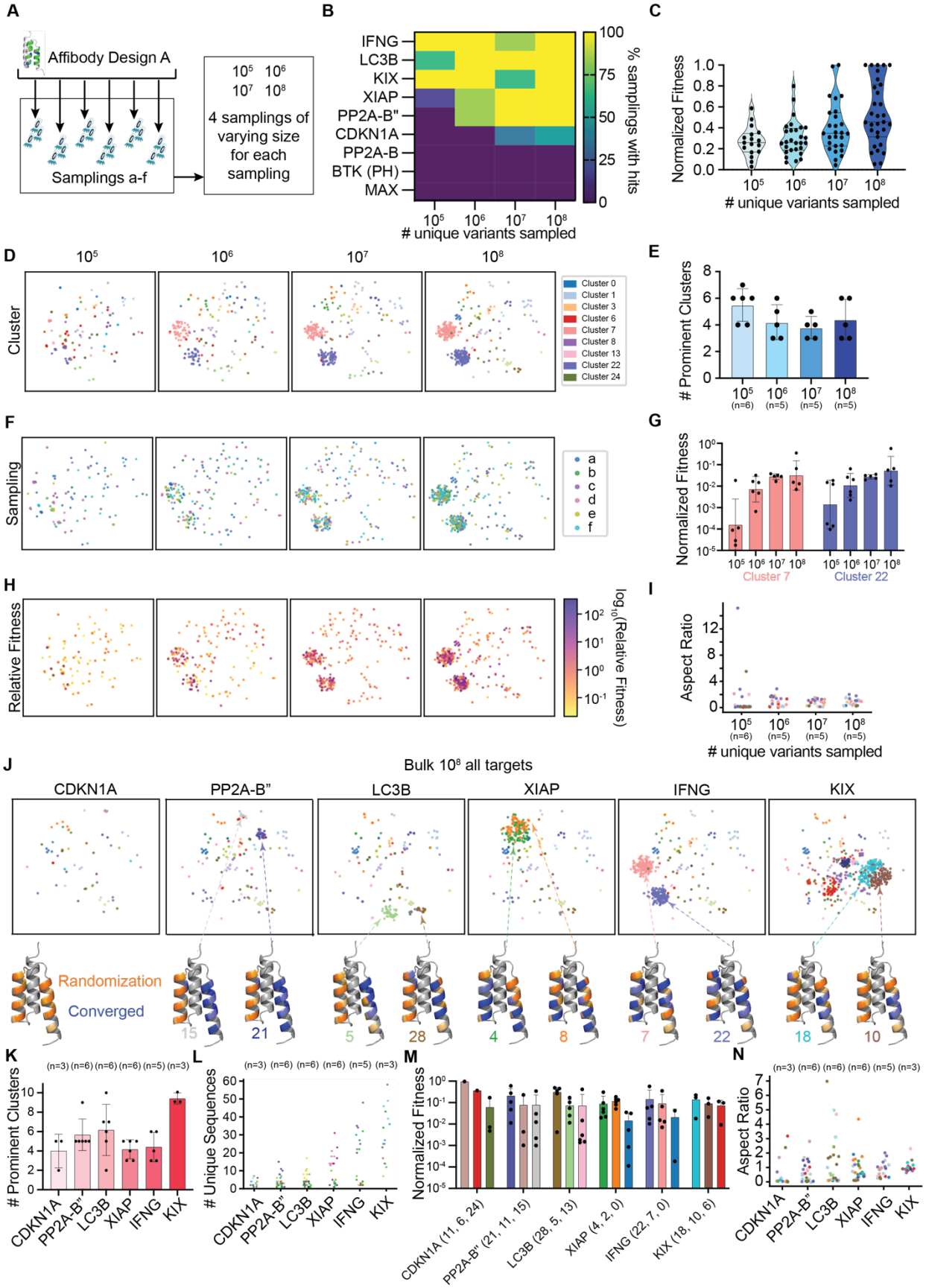
A
Affibody Design A
Samplings a-f
10^5 10^6
10^7 10^8
4 samplings of varying size for each sampling
B
IFNG
LC3B
KIX
XIAP
PP2A-B"
CDKN1A
PP2A-B
BTK (PH)
MAX
unique variants sampled
% samplings with hits
C
Normalized Fitness
unique variants sampled
D
Cluster
Cluster 0
Cluster 1
Cluster 3
Cluster 6
Cluster 7
Cluster 8
Cluster 13
Cluster 22
Cluster 24
E
Prominent Clusters
F
Sampling
G
Normalized Fitness
Cluster 7
Cluster 22
H
Relative Fitness
log10(Relative Fitness)
I
Aspect Ratio
unique variants sampled
J
Bulk 10^8 all targets
CDKN1A
PP2A-B"
LC3B
XIAP
IFNG
KIX
Randomization
Converged
K
Prominent Clusters
L
Unique Sequences
M
Normalized Fitness
CDKN1A (11, 6, 24)
PP2A-B" (21, 11, 15)
LC3B (28, 5, 13)
XIAP (4, 2, 0)
IFNG (22, 7, 0)
KIX (18, 10, 6)
N
Aspect Ratio

Figure 2. Mapping the PPI landscapes of affibodies and several targets. **A**) Generation of 24 independent samplings (six replicate samplings at four sizes) of design A for selections on a 9-target panel. **B**) Number of unique variants sampled versus % of samplings with hits (heatmap for each target). As an example, we show PP2A-B'' binder success rate versus sampling size shows a Langmuir like curve. **C**) Experimental relative fitness as measured in the split T7 RNAP *E. coli* luciferase assay for the most dominant variant from each selection. Fitness of each binder is normalized to that of the best binder discovered for each target. Binders are categorized by the selection sampling size. **D**) Scatter plots on a t-SNE embedding showing all variants with >0.1% of NGS reads from all samplings that produced a binder for IFNG. Each sequence is colored by cluster association. **E**) Bar plot indicating the number of prominent clusters (>1% of NGS reads) for sampling at each size of unique variants for IFNG, for all samplings in a selection where the top variant was tested. **F**) Scatter plots on a t-SNE embedding, where color indicates from which sampling a variant was derived. **G**) Bar plot showing the relative fitness (see Methods for calculation) of the two most dominant clusters across sampling size, colored by cluster identity. Each bar represents the geometric mean and standard deviation (SD). **H**) Scatter plots on a t-SNE embedding where color indicates relative fitness (see Methods). **I**) Scatter plot showing the aspect ratio for each cluster from each selection (color indicates cluster identity). **J**) Scatter plots on t-SNE embeddings for all 10^8-unique variant-sized samplings across all targets tested, colored by cluster assignments. For select clusters (indicated), the convergence (blue) of randomized positions (orange) on the affibody scaffold (gray) are shown to indicate the position of binding motifs (darker color indicates greater convergence or randomization; see Fig. S10 for convergence analysis). The randomized positions for the design are indicated on the left-most affibody structure. **K**) Bar plot showing the number of prominent clusters for all 10^8 samplings across all targets, for all samplings in a selection where the top variant was tested. **L**) Bar plot showing the number of unique variants in the prominent clusters for all 10^8 samplings across all targets, for all samplings in a selection where the top variant was tested. Each bar represents the geometric mean and SD. **M**) Bar plot showing the fitness of the top three prominent clusters for all 10^8 samplings across all targets where fitness is relative within a target. **N**) Scatter plot showing the aspect ratio of prominent clusters across all 10^8 samplings for each target.

Characterizing diverse affibody-target PPI landscapes

Given the reproducibility between replicate samplings and the high hit rate at 10^8 unique variants sampled (binders were enriched for 6 of 9 targets), we next performed selections with one sampling of design A at 10^8 unique variants across a broad, 96-target panel (**Table S2**). This panel contains proteins of varying size, structural motifs, disordered content, function, localization, and surface area (**Table S2**). Out of the 96 targets, 49 gave hits from the 10^8 unique variants (**Fig. 3A**) (**Fig. S12** as described in **Supplementary Information**). To further stratify those 49 targets by affibody binder density, we performed a selection with a smaller, independent 10^6 unique variant sampling. For the remaining targets that failed to enrich binders with the initial 10^8 sampling, we performed selection with a larger ~10^{10} unique variant sampling (phage titer, NGS, and binder validation were performed as described in **Supplementary Information**, **Source Data 3, Figs. S12, 13**). From the success or failure of these selections, we separated targets into four

categories: high density (hits for 22 targets at 10^6), moderate density (hits for 27 targets at 10^8 but not 10^6), low density (hits for 5 targets only at 10^{10}), and undetectable density (no hits for 42 targets at 10^{10}). Unexpectedly, increasing to the sampling size to 10^{10} produced hits for only 5 out of 47 targets with densities $<10^{-8}$. This suggests that routine sampling with large sampling sizes is unlikely (~10% of cases) to identify additional binding motifs beyond those already represented at 10^8 sampling sizes, even as the fitness of variants identified from 10^{10} sampling sizes are 3.3-fold higher than those at 10^8 (**Fig. S14**). This is a strong indication that targets with undetectable binder density are either intrinsically resistant to binding or are structurally incompatible with the affibody scaffold.

As we observed with our 9-target panel, the top variant identified in selections with larger sampling sizes generally has higher fitness compared to variants identified in selections with lower sampling sizes (**Fig. 3A**). Again, we performed hierarchical clustering using the data from the selections with 10^6 and 10^8 sampling size on targets with binder density greater than 10^{-8} (**Fig. S15**, **Table S4, Source Data 4**). We observed 0-8 prominent clusters for sampling at 10^8, and the range in number of prominent clusters stayed consistent across three example subsets (10^6 sampling/10^{-6} density (yellow), 10^8 sampling/10^{-6} density (green), 10^8 sampling/10^{-8} density (blue); **Fig. 3B**). This suggests that there is a general ceiling in the number of binding motifs for target-affibody pairs, consistent with our 9-target panel. Because our samplings were performed close to the binder density for many targets and were not performed in replicate, we found fewer sequences (often <10) for each cluster, precluding robust analysis of binding motif convergence for most clusters (**Fig. S15**). As with our initial 9-target panel, relative fitness increased with higher density sampling (**Fig. 3C**) and cluster aspect ratios were found to be less than 3 in most cases (**Fig. 3D**). In only 3 cases (of 18 possible), we observed a different dominant cluster when sampling at 10^8 verses 10^6 (**Table S4**); consistent with only observing LC3B (1/6) having this feature in the 9-target panel. The 96-target panel data, which is remarkably consistent with the 9 target panel data, demonstrates the general features of the affibody-target PPI topology, which likely extends to untested target panels: ~50% of targets with a density $>10^{-8}$, ~10% of targets with a binder density $<10^{-8}$ have a density $>10^{-10}$, targets have 1-8 prominent clusters (unique binding motifs), and 4% of prominent clusters (1/6 targets) have aspect ratios >5 (indicative of a more complex binding motif that would not be identified at lower sampling sizes).

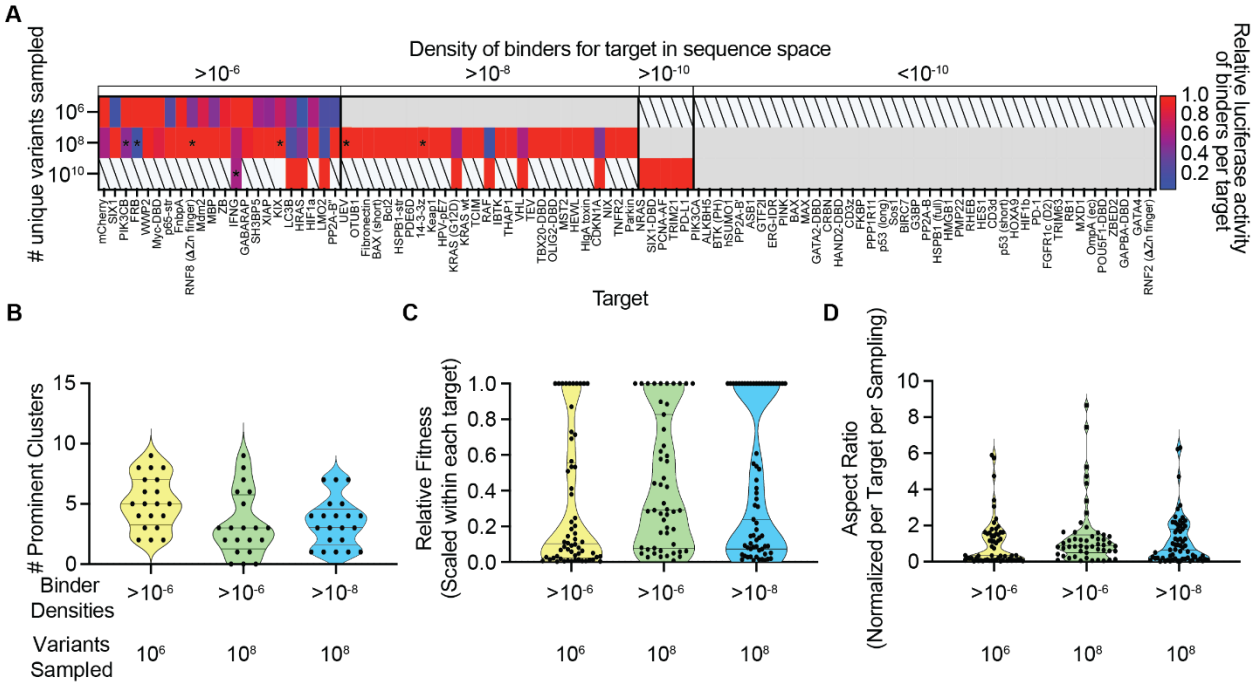

Figure 3: Quantifying the generalizability of affibody-target PPI landscapes. A) Heat map splitting the 96-target panel into four categories of binder density with respect to design A: $>10^{-6}$, $>10^{-8}$, $>10^{-10}$, and $<10^{-10}$, based on their hit rate according to the number of variants tested. The relative luciferase activity of binder per target was calculated by subtracting the background (negative control signals) from those of each binder, then normalizing to the best binder discovered for a target. Boxes in gray indicate no hit, boxes with slashes indicate not tested due to presumably predictable outcomes. Boxes with asterisks indicate selections where a variant less than three-fold abundant compared to the top variant was tested. **B**) Number of prominent clusters in each selection, sorted by binder density category and number of unique variants sampled. Prominent clusters are defined by those possessing >1% sequences in each selection. **C-D**) For the top 3 most prominent clusters, **C**) distribution of the relative fitness of clusters in each successful selection, defined as the fraction of total absolute counts per target-sampling combination attributable to each cluster. **D**) Violin plot of the aspect ratios of clusters in each successful selection, calculated as relative fitness divided by the fraction of unique sequences in the cluster. An aspect ratio near 1 indicates a proportional relationship between a cluster's height and its sequence diversity.

Assessing effects of randomization design on binder topology

To assess how specific design features of the scaffold variation influence binder discovery, we constructed eight additional affibody designs (**Table S5, 16**) based on eight distinct design considerations: different randomized positions (Designs A–C; **Table S1**); different amino acid usage (Designs B, D–F); and with core hydrophobic mutations (Designs G–I) (**Fig. 4A, Tables S1, 5**). Additionally, we constructed a fully randomized 30-mer NNY design to serve as an undefined scaffolding structure reference (Design J; ~2×10^{35} theoretical variants, **Fig. S16**), benchmarking the baseline density of binder variants from complete randomness. A 10^8 sampling of each design was screened across the 96-target panel (**Fig. 4B, Fig. S17**), revealing substantial

variation in hit counts (e.g., 44 in the sampling of Design B vs. 1 in the sampling of Design E) and hit patterns (18 targets hit in ≥5 libraries; 10 targets in only 1). Targets with higher binder density with Design A more consistently yielded hits across alternative designs, suggesting these represent intrinsically facile targets (**Fig. 4C**).

We compared how varying the randomized positions influenced binder discovery. Design B features mutations targeting the surface-exposed residues on one face of the canonical three-helix affibody bundle, while Design A includes additional positions randomized with a restricted codon set (informed by prior yeast display work across seven targets) (*53*). Designs A and B showed near-complete overlap in targets hit, suggesting that the modified randomizations in Design A had limited impact on binder density for most targets (**Fig. S18**). Exceptions included IBTK, VHL, and the OLIG2 DNA-binding domain, which hit with the sampling of Design A but not Design B (>10^{-8} density in Design A), and NRAS, G3BP, and PP2A-B', which hit in the sampling of Design B but not Design A. Since these hits likely lie near the binder density threshold, these differences may reflect stochastic variation rather than systematic effects. Of the 41 shared hits, 8 had Design A binders with 50% or less activity of Design B binders, while 12 showed 50% higher activity from Design A; the remainder were comparable (**Fig. S19**). Overall, the additional randomization in Design A offered only marginal differences relative to the canonical Design B (**Fig. 4D, S18, 19**). In contrast, comparing Designs B and C, which differ in the number of helices randomized, showed that randomizing two helices is strongly beneficial: the sampling of Design B produced 44 hits, versus 18 for Design C (**Fig. 4B, 4D, S18, 19**). Only two targets, CD3d and GATA2 DNA-binding domain, produced binders exclusively with the sampling of Design C. Among the 15 shared hits, Design C binders outperformed Design B binders by >50% in six cases, while Design B outperformed Design C by >50% in just two cases (**Fig. S18**). Thus, while single-helix randomization reduces binder density, likely by restricting possible binding motifs, binders from Design C often show equal or better affinity suggesting that Design C has improved sampling of single helix binding motifs (i.e., **Fig. 2I** PP2A-B''). Therefore, two-helix randomization generally outperforms one-helix designs by accessing more possible binding motifs but at the cost of reduced sampling of single helix motifs.

Next, we investigated how amino acid diversity at canonically randomized positions affects binder discovery. Design D uses NNK codons (all 20 amino acids + stop), while Design B uses NNY (no stop, Glu, Lys, Met, Gln, or Trp). Designs E (VBY codons, which represent postulated primordial amino acids (*57*)) and F (KMY codons, encoding Tyr, Ala, Ser, and Asp to test Tyr's privileged role in PPIs (*58*)) include further restricted sets (**Fig. 4A**). We found that reduced amino acid diversity decreased the number of targets with binder densities >10^{-8} (**Fig. 4B**) and yielded

lower-fitness binders when present (**Fig. 4E**). Samplings of Designs B and D (NNY and NNK codons) performed comparably (42 and 44 hits, respectively, with similar fitness), while the samplings of Designs E and F (VBY and KMY codons) performed poorly, producing only one and six low-fitness hits, respectively **(Fig. 4E).** In every case where the samplings of Designs E and F produced a hit, the samplings of Designs B and D did as well. A key exception was a fibronectin binder from Design F (KMY codons), which had affinity comparable to those from more diverse libraries (**Fig. 4B**). Tyr does appear to play a beneficial role: while the sampling of Design E (VBY codons), which lacks aromatic or acidic residues, yielded just one hit, the sampling of Design F (KMY codons), which included Tyr but no other hydrophobic side chains, had six hits. Among overlapping targets hit with samplings of Designs B and D, just three Design B hits showed >50% higher fitness than their Design D counterparts, whereas nine Design D hits outperformed Design B by >50% (**Fig. S19**). Notably, two targets hit only with Design D: MAX and CRBN. Their binders contained Trp residues in the randomized region, along with Glu and Lys (these three residues are not encoded by NNY). These data suggest that amino acid diversity significantly influences both binder density and fitness. Nevertheless, when all amino acid categories are represented, additional chemical redundancy (e.g., Lys vs. Arg, Glu vs. Asp) offers only marginal gains, likely due to the conflicting balance between additional possible motifs and dilution of all motifs.

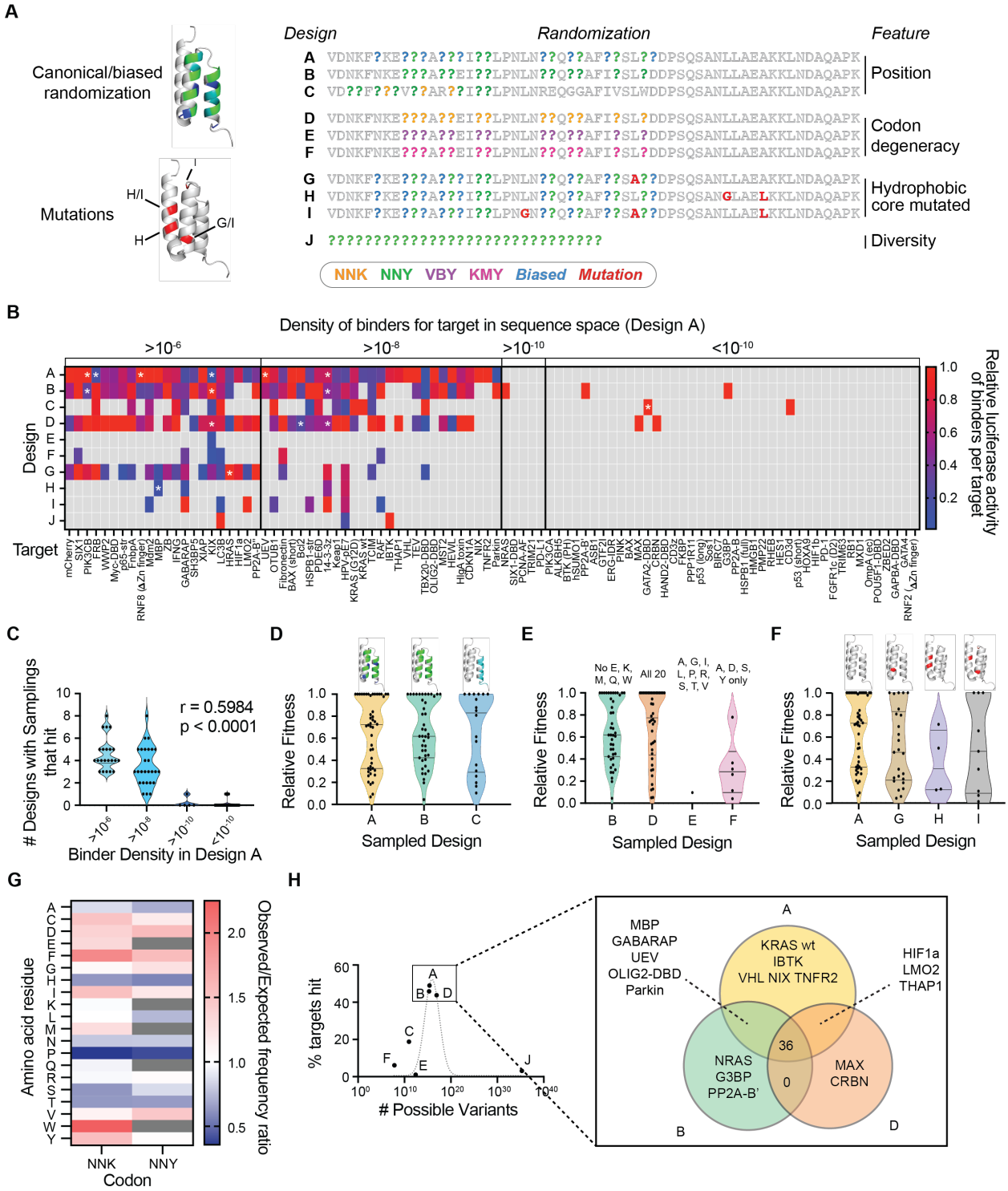

Figure 4: **Elucidating the impact of randomization design on target-affibody binder density**. **A**) Designs (**Table S1, Fig. S16**) for nine affibodies (Designs A-I) and a tenth fully random NNY 30-mer (Design J) to assess the relative importance of randomization position, codon diversity, and scaffold stability. **B**) Heat map indicating the relative binding affinity across all binders for a given target (based on the *E. coli* RNAP complementation luciferase assay) normalized to the

highest affinity binder. Boxes with asterisks indicate selections where the variant stochastically was was less than three-fold as abundant as the top variant. **C**) Violin plot comparing binder density category and number of designs that hit with a 10^8 sampling (two-tailed Pearson correlation, 95% C.I. 0.3972 to 0.7013, $p < 0.0001$). **D-F**) Violin plots showing the relative fitness (directly measured in the binding luciferase assay) of all hits discovered in indicated sampled design **G**) Position-independent amino acid usage for degenerate codons NNK and NNY, calculated from all binder sequences across all designs containing NNK or NNY codons. The frequencies are expressed as an aggregated ratio between observed and expected frequencies, across all positions containing an NNK or NNY codon. **H**) Scatterplot of sequence space diversity versus target hit rate. A dotted gray line shows a lognormal-like relationship between diversity and hit rate, included as a visual guide only due to limited (7) data points. Venn diagram of the targets hit by the top three-performing Designs A, B, and D, with targets where not all three designs hit listed.

Finally, in Designs with mutated hydrophobic cores, which varies the average expression level (a proxy of stability) of randomly isolated variants (**Fig. S20**), we observed a precipitous loss of binder density and quality (**Fig. 4F**). As a reference, the most mutated hydrophobic core scaffold designs (Designs H and I) performed almost as poorly as the completely NNY-randomized 30-mer Design J, which has no scaffolding to structure potential binders (**Fig. 4A, 4B**). However, we noted several interesting exceptions where binders originating from Designs G or I exhibit affinities rivaling those without mutated hydrophobic cores: GABARAP, PP2A-B'', HPV-pE7, TBX20-DBD, and HRAS (**Fig. 4B, S19**). Therefore, as expected, stability and structure play an absolutely crucial role in binder density (*37, 59*), and even a few core mutations can negate the effects of binder density increases imparted by a scaffold protein.

The simplest randomization design is nearly optimal

Previous efforts to improve affibody design strategies (Designs A vs. D) and reduce codon redundancy (Designs B vs. D) yielded, at best, marginal improvements in binder density and fitness. Analyzing amino acid usage from hits in NNK and NNY-randomized codons across binders for all targets (**Figs. 4G, S21, 22**), we found that every amino acid was used. Notably, Trp (NNK), Phe (NNK), and Tyr (NNY) were enriched by 125%, 95%, and 54%, respectively, compared to expected frequencies, while Pro was consistently depleted by 60–65% (NNK and NNY). Surprisingly, His was depleted ~40%, whereas Gly, despite being potentially destabilizing to α-helices, was not. When grouped by physicochemical categories (hydrophobic, charged, aromatic, small polar/nonpolar), usage remained similar between NNK and NNY libraries. These results are consistent with the finding that reducing diversity from NNK to NNY provided minimal benefit, but that further restrictions (e.g., VBY- or KMY-encoded libraries) severely reduced binder density (**Fig. 4E**). When comparing size of the theoretically possible variants to hit rates from 10^8

variant samplings across the 96-target panel (excluding core-mutated libraries), hit rates increased with more randomized positions and codon diversity but plateaued around 51%, with no gain beyond a diversity of ~10^{15} (**Fig. 4H**). Notably, Designs E and F were sampled nearly to completion yet yielded poor hit rates, suggesting that their limited performance stems from lack of binding motifs within the set of possible variants. In contrast, Design J, the fully random 30-mer, was sampled at only ~10^{8} variants out of a theoretical ~10^{35} (i.e., <10^{-25}% of the total space) and produced three hits reflecting the low sampling (as binding motifs likely are contained in the set of possible variants for nearly all targets). These results indicate the existence of both an optimal diversity window and a practical ceiling for scaffold randomization design. We hypothesize that two forces drive this optimization: increasing diversity at the position and codon levels increases the possible binding motifs in the design; however, additional randomizations or redundancies in amino acid types dilute each individual binding motif. Therefore, while modifying randomization positions or codons is unlikely to enhance binder density, we anticipate that mutations that improve scaffold stability could lead to higher binder densities and improved fitness.

Biochemical characteristics of targets are not predictive of binder density

We next asked whether broad target classifications correlate with binder density categories. The bulk of our 96-target panel can be grouped by structural and functional features: kinases/phosphatases, E3 ligases, transcription factors, extracellular membrane domains, intrinsically disordered proteins (>50%), and targets with predominant α-helical or β-sheet content (>30%) (**Table S6**). Median binder density did not differ significantly across classes (Kruskal-Wallis H = 10.81, df = 7, p = 0.147) (**Fig. S23**), though individual targets in generally difficult categories, such as Mdm2 and RNF8 (E3 ligases) and SIX1, LMO2, and p65 (transcription factors), achieved high binder density (>10^{-6}), suggesting intrinsic target-specific properties outweigh class-level trends.

To explore possible determinants of binder density, we ranked difficult-to-bind targets by the number of designs yielding binders: *hard* targets only hit with large libraries (e.g., 10^{10}-sized Design A) or single instances with our varied Designs at 10^{8} sampling size; *scarce* targets had no hits across any sampling (**Table S7**; ordering in **Fig. 3A, 4B**). Using this ranking, we tested correlations between binder difficulty and target properties. Expression level in PANCS (**Fig. S24**) (Pearson's correlation, r = –0.2775, p = 0.0241) and percent disorder (r = 0.2439, p = 0.0485) showed modest significance; other features, including length, solvent-accessible surface area, secondary structure content, and AlphaFold pTM or ipTM scores, were not significantly correlated

with binder difficulty (**Fig. S25, Table S8**). Solvent-exposed residue distributions were also not different across difficulty groups (**Fig. S26**). A multiple linear regression incorporating all variables explained only 23% of the variance in binder difficulty rank ($R^2 = 0.2301$, $p = 0.0270$), with disorder ($\beta = 0.285$, $p = 0.0087$) and expression ($\beta = -9.7\times10^{-6}$, $p = 0.0402$) as the only significant contributors (**Table S9**). Similarly, Principal Component Analysis (PCA) of all eight variables failed to resolve clusters by binder density (**Fig. S27**). Thus, while disorder and expression are mildly associated with target tractability, standard biophysical and structural parameters are not, pointing to the likely influence of additional, untested features, such as scaffold-target shape complementarity.

Contrastive representation learning enhances binder identification while revealing generalizable interaction landscapes.

Given the observed experimental reproducibility in PPI landscapes, we aimed to see if machine learning trained on the PANCS-Binders data could be used to generalize and further populate the landscapes. Critically, our data provides both positive (binders) and negative (non-binders) examples for each target protein (**Fig. 5A**), in contrast to natural PPIs which represent only positive examples. As a first step toward understanding how PANCS-Binder data can enable computation and further characterizing the fitness landscape, we used a simple approach to classify binders and non-binders. Namely, we embedded sequences with ESM-2-650M (*63*) and trained a neural network by minimizing a triplet contrastive loss with margin (*61, 62*) so as to position binders proximal and non-binders distal to their respective target proteins in a shared low-dimensional (latent) space (**Fig. 5B, 5C**); the target and binder embeddings were subsequently concatenated and projected for classification (Methods). The training (80%) and validation (20%) data were drawn from for the selections with Designs A-J (10^8 sampling size) for each of the 96 targets (58 of which have hits in at least one sampling) in **Fig. 4B**. The area under the validation receiver operating characteristic (ROC) curve was well above that for random guessing and above a procedure that classified binders and non-binders based on the ESM-2-650M embeddings without the triplet contrastive loss (**Fig. 5D** and **Fig. S28**), which suggests that the triplet contrastive loss yields a meaningful latent space.

We then used the data for the 9 target panel in Figure 2, which contains 24 independent samplings (six replicate samplings of 10^5, 10^6, 10^7, and 10^8 unique variants) from Design A as a test set. We quantified discrimination between binders and non-binders by measuring the cosine distances in the latent space between target proteins and their corresponding binders and non-binders (**Fig. 5E, S29**). Permutation tests on the distance distributions confirmed that distance

distribution means (i.e., differences between target-binder and target-non-binder distances on held-out test sequences) were statistically different across all targets ($p < 0.0001$ for IFNG, LC3B, KIX, XIAP, and PP2A-B"; $p < 0.001$ for CDKN1A). Interestingly, we note that model performance is strongly correlated with target difficulty, as defined by the number of hits obtained across 10^8 unique variant samplings (**Fig. S30**). Because the test data are from samplings that are independent of those in the training and validation data, these results show that the model can identify patterns that are consistent for a design. These patterns can be visualized by projecting the latent space (**Fig. 5F and Fig. S31**). The results are consistent with but improve on those in **Fig. 2I** in that now there is a meaningful notion of distance to the target, which in principle could be exploited to search for new binder sequences.

We further characterized the model's ability to generalize by plotting the mean distance as a function of the minimum number of edits separating the training and test sequences (**Fig. 5G**). Almost all the test sequences are at least 5 edits from the training sequences, and the model can predict binders that are up to 9 edits from the training sequences (reflecting the ability of the model to identify the minimal binding motif while discarding the irrelevant positions - those that are not converged in **Fig. 2I**). To assess the minimum data requirements for achieving significant separation in the latent space and reliable classification of binders and non-binders, we systematically subsampled the training data, maintaining balanced representation of targets and again trained the model, following the same procedure. To ensure consistency when comparing different models, the minimum edit distance is defined with respect to the entire training set. There is a striking transition at 50% of the data in the mean distance between binders and non-binders for all edit numbers (a similar trend was observed for AUROCs, **Fig. S32**). This transition reinforces the idea that large sampling sizes (to produce multiple sequences with a shared binding motif) are needed to meaningfully characterize the fitness landscape.

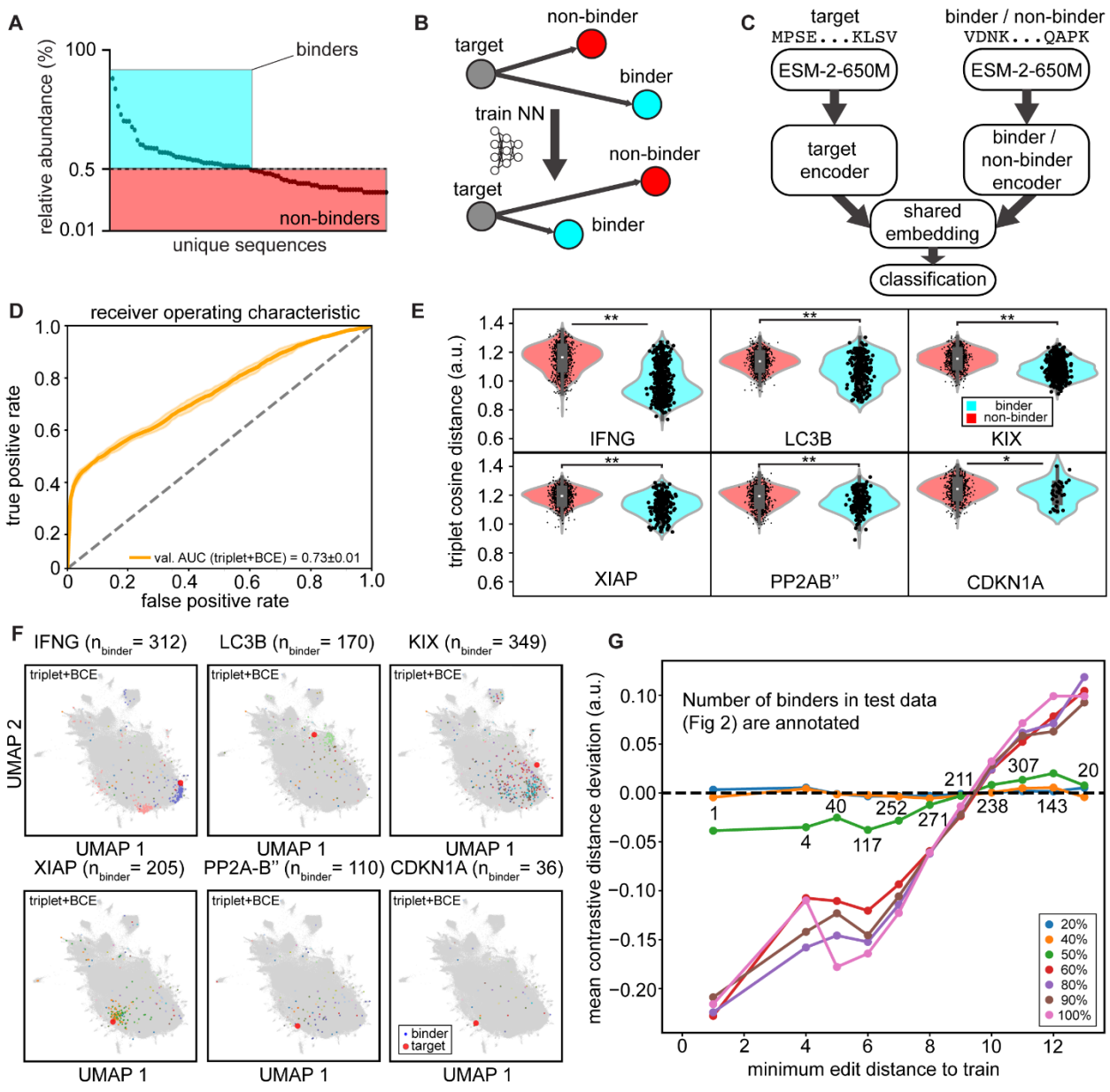

Figure 5: Supervised contrastive learning via PANCS data facilitates binder identification and discovers a shared latent space between targets and binders. **A**) NGS data provide relative abundances for sequences that bind the target protein. Binders (cyan) and non-binders (red) are classified using a 0.5% relative abundance threshold (number of NGS reads per variant divided by total reads from one selection round). **B**) Triplet contrastive loss is employed to co-embed target protein sequences with binders, positioning targets and binders proximally in embedding space while separating targets from non-binders. **C**) Target and binder/non-binder sequences are converted to numerical representations using ESM-2-650M. Target and binder/non-binder encoders are trained on PANCS data, generating a shared latent space subsequently used for classification tasks. **D**) Receiver operating characteristic (ROC) curves for models trained on PANCS data. Validation performance is shown for triplet + binary cross-entropy (BCE) loss. Area under the curve (AUC) values for the validation data are shown inset. Shading corresponds to the standard deviations across a 5-fold split. **E**) Cosine distances between targets and binders/non-binders in triplet models for six held-out test targets (from **Fig. 2**): IFNG, LC3B,

KIX, XIAP, PP2A-B'', and CDKN1A. Violin plots show distribution separation between binders (cyan) and non-binders (red). Individual data points are displayed with x-axis jitter for visualization. Permutation test results comparing distribution means are indicated (**: $p < 0.0001$; *: $p < 0.001$). **F**) Two-dimensional UMAP projections of latent space co-embeddings of triplet + BCE models. Test data (from **Fig. 2**) show targets (large red circles) co-localized with corresponding binders (small colored circles). Target labels indicate binder counts, with binders colored by cluster membership. All data points are shown in gray. **G**) Mean contrastive distance deviation (defined as binder distance−non-binder distance) for test data at varying minimum edit distances to the nearest training sequence. Models trained on subsets of training data (percentages colored) are shown. The number of test sequences are numbered on the plot. In all cases, minimum edit distances to train are defined over the full training set to enable comparison.

Discussion

In this work, we addressed two perennial questions concerning the emergence of *de novo* PPIs. First, in terms of how likely a random variant of one protein is to interact with another random target, we quantified that number across a range of targets for interaction with a three-helix bundle and found that it can vary between 1 in 10,000 to over 1 in 10,000,000,000, but at the median across targets, binding occurs at a frequency of approximately 1 in 100,000,000. These values correspond to minimal binding motifs of 2-8 specified amino acids within the randomization design. Second, because this likelihood varies over 5 orders of magnitude, we address the relative dependence of this value on the features of how the binder is randomized versus the features of the target. We found that target-specific effects drive the probability of finding a binder: targets maintained similar likelihood across multiple randomization designs, indicating that target-intrinsic features are the primary drivers of success in selections more broadly – meaning the topology of the PPI is driven more so by the target than the binder. However, we were unable to pinpoint specific features of targets that significantly impact the likelihood of the topology, indicating "bindability" of a target with a given scaffold is more complex than easily quantifiable biophysical features. These results have implications in natural evolution. In natural evolution, proteins must navigate sequence space though sequential mutations (co-evolution) or via high diversity mutations (somatic hypermutation). Our results suggest that, for the evolution of a *de novo* interaction PPI, certain proteins with simple, 2-4 amino acid binding motifs can, through a limited number of mutations, more readily acquire additional novel binders. In other words, these proteins appear poised for evolving new interactions. This is buoyed by the observation that the "easy" targets in our classification, such as KIX and LC3B, are generally known to bind many proteins in nature, indicating an inherent "bindability" feature (requiring just 2-3 amino acids). "Hard" targets likely contain more complicated motifs (6+ amino acids or strict shape constraints). While hard to predict, this "bindability" can be experimentally determined using our pipeline. Finally, that randomization to include all 20 amino acids was nearly optimal with designs with

biased randomizations having only marginal improvement; that the simplest randomization designs tended to produce the best outcomes points to the optimization of natural proteins and codons more generally for evolvability.

The statistical likelihood of a given protein binding a target and broader topology of the scaffold-target pair also has a profound impact on binder discovery. First, different methods have different maximum sampling sizes (10^{15} for ribosome display vs. 10^3 for ELISA based screening). If the binder density and method are mis-matched, then the selection will fail. Indeed, it is rarely known how “difficult” (dense) a target is at the outset of most discovery campaigns. Second, as binder density is dependent on binder motif frequency, randomization designs should be performed with this in mind. For example, nanobody libraries often include 16+ randomizations across 2-3 loops; however, we predict two designs would likely be superior: the CD3 loop with 8 NNK codons with flanking flexible residues in the loop (~10^9 variants) allowing for up to 7 specified amino acids for single loop binding motifs and a separate design where each 10-12 NNK codons are separated across two or three loops for multi loop binding motifs. Both designs offer different potential binding motifs, analogous to those allowed in the 16+ randomized designs, but with higher motif frequency for each motif compared to the single high diversity design. Third, the topological characteristics of the landscape can also play a critical role. In display and sorting based selections with threshold-based sorting functions, selections with only low aspect ratio clusters will have outputs dominated with weak binding variants simply because these types of variants greatly outnumber the high affinity variants. However, if there are only a few sequences per cluster because the sampling is near the binder density or if targets have a high aspect ratio cluster, then display/sorting selections should have reduced noise and perform better. Fourth clustering of selection outcomes provides critical evidence of the evolvability of variants for affinity maturation. Because variants amplify in an affinity-dependent manner in PANCS-Binders, the differences in reads (population) correlate with differences in affinity. High aspect ratio clusters have only lightly sampled binding motifs, but the few variants we have identified demonstrated high affinity; therefore, these peaks would be candidates for optimization because the sequence space appears both fruitful and underexplored. Additionally, large clusters, which have significant sampling, can be used to estimate the likely ruggedness of the binding motif to inform choice of affinity maturation strategy (i.e., epPCR or PACE which can move only a couple mutations at a time making rugged landscapes challenging vs selection of positions for randomization in second generation libraries which can move across a greater range of sequence space to further map rugged landscapes). While we have focused here on *de novo* binder discovery using the affibody scaffold, we suspect that motifs that have been broadly adapted by natural evolution for binding

to a wide range of proteins (i.e., antibodies, WD40 motifs, C-type lectins, etc.) have likely been selected because they have (or evolved to have) both higher average binder density and a selection strategy matched to the types of landscapes involved.

Recently, *in silico de novo* protein binder design has emerged as a powerful approach for binder discovery. Prominent state-of-art approaches primarily rely on structure-based approaches such as co-folding via diffusion models (*60*) and sequence optimization via AlphaFold2 (*22*). Such state-of-art open-source models are primarily trained on PDB-derived structures, thus, the incorporation of more PPI data has the potential to further accelerate progress in this space (*61*). PANCS-Binders holds unique potential to produce high-quality training data to enable various machine learning applications due to its ability to efficiently identify binders from diverse customizable phage libraries with billions of variants against a wide array of proteins. Data from PANCS-Binders here reveal the rank-order of variant fitness for tens-to-hundreds of binding variants and extensive sets of identified non-binders to facilitate supervised contrastive learning, which as demonstrated here, can lead to superior performance relative to non-contrastive approaches. Our approach further produces an interpretable latent space which can in principle be adapted to serve as an encoder to generate conditional embeddings for generative diffusion models (*62, 63*), enabling *in silico* generation that respects the underlying PPI landscape. Finally, binder density has profound impacts for how we should evaluate *de novo* design strategies: if thousands or more of designed variants are required to be tested experimentally to find one that works (*20*), then designs must be evaluated in comparison to a similar size of random variants. We argue for robust benchmarking by evaluating the difficulty of a target using binder density with traditional scaffolds as well as the evaluation of designs from the top and bottom design quartiles as a necessary first order approximation of whether a design strategy is efficacious or whether a target is simply easy.

SUMMARY

We propose a novel framework for characterizing PPI fitness landscapes as applied *to de novo* binder discovery and use PANCS-Binders selection results to map the topologies of these highly diverse landscapes. Based on replicate samplings of varying sizes, we mapped the fitness landscapes for several targets in detail (IFNG, LC3B, KIX, XIAP) demonstrating the highly reproducible nature of the fitness landscape. Analogous experiments for dozens of targets indicate that 51% of targets have a density $>10^{-8}$ and 5 affibody binding motifs (average number of prominent clusters); the aspect ratios of almost all clusters were <5, suggesting that high affinity binders come from optimal variants containing a common binding motif rather than from rare high

affinity binding motifs. By systematically varying key features of affibody randomization design, we demonstrate a fundamental trade-off: designs with too little diversity risk excluding binding motifs, while those with excessive diversity dilute binding motifs, suggesting limited routes for optimizing designs for the affibody scaffold. Additionally, while target identity has dramatic impacts on binder density ($>10^{-5}$ to $<10^{-10}$) and topology (3-7 prominent clusters), we were unable to identify target characteristics capable of making a predictive model for target difficulty. Although these conclusions are grounded in data only from the affibody scaffold, we expect that other scaffolds are likely to behave similarly; however, the extent to which target difficulty is a cross-scaffold feature is currently unknown. Thus, although we predict similar general topological features, we do not expect other scaffolds will necessarily behave similarly with the target panel used here. Our findings are of consequence for improving *de novo*, experimental binder discovery, in benchmarking and improving *in silico* binder design strategies, and for understanding the natural evolution of *de novo* PPIs.

Methods

Cloning Plasmids

All plasmids (**Table S10**) were generated with Gibson Assembly (GA) of PCR fragments made with Q5 DNA polymerase (NEB) following the manufacturer's protocol. Primers and target inserts were ordered from Integrated DNA Technologies (**Table S11**). To make plasmids, GA reactions were transformed into DH10β chemically competent *E. coli* with the following protocol: Incubation on ice for 30 min, heat shock at 42°C for 45s, 2 min recovery on ice, and addition of 2xYT media and outgrowth with agitation at 37°C for 1h prior to plating on LB agar supplemented with kanamycin (40µg/mL in water), chloramphenicol (33µg/mL in EtOH), and carbenicillin (100µg/mL in water) as appropriate, and overnight incubation at 37°C. Correct insertion was confirmed via Sanger sequencing.

Generating Bacterial Strains

To create multi-transformed bacterial strains for PANCS (+/-AP) and the split T7 RNAP *E. coli* luciferase assay (2-22/C-term-target/N-term-binder), S1030 *E. coli* strains were made chemically competent by preparing a 1/100 subculture of cells at OD600 ~0.3, and pelleting and washing with cold $CaCl_2$- glycerol buffer (60mM $CaCl_2$, 10mM HEPES pH 7.0, 15% glycerol) twice prior to resuspension in the same buffer. Plasmids were transformed one at a time and transformed into single- or double-transformant chemically competent S1030 *E. coli* cells.

Plaque Assays

To perform activity-independent plaque assays to measure titer, 2µL of phage was added to 100µL of S1030-1059 *E. coli* cells at OD600 ~0.3, and a 50X serial dilution of each phage stock was performed by adding 2µL of the previous phage-*E. coli* mixture to the next, for a total of four dilutions. The mixture was combined with 750µL of 50°C top agar (7 g/L agar, 25 g/L LB) and plated evenly onto a four-quadrant bottom agar (15g/L agar, 15g/L LB) plate. After overnight

incubation at 37°C, plaques were counted in the quadrant containing 10-200 plaques to determine the phage titer. The phage titer was calculated with the following equation:

$$Titer\left(\frac{PFU}{mL}\right) = \frac{(Dilution\ factor\ of\ stock\ phage)(\#\ plaques)(50)^{(quadrant\ \#\ -1)}}{0.002mL\ stock\ phage}$$

To perform activity-dependent plaque assays, the S1030-1059 strains were swapped for S1030 double-transformed +/- AP strains, consisting of the desired and a different target strain to confirm the presence or absence of on/off-target activity. In addition, to confirm the fidelity of the selections, plaque assays were performed with naïve S1030 strains to verify the absence of cheater phage in the absence of plaques.

Production of Randomization Design Samplings

Affibody designs followed those of prior studies (*53, 57, 58, 64, 65*) and rational design. PCR conditions using Q5 DNA polymerase (NEB) were optimized for each design such that approximately 20-100µg product was formed at large scale, during which randomization schemes were introduced into an affibody template phage via primers containing degenerate codons (IDT) (**Tables S10, S11**). Following cleanup using the Wizard kit (Promega), restriction enzyme digestion with DpnI and NheI-NF (NEB), gel extraction and cleanup with the Zymo DNA clean and concentrate (Zymo DCC-5) kit, the digestion products were treated with T4 DNA ligase overnight. Ligated products were purified and concentrated with the Zymo DCC-5 kit prior to transformation via electroporation into S1030-1059 cells. The cells were added to 50mL SOC media warmed to 37°C and incubated with shaking for 2h, during which timepoints were collected to assess the titer and sampling size by plaque assay. At the end of the outgrowth, the cells were pelleted, and the supernatant was stored. To amplify the libraries, a 450mL S1030-1059 culture was grown to OD600 ~0.6, and the phage supernatant was added. Following incubation for 6h, the cells were pelleted and the supernatant sterile-filtered and stored at 4°C. All libraries were subject to PANCS with naïve S1030 cells and a no-fusion negative control to ensure that libraries did not contain neither gIII nor N-term wild-type cheaters.

Replication Assay

To determine the expression levels of new targets explored in this chapter in the PANCS system by measuring the phage replication rate, phage encoding the wild-type N-terminus of T7 RNAP were prepared to a diluted stock of 500 PFU/µL. 1000µL of phage were added to 1mL subcultures of S1030 transformed with each +AP in stationary phase, and incubated for 12h at 37°C with agitation. A no-culture (i.e. media only) control was included as a negative control. The cells were then pelleted, and the titer of supernatant was measured via plaque assay or qPCR. To calculate the replication rate of each target in the PANCS system, the titer of each output was divided by that of the media-only negative control.

PANCS

PANCS was performed in deep 96-well plates in experiments where 10^8 or fewer unique variants were tested. 1/10 subcultures of S1030 *E. coli* cells double-transformed with +/- APs

were grown to OD600 ~0.5, then seeded with the number of unique variants of interest (10^5 -10^8 PFU) in the first passage to obtain a final volume of 1mL per well and incubated with shaking for 8-12h. To continue the PANCS for a total of 6 passages, 5% of phage supernatant (50µL of 1mL) were passaged to the next round of selection following the same procedure.

To create our 10^{10} unique variant sampling of Design A, eight separate transformations (of ~1-2*10^9 unique variants each) were combined to generate a composite, 10^{10} sampling of Design A and used to inoculate 500mL cultures of +/-AP S1030 strains, as described before (*16*). Subsequent passages were performed in progressively smaller volumes: 100mL (P2), 15mL (P3), and 5mL (P4+), while maintaining both a 1:1 phage-to-cell ratio and 5% passaging rate.

qPCR

qPCR was utilized to determine the endpoint titer in a high-throughput manner following each PANCS experiment. The qPCR was optimized by testing various sets of primers that recognized different sites in the phage genome and determining the linear range and accuracy of the assay with the different primers and dilutions of phage. The primers VC-525 and VC-526 were selected to amplify a small section of the M13 phage gIII for assay generalizability to different libraries (**Table S11**). 20µL qPCR reactions were assembled in the following way: 3µL water, 1µL each primer at 5µM, 10µL Power Up SYBR Mix (ThermoFisher, Cat#: A25777), and 5µL PANCS endpoint phage diluted 200X. The qPCR reactions were run on a QuantStudio6Pro with the following protocol: 95ºC 10m denaturation, 40 cycles of 95ºC 20s, 60ºC 20s, and 72ºC 20s, final elongation 72ºC 20s, then a melt curve protocol of 95ºC 10s, 65ºC 1m, and 97ºC 1s to ensure that only one amplicon exists. Each qPCR run included a standard curve of phage with known titers confirmed by plaque assay, later used to calculate the endpoint titers of each selection.

Next Generation Sequencing (NGS)

NGS samples were prepared by appending adaptor and barcode sequences to PCR amplicons starting from the linker sequence and ending just downstream of the stop codon positioned following the affibody scaffold, using primers MS-1144-MS-1156 and MS-1346-1357 (**Table S11**). The PCR protocol consisted of a 10 min initial denaturation step at 98ºC, and 30 cycles of 15 s 98ºC denaturation, 15 s 63 ºC annealing, and 1 min 72 ºC extension steps, and a 5 min final 72 ºC extension step. The PCR products were pooled with other barcodes within the same NGS sample and purified with the Zymo DCC-5 kit prior to measuring their concentrations with the Qubit dsDNA High Sensitivity kit (ThermoFisher). Each NGS sample was diluted to 20ng/µL and 500ng total barcoded PCR products were submitted for analysis with the Amplicon EZ service at Genewiz (Azenta). Sequencing data was analyzed by BB Merge to merge all paired end reads, then MATLAB was used to separate the results for each selection by barcode (Step1) and translate the sequences (Step2) (*66, 67*).

Split-T7 RNAP *E. coli* Luciferase Assays

We followed the split-RNAP *E. coli* luciferase assay protocol detailed in our previous publication (*16*). In summary, a double-transformant S1030 strain (2-22/C-term-target lux) was made chemically competent as described above and transformed with N-term binder and non-binder plasmids (**Tables S9, S10**). Colonies for each condition were picked and grown in 1mL of LB media with kanamycin, chloramphenicol, and carbenicillin overnight at 37ºC with shaking. The

next day, 1/20 subcultures for each colony were prepared in LB media supplemented with antibiotics and L-arabinose (2 mg/mL) and cultured for 3.5 h before measuring OD600 and luminescence signals using a BioTek Synergy Neo2 plate reader. To compare the activities of binders and non-binders, we computed the quotient of luminescence signal and OD600 and normalized all binder signals to those of the non-binder signals.

Clustering and Fitness Analysis

For each collection of binders (data presented in **Fig. 2** and **Fig. 3** individually), sequences were one-hot encoded in the variable positions; PCA was performed on the one-hot encoded representation and the top 100 PCs were used to compute the 2-dimensional t-SNE embedding. We computed the associated Hamming distance matrix and performed hierarchical clustering with Ward linkages. For Fig. 2 data, we used 30 clusters, and 100 clusters for Fig. 3. Furthermore, using the Hamming distance and the top 100 PCs resulted in identical clusters.

From the experimentally measured fold-change of top variants for each target in each sampling, we extrapolated the fitness of other variants identified in the NGS of that selection using the ratios of relative concentrations linearly:

$$f_{i,k} = f_{top,\,k}\frac{C_{i,\,k}}{C_{top,\,k}}$$

Where and $C_{top,\,k}$ are the relative concentrations and of variant i and the top variant; $f_{i,\,k}$and $f_{top,\,k}$are the extrapolated fitness for variant i and the measured fitness for the top variant. The total fitness is the sum of fitness for binders within a cluster and the normalized fitness is normalized over the number of binders.

Amino acid distribution analysis

We followed a two-step approach to analyze the position-dependent amino acid usage at randomized positions in the affibodies using the Python libraries pandas, numpy, itertools, and Bio. First, the expected frequency of each amino acid for a given codon position was computed based on the design strategy. For each codon position, we considered all possible codons that could be generated from the nucleotides at each position, and the expected percent frequency of each amino acid was calculated by determining the fraction of codons that encode each amino acid. Observed percent frequencies were determined from the NGS binder sequences as reported in Step2 analysis, where we considered only sequences >1% abundance in their respective selections. The percent difference was calculated for each amino acid at every degenerate codon position using the formula:

$$Equation\ 4, Percent\ difference = \frac{Observed\ frequency - \ Expected\ frequency}{Expected\ frequency}\ x\ 100$$

To assess the usage bias of each amino acid across the degenerate codons NNK and NNY, we identified positions in our designs that contained NNK or NNY and aggregated the percent differences of observed/expected amino acid frequencies across all binder sequences within a given design and across all designs.

Statistical Analysis between Target Properties and Binder Density

The statistical analyses: Pearson's and Spearman's correlations, Kruskal-Wallis test for multi-group target comparisons, Dunn's post-hoc pairwise tests, log-normal curve fitting, and multiple linear regression were all performed in GraphPad Prism 10 to assess meaningful relationships between target properties and binder density, analyze class-wide influences on target difficulty, and understand the relationship between sequence space diversity and hit rate, respectively. To process the split-T7 RNAP *E. coli* luciferase assay and qPCR data, Microsoft Excel was used to calculate fold changes in luminescence/OD_{600} over negative control signal and convert C_q values to titers based on a standard curve of phage with known PFU/mL.

Data Preprocessing for Machine Learning

NGS data from PANCS provided affibody binder sequences for each target protein studied. For each target, all sequences returned from NGS were analyzed, with read counts aggregated for identical proteins encoded by different genes. Quality control filtering removed sequences that were not exactly 58 amino acids in length or lacked the required affibody framework regions (N-terminal VDNK and C-terminal DPSQSANLLAEAKKLNDAQAPK). The total sequence count for each selection was calculated by summing all remaining filtered sequences. Relative abundance was defined as the number of reads for each unique protein sequence divided by the total sequence count for that selection. Across all targets in all libraries, sequences with relative abundance ≥0.5% were classified as binders, while those below this threshold were designated as non-binders. True non-binders were identified through two approaches: (1) sequences present in pre-selection sampling NGS data but absent from target-specific selections, and (2) sequences that appeared in selections for one target but not others within a sampling. This classification strategy was necessary because the sequenced pool represents only a subset of the theoretical design space, which is too large for exhaustive exploration.

Training Details for Neural Networks

Protein sequences were passed through the pre-trained ESM-2-650M model to obtain an embedding of the size 1280 x L, where L is the number of amino acids in the protein; for each sequence, the embedding was averaged over L to obtain a vector of size 1280 as input for the downstream model. We trained two encoders: one for affibodies and another one for targets. For each encoder, we used 3 hidden layers of a multi-layer perceptron (MLP) of size 640 with rectified linear unit (ReLU) activations and an output dimension of 1280. The newly encoded embeddings of both affibody and target were then concatenated and projected before passing a sigmoid layer for classification.

For the triplet model, our loss function was the sum of equally weighted binary cross-entropy loss and triplet loss with margin as implemented in PyTorch. We used an initial margin of 0.4 and a cosine annealing strategy for updating the margin, allowing a gradual change from 0.4 to 0 every 10 epochs. To enhance separation of binders and nonbinders in the latent space, we used a curriculum learning strategy based on semi-hard triplet mining, which was achieved by computing the distances between the negatives/nonbinders and the target in the shared latent space, enabling the identification of nonbinders that are closest to the anchor. For each epoch,

we used all (anchor/target, positive/binder) pairs and sampled 100 negatives randomly. We used a curriculum of [0, 50, 75] percentiles and updated the curriculum whenever the model stopped improving. We used a constant learning rate of 1e-6 and trained the model for 130 epochs. For the BCE model, the loss function is simply the binary cross-entropy; we used a constant learning rate of 1e-5 and trained the model for 50 epochs. In both cases, we used an Adam optimizer.

Visualization of Shared Embeddings

To visualize the shared latent space, we computed the top 100 PCs from the embeddings of all binders (train, validation, and test) to obtain a 2D UMAP embedding. We then used the same PC projection and UMAP transformation to embed the targets. We used the same cluster identities as established in Fig. 2 for coloring.

Validation on Test Data

To construct the comparison of violin plots, we randomly sampled 1000 negatives for each target. For computing the area under the receiver operating curve (AUROC), we randomly sampled 100 negatives per positive. The sequences were passed through the neural networks with the lowest validation loss from both training strategies. Cosine distances were then computed and displayed on violin plots. For each target, permutation tests with 10000 replicates were performed between the binders and non-binders using the mean as the test statistic.

Acknowledgments

We thank the members of the Dickinson, Vaikuntanathan, and Dinner labs for providing constructive and insightful discussions about this study. We thank the University of Chicago Sanger Sequencing Core and Azenta's Genewiz Amplicon-EZ service for sequencing support. We thank Eddy Pineda and Jingzhou Yang for providing transcription factor and E3 ligase +AP plasmids. We thank Dr. Somayeh Ahmadiantehrani for assistance with manuscript preparation. We thank Prof. Samantha Riesenfeld for providing GPUs for training the neural networks.

Funding:

This work was supported by the National Institute of General Medical Sciences (R35GM119840 to BCD, R35GM147400 to SV, R35GM136381 to ARD) of the National Institutes of Health. MJS was supported by the National Institute of General Medical Sciences (F32GM147968) and SSL was partially supported by an NIH training grant (T32GM144290) and the Eckhardt Graduate Scholarship from Department of Chemistry at the University of Chicago. CFG was supported by the Seymour Goodman Graduate Fellowship from the Department of Chemistry at the University of Chicago. AN was supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences, LLC program.

Author contributions:

Conceptualization: SSL, MJS, CFG, AN, ARD, SV, BCD Funding acquisition: SSL, MJS, CFG, AN, SV, BCD Investigation: SSL, MJS, CFG, AN, JAP, ST Methodology: SSL, MJS, CFG, AN,

CB, JAP, ST Supervision: ARD, SV, BCD, Writing – original draft: SSL, MJS, CFG, AN, BCD Writing – revision, review, & editing: SSL, MJS, CFG, AN, ARD, SV, BCD

Competing interests:
BCD is an inventor on the patent describing the split RNAP biosensors. The University of Chicago has filed a provisional patent on the PANCS-Binders technology with MJS and BCD listed as inventors. BCD is a founder and holds equity in Tornado Bio, Inc. The other authors declare no competing interests.

Data and materials availability:
All data are available in this manuscript or supplementary materials. All physical vectors will be made available on reasonable request.

Supplementary Materials

The PDF file includes:
Figs. S1-32
Tables S1-11

Other Supplementary Material for this manuscript includes the following:
Data S1
Data S2
Data S3
Data S4

References and Notes

1. T. Hunter, Signaling--2000 and beyond. *Cell* **100**, 113-127 (2000).
2. S. K. Miryala, A. Anbarasu, S. Ramaiah, Discerning molecular interactions: A comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* **642**, 84-94 (2018).
3. P. Porras *et al.*, Towards a unified open access dataset of molecular interactions. *Nat Commun* **11**, 6144 (2020).
4. A. Bandrowski, M. Pairish, P. Eckmann, J. Grethe, M. E. Martone, The Antibody Registry: ten years of registering antibodies. *Nucleic Acids Res* **51**, D358-D367 (2023).
5. B. Z. Stanton, E. J. Chory, G. R. Crabtree, Chemically induced proximity in biology and medicine. *Science* **359**, (2018).
6. P. J. Carter, G. A. Lazar, Next generation antibody drugs: pursuit of the 'high-hanging fruit'. *Nat Rev Drug Discov* **17**, 197-223 (2018).
7. L. Ou, M. T. Setegne, J. Elliot, F. Shen, L. M. K. Dassama, Protein-Based Degraders: From Chemical Biology Tools to Neo-Therapeutics. *Chem Rev* **125**, 2120-2183 (2025).
8. F. Pazos, J. A. Ranea, D. Juan, M. J. Sternberg, Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. *J Mol Biol* **352**, 1002-1015 (2005).
9. B. T. Seet, I. Dikic, M. M. Zhou, T. Pawson, Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* **7**, 473-483 (2006).
10. J. M. Di Noia, M. S. Neuberger, Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* **76**, 1-22 (2007).

11. S. Handa *et al.*, RNA control of reverse transcription in a diversity-generating retroelement. *Nature* **638**, 1122-1129 (2025).
12. S. S. Sidhu, H. B. Lowman, B. C. Cunningham, J. A. Wells, Phage display for selection of novel binding peptides. *Methods Enzymol* **328**, 333-363 (2000).
13. V. C. Xie, M. J. Styles, B. C. Dickinson, Methods for the directed evolution of biomolecular interactions. *Trends Biochem Sci* **47**, 403-416 (2022).
14. A. Wellner *et al.*, Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat Chem Biol* **17**, 1057-1064 (2021).
15. D. N. Philpott *et al.*, Rapid On-Cell Selection of High-Performance Human Antibodies. *ACS Cent Sci* **8**, 102-109 (2022).
16. M. J. Styles *et al.*, High-throughput protein binder discovery by rapid in vivo selection. *bioRxiv*, (2025).
17. B. T. Porebski *et al.*, Rapid discovery of high-affinity antibodies via massively parallel sequencing, ribosome display and affinity screening. *Nat Biomed Eng* **8**, 214-232 (2024).
18. A. Baryshev *et al.*, Massively parallel measurement of protein-protein interactions by sequencing using MP3-seq. *Nat Chem Biol* **20**, 1514-1523 (2024).
19. D. Younger, S. Berger, D. Baker, E. Klavins, High-throughput characterization of protein-protein interactions by reprogramming yeast mating. *Proc Natl Acad Sci U S A* **114**, 12166-12171 (2017).
20. L. Cao *et al.*, Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551-560 (2022).
21. N. R. Bennett *et al.*, Atomically accurate de novo design of antibodies with RFdiffusion. *bioRxiv*, (2025).
22. M. Pacesa *et al.*, BindCraft: one-shot design of functional protein binders. *bioRxiv*, (2025).
23. N. Spisak, A. M. Walczak, T. Mora, Learning the heterogeneous hypermutation landscape of immunoglobulins from high-throughput repertoire data. *Nucleic Acids Res* **48**, 10702-10712 (2020).
24. T. C. Yu *et al.*, A biophysical model of viral escape from polyclonal antibodies. *Virus Evol* **8**, veac110 (2022).
25. A. L. Ferguson *et al.*, Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* **38**, 606-617 (2013).
26. A. Acevedo, L. Brodsky, R. Andino, Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686-690 (2014).
27. T. N. Starr *et al.*, Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* **377**, 420-424 (2022).
28. B. B. Larsen *et al.*, Functional and antigenic landscape of the Nipah virus receptor binding protein. *bioRxiv*, (2024).
29. J. Li, Z. Yi, M. C. Laskowski, M. Laskowski, Jr., C. Bailey-Kellogg, Analysis of sequence-reactivity space for protein-protein interactions. *Proteins* **58**, 661-671 (2005).
30. X. Lian *et al.*, Deep-learning-based design of synthetic orthologs of SH3 signaling domains. *Cell Syst* **15**, 725-737 e727 (2024).
31. A. M. Bendel *et al.*, The genetic architecture of protein interaction affinity and specificity. *Nat Commun* **15**, 8868 (2024).
32. J. Zhao, Y. Li, N. Terasaka, H. Aikawa, H. Suga, Diversity Scale of Library Matters: Impact of mRNA Library Diversity Scales on the Discovery of Macrocyclic Peptides Targeting a Protein by the RaPID System. *ACS Cent Sci* **11**, 431-440 (2025).
33. K. Nord *et al.*, Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nat Biotechnol* **15**, 772-777 (1997).
34. P. A. Lofdahl, O. Nord, L. Janzon, P. A. Nygren, Selection of TNF-alpha binding affibody molecules using a beta-lactamase protein fragment complementation assay. *N Biotechnol* **26**, 251-259 (2009).

35. T. Bridge *et al.*, Site-Specific Encoding of Photoactivity in Antibodies Enables Light-Mediated Antibody-Antigen Binding on Live Cells. *Angew Chem Int Ed Engl* **58**, 17986-17993 (2019).
36. X. Wang *et al.*, SYNBIP: synthetic binding proteins for research, diagnosis and therapy. *Nucleic Acids Res* **50**, D560-D570 (2022).
37. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* **103**, 5869-5874 (2006).
38. A. J. Faure *et al.*, The genetic architecture of protein stability. *Nature* **634**, 995-1003 (2024).
39. V. H. Odegard, D. G. Schatz, Targeting of somatic hypermutation. *Nat Rev Immunol* **6**, 573-583 (2006).
40. S. A. McMahon *et al.*, The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat Struct Mol Biol* **12**, 886-892 (2005).
41. I. Fragata, A. Blanckaert, M. A. Dias Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol Evol* **34**, 69-82 (2019).
42. M. Sandhu *et al.*, Computational and Experimental Exploration of Protein Fitness Landscapes: Navigating Smooth and Rugged Terrains. *Biochemistry* **64**, 1673-1684 (2025).
43. D. M. Fowler, S. Fields, Deep mutational scanning: a new style of protein science. *Nat Methods* **11**, 801-807 (2014).
44. A. Papkou, L. Garcia-Pastor, J. A. Escudero, A. Wagner, A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860 (2023).
45. J. Chen, D. R. Woldring, F. Huang, X. Huang, G. W. Wei, Topological deep learning based deep mutational scanning. *Comput Biol Med* **164**, 107258 (2023).
46. J. Zinkus-Boltz, C. DeValk, B. C. Dickinson, A Phage-Assisted Continuous Selection Approach for Deep Mutational Scanning of Protein-Protein Interactions. *ACS Chem Biol* **14**, 2757-2767 (2019).
47. C. J. Markin *et al.*, Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. *Science* **373**, (2021).
48. N. C. Wu *et al.*, Major antigenic site B of human influenza H3N2 viruses has an evolving local fitness landscape. *Nat Commun* **11**, 1233 (2020).
49. H. K. Haddox, A. S. Dingens, S. K. Hilton, J. Overbaugh, J. D. Bloom, Mapping mutational effects along the evolutionary landscape of HIV envelope. *Elife* **7**, (2018).
50. K. S. Sarkisyan *et al.*, Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397-401 (2016).
51. M. S. Subbanna, M. J. Winters, M. Ord, N. E. Davey, P. M. Pryciak, A quantitative intracellular peptide-binding assay reveals recognition determinants and context dependence of short linear motifs. *J Biol Chem* **301**, 108225 (2025).
52. A. Ernst *et al.*, Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst* **6**, 1782-1790 (2010).
53. D. R. Woldring, P. V. Holec, L. A. Stern, Y. Du, B. J. Hackel, A Gradient of Sitewise Diversity Promotes Evolutionary Fitness for Binder Discovery in a Three-Helix Bundle Protein Scaffold. *Biochemistry* **56**, 1656-1671 (2017).
54. J. R. Meyer *et al.*, Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428-432 (2012).
55. V. C. Xie, J. Pu, B. P. Metzger, J. W. Thornton, B. C. Dickinson, Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *Elife* **10**, (2021).
56. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).
57. M. Su, S. J. Roberts, J. D. Sutherland, Initial Amino Acid:Codon Assignments and Strength of Codon:Anticodon Binding. *J Am Chem Soc* **146**, 12857-12863 (2024).

58. F. A. Fellouse, C. Wiesmann, S. S. Sidhu, Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc Natl Acad Sci U S A* **101**, 12467-12472 (2004).
59. M. Nosrati, M. R. Housaindokht, New insights into the effect of mutations on affibody-Fc interaction, a molecular dynamics simulation approach. *J Struct Biol* **215**, 107925 (2023).
60. J. L. Watson *et al.*, De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089-1100 (2023).
61. P. Gainza, R. D. Bunker, S. A. Townson, J. C. Castle, Machine learning to predict de novo protein-protein interactions. *Trends Biotechnol*, (2025).
62. J. H. a. T. Salimans, Classifier-Free Diffusion Guidance. *arXiv*, (2022).
63. A. R. a. P. D. a. A. N. a. C. C. a. M. Chen, Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv*, (2022).
64. E. Gunneriusson, K. Nord, M. Uhlen, P. Nygren, Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling. *Protein Eng* **12**, 873-878 (1999).
65. A. D. Keefe, J. W. Szostak, Functional proteins from a random-sequence library. *Nature* **410**, 715-718 (2001).
66. B. Bushnell, J. Rood, E. Singer, BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* **12**, e0185056 (2017).
67. I. Rentero Rebollo, M. Sabisz, V. Baeriswyl, C. Heinis, Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res* **42**, e169 (2014).