# Massively parallel protein-protein interaction measurement by sequencing (MP3-seq) enables rapid screening of protein heterodimers

**Authors:** Alexander Baryshev[1*], Alyssa La Fleur[2*], Benjamin Groves[1], Cirstyn Michel[3], David Baker[4,5,6,7], Ajasja Ljubetič[4,5,8,$] and Georg Seelig[1,2,$]

1 Department of Electrical & Computer Engineering, University of Washington, Seattle, WA 98195, USA.
2 Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195, USA.
3 Department of Bioengineering, University of Washington, Seattle, WA 98195, USA.
4 Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.
5 Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.
6 Department of Bioengineering, University of Washington, Seattle, WA, USA
7 Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.
8 Department for Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana SI-1000, Slovenia.
*equal contributions
[$]correspondence: gseelig@uw.edu, ajasja.ljubetic@gmail.com

**Abstract**: Protein-protein interactions (PPIs) regulate almost all cellular processes, and engineered PPIs have cell and gene therapy applications. Identifying all disruptive variants in human PPIs or characterizing all possible interactions in panels of rationally designed proteins requires experimental workflows that can scale to thousands of interactions. We here introduce massively parallel protein-protein interaction measurement by sequencing (MP3-seq), an easy-to-use and highly scalable yeast-two-hybrid (Y2H) approach to address this need. In MP3-seq, DNA barcodes are associated with specific protein pairs; enrichment of these barcodes during yeast selection can be read by sequencing to provide a direct measure of interaction strength. We show that MP3-seq is quantitative over several orders of magnitude of interaction strengths and can scale to measure over 100,000 interactions at once. We apply MP3-seq to characterize interactions between several families of rationally designed heterodimers and develop a greedy algorithm to reduce large-scale MP3-seq screens to identify sets of potentially orthogonal heterodimers. Finally, we use MP3-Seq to delve into and verify the elements that confer specificity to interactions between coil-hairpin-coil binders with designed hydrogen-bond networks.

## Introduction

Synthetic protein binders that can mediate interactions between other proteins or cells have the potential to revolutionize fields from cell therapy (Cho et al., 2018) to synthetic biology (Chen et al., 2020; Chen and Elowitz, 2021; Gao et al., 2018; Groves et al., 2016) and material science (Ben-Sasson et al., 2021; Gonen et al., 2015; Ljubetič et al., 2017b). Early work in synthetic biology often relied on natural interaction domains such as SH3 and PDZ (Kuriyan and Cowburn, 1997). However, such components provide a poor starting point for rationally designing large-scale assemblies due to crosstalk. Consequently, synthetic protein circuits' scale and complexity still pale compared to biological PPI networks. In order to scale up synthetic protein-based circuits, we need large-scale libraries of modular interaction domains.

Fulfilling this need through rational heterodimer design has yielded promising results, but creating large, totally orthogonal sets of dimers remains challenging. Early rational design work focused on coiled-coil dimers (1x1s): while most PPIs involve backbone shape complementarity, 1x1 binding is primarily determined by complementary electrostatic and hydrophobic interactions at specific heptad positions. Because the biophysical rules guiding these interactions are comparably well understood and can be captured by predictive models (Fong et al., 2004; Potapov et al., 2015), orthogonal sets of up to six 1x1 heterodimers have been generated (Lebar et al., 2020; Thompson et al., 2012). Recently, a set of orthogonal 1x1s (consisting primarily of homodimers) was identified in a high-throughput screen (Boldridge et al., 2020). However, the restricted geometry required for 1x1 interactions leads to some degree of interaction promiscuity which fundamentally hampers designing larger orthogonal sets. The Baker lab expanded the coiled-coil toolbox by introducing helical bundle heterodimers, with protomers consisting of two coils connected by a hairpin loop, where interactions are determined by designed hydrogen bond networks between the bundles (2x2s) (Boyken et al., 2016; Chen et al., 2019). This multi-helix bundle approach increases the alphabet of possible coil interactions and could pave the way to generating larger orthogonal sets. However, reliably minimizing off-target interactions (e.g., due to proteins associating in an unexpected register or orientation) remains challenging as these are not captured in the typical biophysical design objective to produce 2x2s. An effective alternative is to prepare diverse libraries of *de novo* binders designed to have strong on-target interactions and then experimentally measure an all-by-all matrix of interactions. Afterward, an orthogonal set can be extracted (Brodnik et al., 2019).

Mass spectrometry and protein-array based methods can measure PPIs but require laborious protein purification steps (Rao et al., 2014; Smits and Vermeulen, 2016). Yeast- or phage-display methods use next-generation sequencing technologies to increase throughput, but are limited to "several-versus-many" screening (Rouet et al., 2018). Alpha-seq, a recent high-throughput method that takes advantage of the yeast mating pathway, overcomes this limitation and allows library-on-library screening (Younger et al., 2017). However, throughput is still limited; moreover, not all proteins of interest will correctly fold when displayed on a surface. Yeast two-hybrid (Y2H) methods are a powerful alternative to surface display for characterizing PPIs (Fields and Song, 1989). In Y2H, one protein of interest is fused to a DNA binding domain (DBD) and the second is fused to a transcriptional activation domain (AD). If the proteins interact, a functional transcription factor is reconstituted and drives the expression of a growth-essential enzyme. Early Y2H approaches tested small numbers of PPIs using plate-based selection, but lab automation and pooling strategies (Luck et al., 2020; Weile et al., 2017) have enabled proteome-scale screens. To further address Y2H scaling issues, high-throughput Y2H (HT-Y2H) and enzyme complementation methods have been developed leveraging next-generation sequencing to read out interaction strength (Diss and Lehner, 2018; Erffelinck et al., 2018; Jin et al., 2007; Rajagopala and Uetz, 2009; Trigg et al., 2017; Weimann et al., 2013; Yachie et al., 2016; F. Yang et al., 2018; J.-S. Yang et al., 2018; Yu et al., 2011). Still, most of these methods require library construction in *E. coli* before interaction screening in yeast or rely on yeast mating and thus necessitate two separate protein libraries to transform MAT$\alpha$ and MATa yeast. High-throughput bacterial two-hybrid assays have been developed, which provide a non-eukaryotic alternative for screening PPIs (Boldridge et al., 2020). Concomitantly with the development of experimental methods, custom workflows for analyzing HT-Y2H data were developed (Banerjee et al., 2021; Velásquez-Zapata et al., 2021).

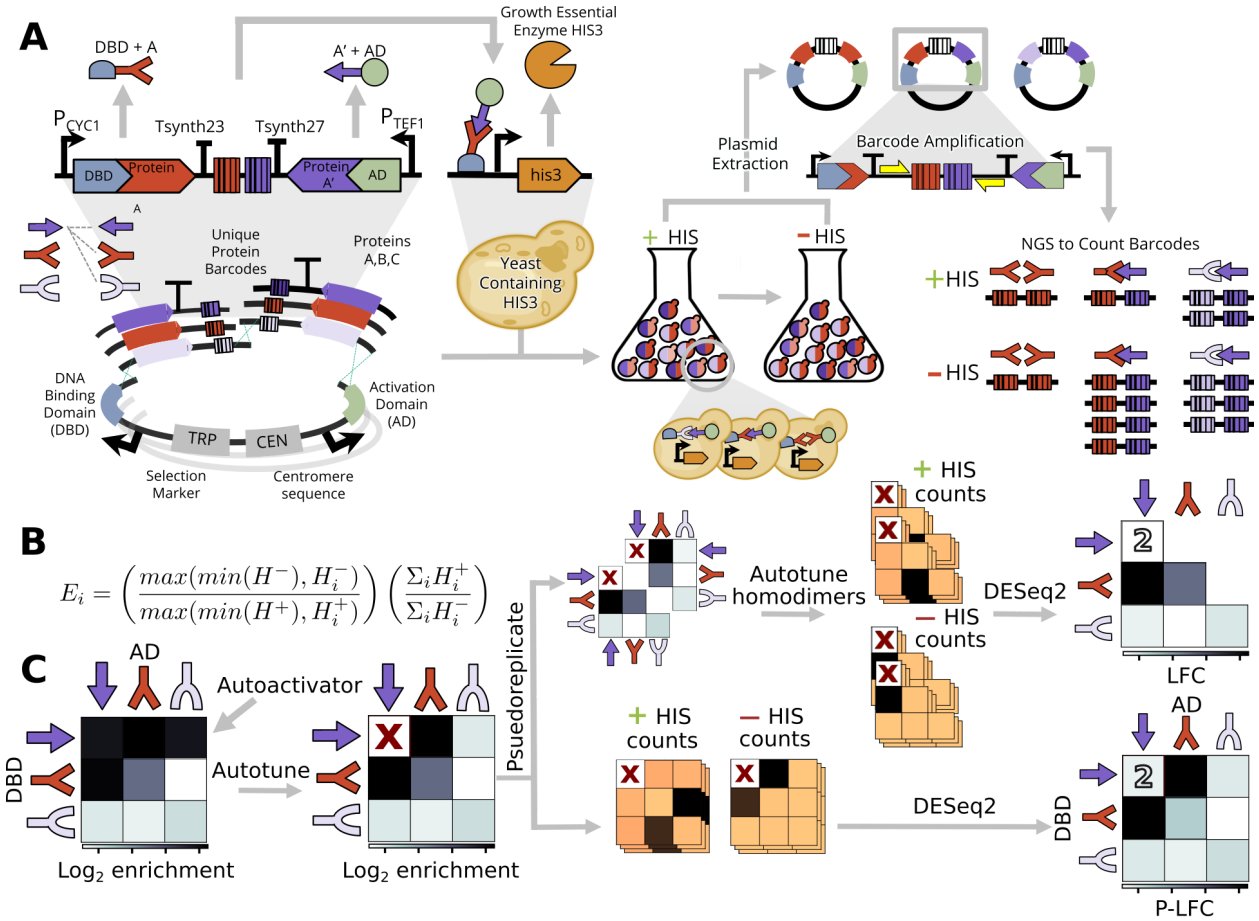



**Figure 1. MP3-seq workflow. A**. Barcoded DNA fragments encoding proteins to be screened, a fragment encoding two terminators, and a backbone fragment bearing a centromere signal, a DBD, and an AD are transformed into yeast cells via electroporation. Homology of fragment 5' and 3' ends allows plasmid assembly in yeast. The centromere sequence ensures the expression of one pair of hybrid proteins per cell on average. Upon transformation, some cells are stored, and some are subjected to growth selection in media lacking histidine. If the hybrid proteins interact, the DBD and AD form a transcription factor that drives HIS3 expression. Therefore, interacting pairs should be enriched in the population post-selection. Cells are lysed, and barcode-containing fragments are amplified and sequenced. **B.** Enrichment can be calculated for each PPI *i* using library size normalized read counts with a pseudo count of the minimum detected value per condition. **C.** After enrichment calculation, each replicate is screened for autoactivators and corrected with Autotune. Replicate pre- and post-selection barcode counts are merged directly with DESeq2 or split into pseudoreplicates and then merged.

Here, we introduced MP3-seq, a massively parallel Y2H workflow for measuring PPIs using sequencing. In MP3-seq, the identity of each protein is encoded in a DNA barcode, and the relative barcode-barcode pair abundance before and after a selection experiment serves as a proxy for interaction strength. Plasmids encoding the protein pairs of interest and their associated barcodes and all other elements required for Y2H experiments are assembled in yeast through homologous recombination. Thus, the experimental workflow bypasses the need for plasmid cloning in E. coli or yeast mating. We first validate MP3-seq using well-characterized coiled-coil heterodimer interactions and synthetic binders for different members of the human BCL2 protein family. Then, we apply MP3-seq to

characterize interactions between rationally designed 2x2 and 1x2 heterodimers and demonstrate that it scales to measuring over 100,000 PPIs in a single experiment. We identify successful designs and use a greedy algorithm to findy potentially orthogonal subsets. Finally, we use MP3-seq to delve into the elements expected to confer 2x2 interaction specificity by screening variations of a successful 2x2 pair.

## Results

**MP3-seq workflow.** In MP3-seq, all molecular components required for measuring interactions between two specific proteins are encoded on a single plasmid (**Figure 1A**). A plasmid library is constructed directly through homologous recombination in yeast to measure all possible interactions for a set of proteins. To do so, we transform haploid MATa-type yeast with a mixture of DNA fragments. One of these fragments is a backbone carrying a centromere sequence, the selection marker, a DBD, and an AD. On average, the centromere sequence ensures that only one pair of proteins is expressed in each cell (Scanlon et al., 2009). We use a Cys(2)His(2) zinc-finger domain of the mouse transcription factor Zif268 (McIsaac et al., 2013) as the DBD and its corresponding promoter to drive the expression of the growth essential enzyme *his3*. The herpes simplex virus-derived protein domain VP16 is used as the AD. Additional fragments each contain one of the proteins of interest and its associated unique barcode separated by a terminator sequence. In the experiments described here, these fragments are ordered as a single oligonucleotide but for longer proteins a barcode can be added to a protein of interest by PCR. Due to their short length and distinct sequences, we used Tsynth23 and Tsynth27 as terminators in most of our experiments (Curran et al., 2015).

After transformation, we perform selection in media without Tryptophan (Trp) to ensure plasmid maintenance. A portion of the cells is frozen, while the rest undergo selection in media lacking histidine (His). Plasmid DNA is extracted from pre and post-His selection cells, and the barcode-containing regions are amplified (**Figure 1A**, right). The barcode-barcode amplicons are sequenced, and their relative enrichment can be calculated from barcode counts to serve as a proxy for interaction strength **(Figure 1B)**. Some proteins' expression or folding may be impacted by fusion with either the AD or DBD; therefore, we test all proteins fused to both the AD and DBD

**MP3-Seq analysis pipeline.** As a first step in the analysis workflow we calculate the enrichment between the pre and post His selection stages of MP3-Seq normalized by library size (**Figure 1B**). Next, we detect autoactivators, and optionally replace their His selection values with values extracted from the functioning orientation (see Autotune in Methods); autoactivation is an error mode in Y2H experiments wherein high enrichment values are observed for a protein of interest for all interaction partners, suggesting that the protein non-specifically activates expression of the selection marker (Shivhare et al., 2021). Typically, this behavior is observed fused to either the AD or the DBD but not in both orientations. Note that if an autoactivator is found, its homodimer is not recoverable with Autotune as there is no second His measurement set to use for correction.

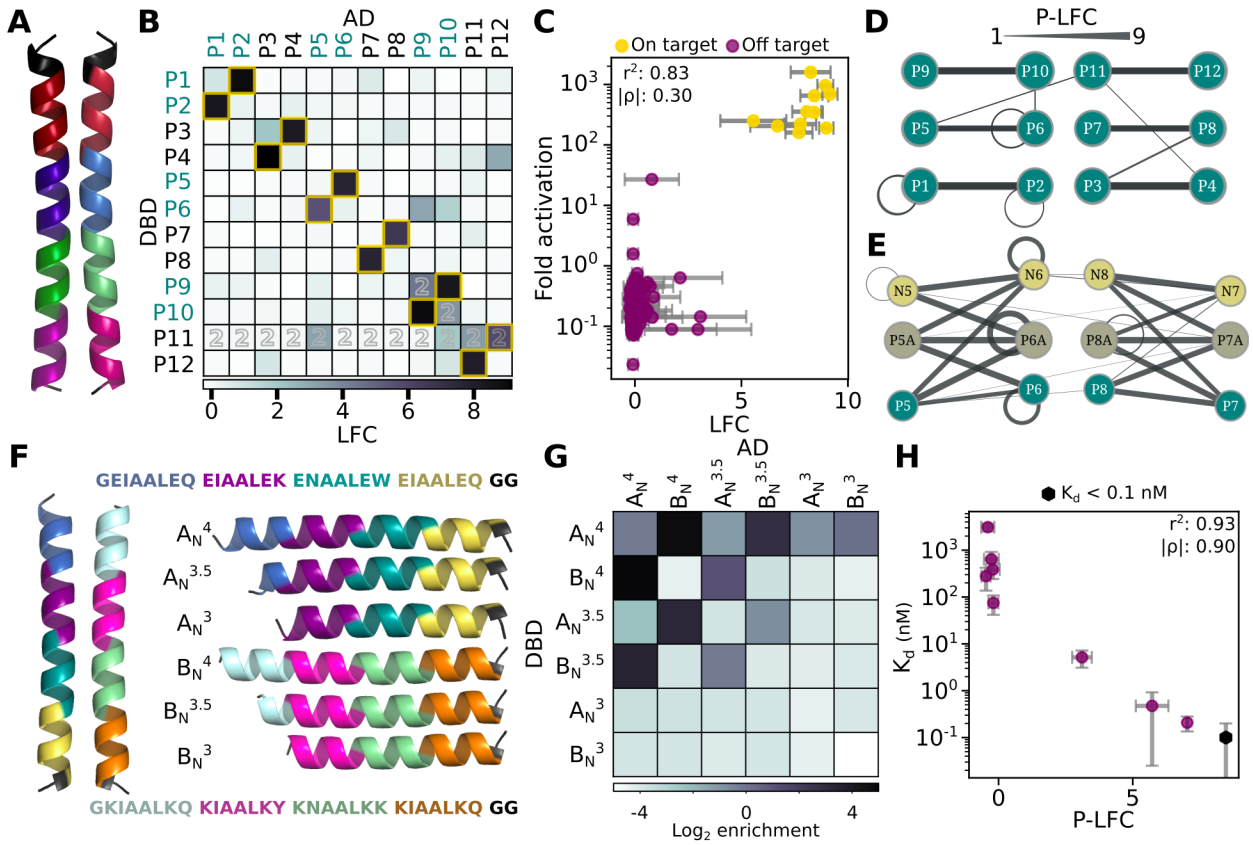



**Figure 2: Validation of MP3-seq with coiled-coil heterodimers. A**. P-series 1x1s from (Lebar et al., 2020) with complementary heptads designed to interact shown in like colors. **B**. MP3-Seq LFC of the P-series coil interactions. All values were calculated from 3 experimental replicates except for those labeled "2," where only 2 were available. Yellow outlines denote on-target interactions. **C.** Correlation of the on- and off-target P-series MP3-seq LFCs with fold activation fluorescence values from (Lebar et al., 2020) **D.** Filtering P-series Jerala interactions and **E.** P, PA, and N series designed coil interactions (Plaper et al., 2021) to include only those with $p_{adj} \leq 0.05$. Line weights correspond to MP3-Seq P-LFCs. **F.** A designed 1x1 and its truncations from (Thomas et al., 2013) **G**. MP3-Seq enrichment for the $A_N$ and $B_N$ coils and their truncations. **H**. Correlation of MP3-Seq P-LFCs with $K_d$ values from (Thomas et al., 2013) for the $A_N$ and $B_N$ truncations.

Following autoactivator removal, enrichment values can be averaged together to obtain interaction strength values. Alternatively, to correct for variation in the read count distribution between experimental replicates resulting from different sequencing depths, selection times, and other experimental factors, we calculate $\log_2$ fold changes (LFC) using the DESeq2 package (Love et al., 2014). DESeq2 calculates differential enrichment across multiple replicates and provides Hochberg-adjusted Wald test p-values, identifying PPIs with LFCs significantly different from an LFC of zero.

In some cases, it is desirable to combine MP3-seq measurements for both fusion orderings (e.g., P1-DBD + P2-AD and P2-DBD + P1-AD -> P1-P2 ). For example, comparing MP3-seq LFCs to $K_d$s collected for the pair of interest (P1-P2). The simplest method for merging these measurements is applying an operation to both, such as averaging or taking the max of the values. Here, we use an alternative approach and treat each fusion ordering as an independent set of measurements, turning them into two

pseudoreplicates. These pseudoreplicates are then combined with DESeq2 to calculate LFCs (**Figure 1B**). We refer to these values as pseudoreplicate LFCs (P-LFCs).

**MP3-seq benchmarking with orthogonal coiled-coil dimers.** To validate MP3-seq, we screened 144 pairwise interactions between six orthogonal 1x1s from Lebar et al. (Lebar et al., 2020) (**Figure 2A**). A pool of 24 DNA fragments where each of the 12 proteins was fused to either the AD or DBD was ordered. Then, all interaction pairs were assembled in a pooled experiment, and interaction strengths were quantified with MP3-seq. We confirmed that selections were occurring as expected by checking pre-His selection barcode counts remained steady over time (**Figure S1A**), that applying different levels of 3-Amino-1,2,4 triazole (3-AT), a competitive inhibitor of His3, affected post-His selection counts (**Figure S1B-C**), and that counts correlated more for post-His than pre-His counts between replicates (**Figure S1E**). LFCs for all orientations were calculated from three experimental replicates which confirmed that the interactions occur almost exclusively between intended partners (i.e., P1:P2, P3:P4, etc., **Figure 2B**). Homodimer plasmid assembly is less efficient in MP3-Seq and generally results in lower input barcode counts for these PPIs, but coverage was sufficient for inclusion in the analysis. This phenomenon is likely due to increased sequence homology, which interferes with the correct ordering of DNA fragments during plasmid assembly. For a more quantitative analysis, we correlated MP3-seq LFCs with luciferase expression assay interaction scores in HEK293T cells from (Lebar et al., 2020) and found good agreement ($R^2$=0.83, **Figure 2C**). **Figure 2D** represents P-LFC MP3-Seq data for these PPIs as a graph, where each protein is a vertex and edges are significant measured interactions weighted by P-LFC.

In a separate all-by-all experiment with 28 proteins, we screened the P-series 1x1s and two sets of 1x1s (N1,N2,N5-N8 and P5A-P8A) derived from the P-series but modified in an attempt to increase the thermodynamic stability of their on-target interactions (Plaper et al., 2021). N1 exhibited strong auto-activation properties when fused to the DBD and these data were infilled with AD fusion measurements using Autotune. In this experiment, we expect related coils to have similar interaction patterns (e.g. P5A can bind to P6A, N6, and P6) which agrees with the interaction graph created from significant MP3-Seq P-LFCs in **Figure 2E**.

To assess if MP3-seq values provide quantitative information about interaction strength, we measured interactions between a set of varying-length 1x1s designed to span a wide range of $K_d$s (Thomas et al., 2013) (**Figure 2F**). We performed an all-by-all measurement; these interactions are not expected to be orthogonal as (Thomas et al., 2013) achieved weaker interactions by truncating the two parent four-heptad binders ($A_N^4$, $B_N^4$) by a half or full heptad (**Figure 2G**). The MP3-seq P-LFCs correlated very well with previously measured dissociation constants over approximately three decades ($r^2$=0.94, **Figure 2H)**. We note that $A_N^4$ was a weak auto-activator when fused to the DBD, and these data were infilled using Autotune for the P-LFC values in **Figure 2H**.

**MP3-seq benchmarking with BCL2 family binders.** To validate MP3-seq outside of 1x1 interactions, we tested a set of proteins previously characterized by biolayer interferometry (Berger et al., 2016) and Alpha-seq (Younger et al., 2017) comprised of six homologous proteins from the BCL2 family (Bcl-2, Bcl-xL, Bcl-w, Mcl-1, Bfl-1, Bcl-B) and nine *de novo* designed inhibitors of those homologs.

A crystal structure of one of the synthetic binders bound to its target is shown in **Figure 3A,** while **Figure 3B** shows the domain organization of the six human proteins. All inhibitors were designed to interact with the BH3 domain binding pocket of the homologs. To better compare with the HT-Y2H method Alpha-seq, a truncated version of Mcl-1 was used. An MP3-seq screen of all BCL2 homologs against all inhibitors is shown in **Figure 3C.** An older version of MP3-seq was used for the first two replicates of this experiment. Unlike the current version, barcodes were inserted into the 3'UTRs of the binders upstream of the terminators and could thus differentially impact mRNA stability and protein levels. However, the inter-replicate correlations suggest that the data is consistent between versions (**Figure S2A**).

Only binders beginning with alpha correspond to the final, specific designs of (Berger et al., 2016), while the others are intermediate or failed designs. This can be seen in **Figure 3C**, with a divide between the largely orthogonal rightmost four columns of the heatmap of MP3-seq P-LFCs and the left, less specific columns.

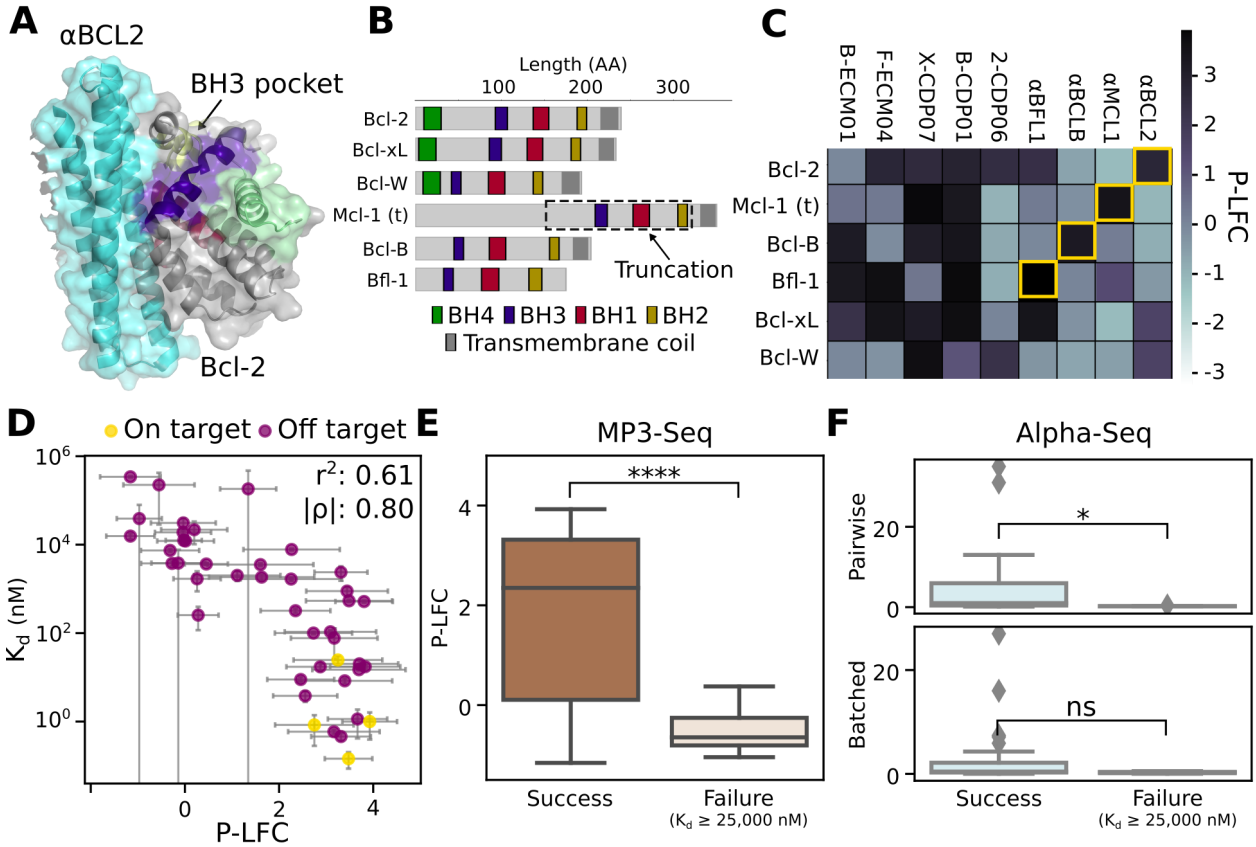



**Figure 3: Validation of MP3-Seq with BCL2 proteins. A)** Colored crystal structure of Bcl-2 and its designed BH3 binding inhibitor from (Berger et al., 2016) (PDB: 5JSN). Binder in blue, Bcl-2 in gray with different domains in the binding region highlighted in color. **(B)** BH3 binding domain annotations of the six human Bcl homologs measured by (Berger et al., 2016; Younger et al., 2017) with the experimental truncation used by Younger and in this work shown. **C)** MP3-Seq P-LFCs for inhibitor and Bcl-2 interactions. Yellow boxes highlight intended on-target interactions **D)** Correlations with $K_d$ measurements from biolayer interferometry. **E)** MP3-Seq P-LFC (**F)** and Alpha-Seq distributions for interactions that failed $K_d$ measurement due to instrument detection limits.

Our data agree well with dissociation constant measurements obtained from biolayer interferometry by (Berger et al., 2016) ($R^2$=0.54, **Figure 3D**). We also exhibit good agreement with Alpha-Seq percent survival for both their low-throughput pairwise assay and high-throughput batched assay ($r^2$: 0.45, $r^2$:0.69, **Figure S3**). Some variation between our results and those published earlier might be due to the different interaction measurement settings. In our case, interactions are measured with proteins expressed in yeast, while biolayer interferometry uses purified proteins, and Alpha-seq displays proteins on the yeast surface.

Some of the BCL2 inhibitors failed to produce $K_d$ values when measured with biolayer interferometry, likely because the interactions were weak and below detection limits. We examined the failed (n=11) versus successfully detected interactions (n=43) and found that the mean MP3-seq value of PPIs for which biolayer interferometry succeeded was significantly greater than for the subset for which biolayer interferometry failed (independent t-test, $H_0$: $\mu_{success} > \mu_{fail}$) **(Figure 3E)**. Similarly, pairwise Alpha-Seq also had the mean of successful interactions significantly greater than that of the failures. However, batched Alpha-Seq, the high throughput version of the assay, did not exhibit this **(Figure 3F).** Together, these results show that MP3-seq can work with globular proteins in addition to coils and that MP3-seq results agree with those obtained by Alpha-seq and biolayer interferometry. The significant difference between the successes and failures provides evidence that MP3-seq excels at separating very weak PPIs from others at high throughput.

**Large-scale assay of designed heterodimers**. To explore the feasibility of using MP3-seq for large-scale screening, we prepared a library of designed heterodimers (DHDs) following (Chen et al., 2019). On-target pairs were designed as three- or four-helix bundles with buried hydrogen bond networks (HB-Nets) connecting all helices and then split into a helix-turn-helix hairpin and a single helix (1x2s) or 2x2s. Initially, we chose 100 designs for testing (DHD1, **Figure 4A**), later adding 52 more designs in a second round of experiments (DHD2, **Figure 4A**). As controls, we selected nine previously tested 2x2s (Chen et al., 2019). We also included several binders derived from these controls through modification of the hairpin loops or truncation of the helices for a total of 35 pairs (DHD0, **Figure 4A**). Finally, we added an additional series derived from a common parent 2x2 through truncating heptads (mALB, **Figure 4A**). For each on-target pair, one protomer is designated "A" and the other "B". Interactions between and within these groups were tested in MP3-seq experiments of varying sizes. We performed two replicates of an all-by-all screen including DHD0, DHD, DHD2, and mALb proteins as well as the BCL2 homologs and cognate binders described above, resulting in a matrix of 337x337 =113,569 interactions. Additionally, we performed three replicates of a screen using only DHD1 with a subset of DHD0, and one screen with DHD0, DHD2, and mALB. The older MP3-seq was used for most experiments in this section. Still, the experiment highlights the large experimental scales achievable with MP3-seq and the inter-replicate correlations suggest that the data is internally consistent (**Figure S2 B-C**).

We applied autotune corrections and symmetrized the read counts to calculate P-LFCs for all possible pairs. On-target interactions, where the A and B protomers were designed to interact, were evaluated for success. In this case, we defined a 'success' as any PPI with a $p_{adj} \leq 0.01$ and a P-LFC $\geq 4$. The number of successes out of the total possible on-target PPIs in each design set can be seen in **Figure 4B.** Every design set had an approximately 20% success rate in the largest screen. This is consistent with the 22% success rate for a-Helical bundles (Ljubetič et al., 2017a). The top 5 P-LFC on-target interactions

for each design category from the are shown in **Figure 4C.** Note that the successes, though strong in the intended interaction, are not necessarily orthogonal with other designs, even when examining only the PPIs of the A and B protomers of the top five designs.

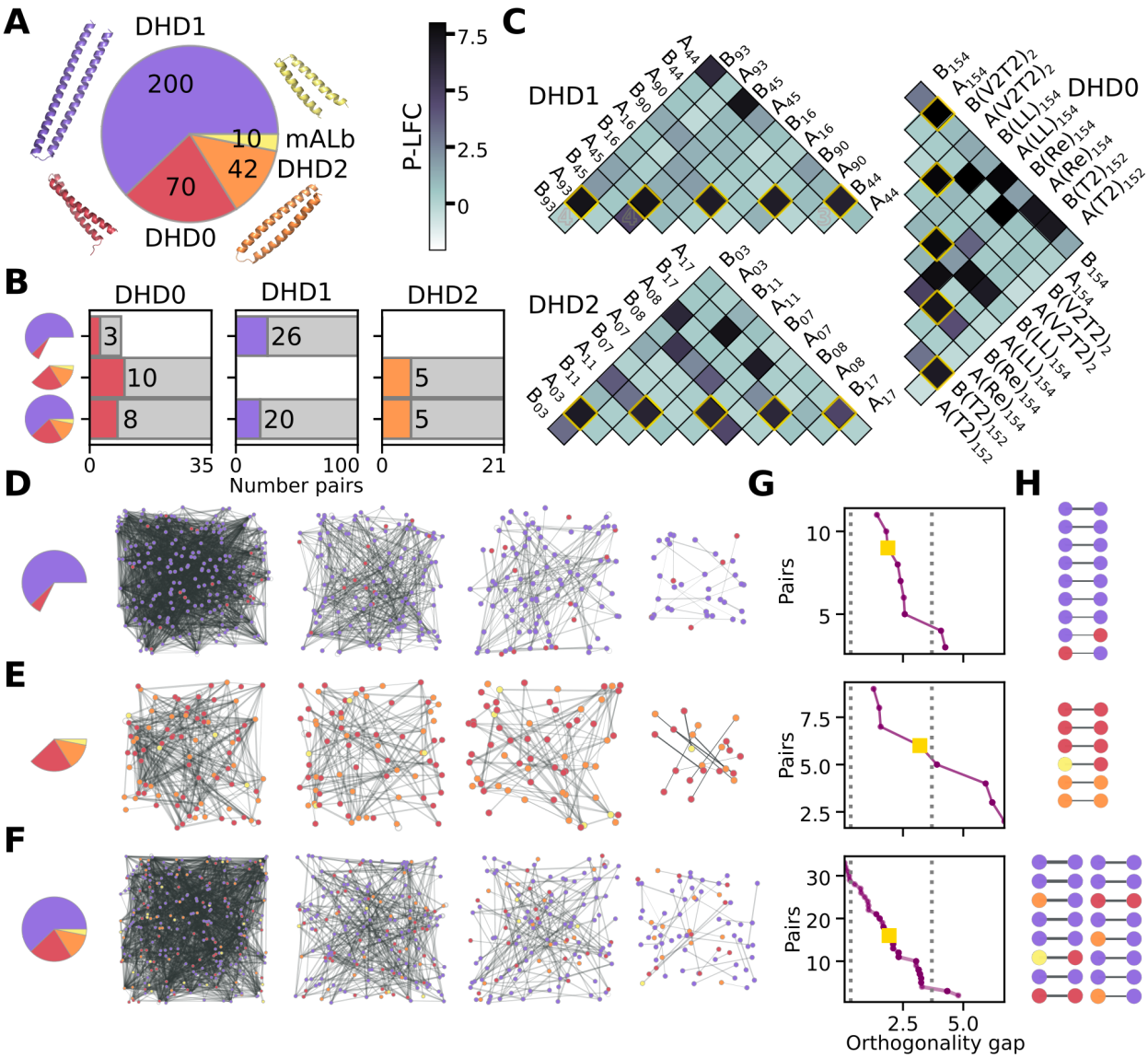



**Figure 4. High throughput screening of designed heterodimers. A.** Summary of the synthetic binder types. **B.** Successful designs for the three main sets. Gray bars are possible pairs for the set, and colored bars indicate the number of successes. **C.** Top 5 successful designs from each set by P-LFC. Yellow boxes show on-target interactions. **D.** From left to right, pre-DUET positive, $p_{adj} \leq 0.01$ P-LFC PPI network for the DHD1 and DHD0 subset, and the network at DUET iterations 50, 75, and 113. **E.** Pre-DUET network for the DHD0, DHD2, and mALb sets, and the network at DUET iterations 10, 15, and 37 (final). **F.** Pre-DUET network for all designs and the network at DUET iterations 50, 75, and 114. **G.** Orthogonality gaps of the DUET final networks without significance filtering. Left and right dashed lines show the MP3-seq orthogonality gaps for the BCL-2 and inhibitor interactions and the P-series interactions, respectively. Yellow squares correspond to half of the starting networks remaining. **H.** DUET final networks reduced to half their final iteration size.

**Finding orthogonal subsets.** For all four sets of designs (DHD1-3, mALB), most measured interactions were off-target interactions between protomers screened in the same experiment (e.g. A1 and B2 of DHD1, instead of A1 and B1). For example, in the all-by-all screen with all of DHD1-3 and mALB, only 33 of 914 total PPIs with $p_{adj} \leq 0.01$ and P-LFC $\geq 4$ were on-target.

Given the number of strong but unintended interactions, we asked whether we could extract potentially useful orthogonal PPI subsets from the data. To this end, we used significant (alpha = 0.01), positive P-LFC values to construct a weighted undirected graph, as in **Figure 2D-E**. Our goal can then be rephrased as finding a graph of degree one vertices (1-regular graph) without self-edges. To do this, we developed a simple scoring function that rewards graphs based on existing orthogonal edges or those over a desired orthogonality cutoff and punishes graphs for non-orthogonal edges. This scoring function was used in a greedy graph reduction method, Deleting Undirected Edges Thoughtfully (DUET), which removes a vertex and its associated edges each iteration until a 1-regular graph remains (see Methods). We applied DUET to different data sets. For DHD1 and the DHD0 subset, we went from 2,001 edges between 202 vertices to 18 DUET pairs **(Figure 4D)**, for the DHD0/DHD2/mALb data from 279 edges between 85 vertices to 11 DUET pairs **(Figure 4E)**, and for all designs (DHD0-2,mALb) from 1,562 edges between 270 vertices to 36 DUET pairs **(Figure 4F)**. Of these, 2, 2, and 4 DUET pairs were on-target for DHD0/DHD1, DHD0/DHD2/mALb, and all designs, respectively.

The DUET final results are only orthogonal if all non-highly significant interactions are considered truly non-interacting. As this may not be the case, we used all P-LFC >0 between DUET pair protomers regardless of significance for a more conservative analysis. First, we removed interactions with protomoers for which the DUET pair P-LFC was lower than the highest non-DUET pair P-LFC. Then, we reduced the remaining DUET pairs by removing whichever pair had the largest non-DUET P-LFC one by one (**Figure 4F**). When reduced to half the starting DUET pairs (squares in **Figure 4F**), the orthogonal sets shown in **Figure 4G** are left, which all have gaps between those of the P-series in **Figure 2** and the BCL2-inhibitor pairs in **Figure 3**.

**Elucidating the rules of specific helix-loop-helix binding.** High-throughput, high-quality data can be used to probe novel design rules to improve protein design. Here, we demonstrate how MP3-Seq can be used to understand how hydrogen bond networks and binder length can confer binding specificity in a designed pair by generating and testing variants of a target design. Our target pair, mALb8, is a high-affinity 2x2 heterodimer with three HB-Nets between its A and B protomers and was included in the mALb set introduced earlier (**Figure 5A**).

First, two mALb protomer truncations were designed, removing one or two turns from the end of each alpha helix (**Figure 5B)**. Truncation 1 (T1) versions are one turn shorter than the original monomers, while Truncation 2 (T2) versions are two full turns shorter (see **Figure S4**). The original binders and the truncations were screened in an all-by-all screen with MP3-Seq and LFCs were calculated from two replicates. It can be seen from **Figure 5B** that truncating the binders by two turns (T2) eliminates binding activity. The relationship between length and binding affinity is highly non-linear, since binding disappears once the minimum length is reached. We can use this data to infer the shortest possible length (3.5 heptades) for 2x2 binders containing two buried HB-Nets. This pattern is similar to that seen for the $A_N$ and $B_N$ single helix truncation experiments in **Figure 1D-E**.

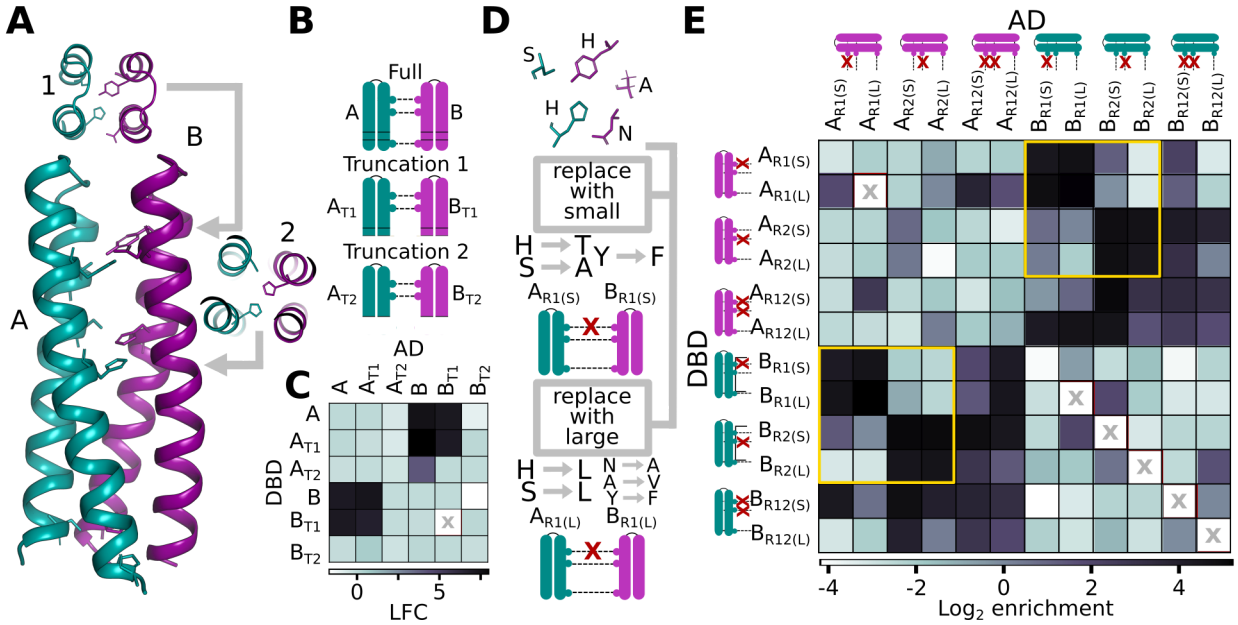



**Figure 5: Systematic exploration of length and hydrogen bond network presence on helix-loop-helix binding.** **A**. The mALb8 helix-loop-helix heterodimer with HB-Nets 1 and 2 shown. **B**. mALb8 and the two shorter variants (T1; one helical turn shorter and T2; two helical turns shorter). **C**. LFC of the truncated mALb8 interactions. Gray X squares indicate insufficient reads. We observed a markedly non linear response to the length of the binder. T2 variants which are 13 residues shorter than the original do not bind. **D**. An example of HB-Net removal using the small and large hydrophobic replacement protocols for the first HB-Net **E**. MP3-seq log2 enrichment for the mALb8 with removed HB-Nets. S and L designate the replacement protocol used, while $R_1$, $R_2$, or $R_{12}$ denote if the first, second, or both HBNets were replaced. Gray X indicates insufficient reads. The yellow boxed regions show that two HB-Net mismatches are enough to specify orthogonality. Hydrophobic residues alone cannot confer specificity.

To probe the specificity conferred by the buried HB-Nets in the 2x2 binders designed with the approach from (Chen et al., 2019) , we have removed either the first, the second, or both buried HB-Nets from the mALb8 dimer. This was done by replacing HB-Net amino acids in both A and B protomers simultaneously with hydrophobic residues that can not form hydrogen bonds. Two Rosetta-design replacement protocols were used ;one using smaller and one using larger hydrophobic residues for replacement. . The first network's replacements in A and B can be seen in **Figure 5D**, while all replaced designs can be found in **Figure S5.** We observe that the designs with removed HB-Nets (R variants) largely do not homodimerize, as interactions between various A and B protomers are low. These results show that, at minimum, two hydrogen bond network mismatches are needed to confer specificity and orthogonality (**Figure 5E,** boxed areas). For example, the binding of $A_{R1}$ and $B_{R2}$ would bury half of the hydrogen bond network on protomer A and half of the hydrogen bond network on protomer B. The binding of $A_{R12}$ and $B_{R2}$ would result in only one HB-Net mismatch on chain B. The binding data suggest that pairs with one HB-NetT mismatch can still exhibit partial binding affinity, for example $B_{R12(S)}$ weakly binds to both $A_{R1}$ and $A_{R2.}$ Additionally, hydrophobic residue size alone is not sufficient to confer orthogonality. For example, both $B_{R1(S)}$ and $B_{R1(L)}$ bind to both $A_{R1(L)}$ and $A_{R1(S)}$.

## Discussion

Here, we introduce MP3-seq, an easy-to-use HT-Y2H method that can measure pairwise PPIs in a single yeast strain. In MP3-seq, the identity of each protein is encoded in a DNA barcode, and the relative barcode-barcode pair abundance before and after a selection experiment serves as a proxy for interaction strength. We also developed a data analysis workflow based on DESeq2 that makes it easy to merge replicates, remove autoactivators, and identify statistically significant interactions. In future work, the MP3-seq workflow could be further generalized by adding autoactivator screening or adjusting the selection scheme to eliminate autoactivators (Shivhare et al., 2021). MP3-seq could also be combined with protein stability and expression assays to confirm that proteins tested in interaction screens are folded correctly and that a negative interaction measurement is not due to reduced protein levels. In particular, it is easy to imagine integrating MP3-seq with other growth selection-based workflows such as Stable-seq (Kim et al., 2013). We expect stability screening to be more important for human protein than for the synthetic binders tested here, which are generally designed to have very stable folds.

We validated MP3-seq using several sets of proteins for which interactions had been previously characterized. In particular, we used a family of human pro-apoptotic proteins BCL2 and their *de novo* designed inhibitors, and three different sets of coiled-coils peptides. We generally found quantitative agreement between our results and those reported previously. We then applied our method to characterize interactions in a pool of *de novo* coil-based heterodimers that contain buried hydrogen bond networks. We showed that our approach could scale to measuring more than 100,000 interactions in a single experiment. Our computational workflow enabled us to identify potentially orthogonal subsets, at various orthogonality gaps, within larger all-by-all screens.

Moreover, by screening interactions between protomers with truncations and modified hydrogen bond networks, we probed important design rules for 2x2 binders with two buried HB-nets. Firstly, we showed that the minimum length for strong binding is 3.5 heptads and that binding affinity drops sharply for shorter designs. Next, we showed that at least two hydrogen bond network mismatches are needed for orthogonality, thereby setting minimum requirements for future sets of orthogonal 2x2 binders with buried HB-nets.

We hope that the increased scale and streamlined workflow of MP3-seq will further accelerate the adoption of HT-Y2H methods and facilitate their use in additional applications. For example, we expect that high-throughput data collected with MP3-seq could be used to train predictive models of PPIs and that such models could be combined with generative algorithms to iteratively design heterodimers with desirable properties or with neural network interpretation to uncover determinants of binding specificity. Moreover, MP3-seq could be applied to characterize interactions between human proteins or to screen variants' impact on PPIs at high throughput.

## Acknowledgments

This work was supported by NIH Award R01GM120379 and ONR Award N00014-16-1-3189 to G.S. A.L. would like to acknowledge funding of European Commission MSCA CC-LEGO 792305 and Slovenian research agency project CC-TRIGGER J1-4406.

**Bibliography**


Banerjee, S., Velásquez-Zapata, V., Fuerst, G., Elmore, J.M., Wise, R.P., 2021. NGPINT: a next-generation protein–protein interaction software. Brief. Bioinform. 22, bbaa351. https://doi.org/10.1093/bib/bbaa351

Benatuil, L., Perez, J.M., Belk, J., Hsieh, C.-M., 2010. An improved yeast transformation method for the generation of very large human antibody libraries. Protein Eng. Des. Sel. PEDS 23, 155–159. https://doi.org/10.1093/protein/gzq002

Ben-Sasson, A.J., Watson, J.L., Sheffler, W., Johnson, M.C., Bittleston, A., Somasundaram, L., Decarreau, J., Jiao, F., Chen, J., Mela, I., Drabek, A.A., Jarrett, S.M., Blacklow, S.C., Kaminski, C.F., Hura, G.L., De Yoreo, J.J., Kollman, J.M., Ruohola-Baker, H., Derivery, E., Baker, D., 2021. Design of biologically active binary protein 2D materials. Nature 589, 468–473. https://doi.org/10.1038/s41586-020-03120-8

Berger, S., Procko, E., Margineantu, D., Lee, E.F., Shen, B.W., Zelter, A., Silva, D.-A., Chawla, K., Herold, M.J., Garnier, J.-M., Johnson, R., MacCoss, M.J., Lessene, G., Davis, T.N., Stayton, P.S., Stoddard, B.L., Fairlie, W.D., Hockenbery, D.M., Baker, D., 2016. Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. eLife 5. https://doi.org/10.7554/eLife.20352

Boldridge, W.C., Ljubetič, A., Kim, H., Lubock, N., Szilágyi, D., Lee, J., Brodnik, A., Jerala, R., Kosuri, S., 2020. A Multiplexed Bacterial Two-Hybrid for Rapid Characterization of Protein-Protein Interactions and Iterative Protein Design (preprint). Molecular Biology. https://doi.org/10.1101/2020.11.12.377184

Boyken, S.E., Chen, Z., Groves, B., Langan, R.A., Oberdorfer, G., Ford, A., Gilmore, J.M., Xu, C., DiMaio, F., Pereira, J.H., Sankaran, B., Seelig, G., Zwart, P.H., Baker, D., 2016. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science 352, 680–687. https://doi.org/10.1126/science.aad8865

Brodnik, A., Jovičić, V., Palangetić, M., Silađi, D., 2019. Construction of orthogonal CC-sets. Informatica 43. https://doi.org/10.31449/inf.v43i1.2693

Chen, Z., Boyken, S.E., Jia, M., Busch, F., Flores-Solis, D., Bick, M.J., Lu, P., VanAernum, Z.L., Sahasrabuddhe, A., Langan, R.A., Bermeo, S., Brunette, T.J., Mulligan, V.K., Carter, L.P., DiMaio, F., Sgourakis, N.G., Wysocki, V.H., Baker, D., 2019. Programmable design of orthogonal protein heterodimers. Nature 565, 106–111. https://doi.org/10.1038/s41586-018-0802-y

Chen, Z., Elowitz, M.B., 2021. Programmable protein circuit design. Cell 184, 2284–2301. https://doi.org/10.1016/j.cell.2021.03.007

Chen, Z., Kibler, R.D., Hunt, A., Busch, F., Pearl, J., Jia, M., VanAernum, Z.L., Wicky, B.I.M., Dods, G., Liao, H., Wilken, M.S., Ciarlo, C., Green, S., El-Samad, H., Stamatoyannopoulos, J., Wysocki, V.H., Jewett, M.C., Boyken, S.E., Baker, D., 2020. De novo design of protein logic gates. Science 368, 78–84. https://doi.org/10.1126/science.aay2790

Cho, J.H., Collins, J.J., Wong, W.W., 2018. Universal Chimeric Antigen Receptors for Multiplexed and Logical Control of T Cell Responses. Cell 173, 1426-1438.e11. https://doi.org/10.1016/j.cell.2018.03.038

Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., Alper, H.S., 2015. Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. ACS Synth. Biol. 4, 824–832. https://doi.org/10.1021/sb5003357

Diss, G., Lehner, B., 2018. The genetic landscape of a physical interaction. eLife 7, e32472. https://doi.org/10.7554/eLife.32472

Erffelinck, M.-L., Ribeiro, B., Perassolo, M., Pauwels, L., Pollier, J., Storme, V., Goossens, A., 2018. A user-friendly platform for yeast two-hybrid library screening using next

generation sequencing. PLOS ONE 13, e0201270. https://doi.org/10.1371/journal.pone.0201270
Fields, S., Song, O., 1989. A novel genetic system to detect protein-protein interactions. Nature 340, 245–246. https://doi.org/10.1038/340245a0
Fong, J.H., Keating, A.E., Singh, M., 2004. Predicting specificity in bZIP coiled-coil interactions. Genome Biol. 5, R11. https://doi.org/10.1186/gb-2004-5-2-r11
Gao, X.J., Chong, L.S., Kim, M.S., Elowitz, M.B., 2018. Programmable protein circuits in living cells. Science 361, 1252–1258. https://doi.org/10.1126/science.aat5062
Gonen, S., DiMaio, F., Gonen, T., Baker, D., 2015. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. Science 348, 1365–1368. https://doi.org/10.1126/science.aaa9897
Groves, B., Khakhar, A., Nadel, C.M., Gardner, R.G., Seelig, G., 2016. Rewiring MAP kinases in Saccharomyces cerevisiae to regulate novel targets through ubiquitination. eLife 5, e15200. https://doi.org/10.7554/eLife.15200
Jin, F., Avramova, L., Huang, J., Hazbun, T., 2007. A yeast two-hybrid smart-pool-array system for protein-interaction mapping. Nat. Methods 4, 405–407. https://doi.org/10.1038/nmeth1042
Kim, I., Miller, C.R., Young, D.L., Fields, S., 2013. High-throughput Analysis of in vivo Protein Stability. Mol. Cell. Proteomics 12, 3370–3378. https://doi.org/10.1074/mcp.O113.031708
Kuriyan, J., Cowburn, D., 1997. Modular peptide recognition domains in eukaryotic signaling. Annu. Rev. Biophys. Biomol. Struct. 26, 259–288. https://doi.org/10.1146/annurev.biophys.26.1.259
Lebar, T., Lainšček, D., Merljak, E., Aupič, J., Jerala, R., 2020. A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. Nat. Chem. Biol. 16, 513–519. https://doi.org/10.1038/s41589-019-0443-y
Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A., 2022. Evolutionary-scale prediction of atomic level protein structure with a language model (preprint). Synthetic Biology. https://doi.org/10.1101/2022.07.20.500902
Ljubetič, A., Gradišar, H., Jerala, R., 2017a. Advances in design of protein folds and assemblies. Curr. Opin. Chem. Biol. 40, 65–71. https://doi.org/10.1016/j.cbpa.2017.06.020
Ljubetič, A., Lapenta, F., Gradišar, H., Drobnak, I., Aupič, J., Strmšek, Ž., Lainšček, D., Hafner-Bratkovič, I., Majerle, A., Krivec, N., Benčina, M., Pisanski, T., Veličković, T.Ć., Round, A., Carazo, J.M., Melero, R., Jerala, R., 2017b. Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. Nat. Biotechnol. 35, 1094–1101. https://doi.org/10.1038/nbt.3994
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8
Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., Choi, D., Coté, A.G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M.F., Kishore, N., Knapp, J.J., Kovács, I.A., Lemmens, I., Mee, M.W., Mellor, J.C., Pollis, C., Pons, C., Richardson, A.D., Schlabach, S., Teeking, B., Yadav, A., Babor, M., Balcha, D., Basha, O., Bowman-Colin, C., Chin, S.-F., Choi, S.G., Colabella, C., Coppin, G., D'Amata, C., De Ridder, D., De Rouck, S., Duran-Frigola, M., Ennajdaoui, H., Goebels, F., Goehring, L., Gopal, A., Haddad, G., Hatchi, E., Helmy, M., Jacob, Y., Kassa, Y., Landini, S., Li, R., van Lieshout, N., MacWilliams, A., Markey, D., Paulson, J.N., Rangarajan, S., Rasla, J., Rayhan, A., Rolland, T., San-Miguel, A., Shen, Y., Sheykhkarimli, D., Sheynkman, G.M., Simonovsky,

E., Taşan, M., Tejeda, A., Tropepe, V., Twizere, J.-C., Wang, Y., Weatheritt, R.J., Weile, J., Xia, Y., Yang, X., Yeger-Lotem, E., Zhong, Q., Aloy, P., Bader, G.D., De Las Rivas, J., Gaudet, S., Hao, T., Rak, J., Tavernier, J., Hill, D.E., Vidal, M., Roth, F.P., Calderwood, M.A., 2020. A reference map of the human binary protein interactome. Nature 580, 402–408. https://doi.org/10.1038/s41586-020-2188-x

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10. https://doi.org/10.14806/ej.17.1.200

McIsaac, R.S., Oakes, B.L., Wang, X., Dummit, K.A., Botstein, D., Noyes, M.B., 2013. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. Nucleic Acids Res. 41, e57. https://doi.org/10.1093/nar/gks1313

Plaper, T., Aupič, J., Dekleva, P., Lapenta, F., Keber, M.M., Jerala, R., Benčina, M., 2021. Coiled-coil heterodimers with increased stability for cellular regulation and sensing SARS-CoV-2 spike protein-mediated cell fusion. Sci. Rep. 11, 9136. https://doi.org/10.1038/s41598-021-88315-3

Potapov, V., Kaplan, J.B., Keating, A.E., 2015. Data-Driven Prediction and Design of bZIP Coiled-Coil Interactions. PLOS Comput. Biol. 11, e1004046. https://doi.org/10.1371/journal.pcbi.1004046

Rajagopala, S.V., Uetz, P., 2009. Analysis of Protein–Protein Interactions Using Array-Based Yeast Two-Hybrid Screens, in: Stagljar, I. (Ed.), Yeast Functional Genomics and Proteomics, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 223–245. https://doi.org/10.1007/978-1-59745-540-4_13

Rao, V.S., Srinivas, K., Sujini, G.N., Kumar, G.N.S., 2014. Protein-Protein Interaction Detection: Methods and Analysis. Int. J. Proteomics 2014, 1–12. https://doi.org/10.1155/2014/147648

Rouet, R., Jackson, K.J.L., Langley, D.B., Christ, D., 2018. Next-Generation Sequencing of Antibody Display Repertoires. Front. Immunol. 9, 118. https://doi.org/10.3389/fimmu.2018.00118

Scanlon, T.C., Gray, E.C., Griswold, K.E., 2009. Quantifying and resolving multiple vector transformants in S. cerevisiae plasmid libraries. BMC Biotechnol. 9, 95. https://doi.org/10.1186/1472-6750-9-95

Shivhare, D., Musialak-Lange, M., Julca, I., Gluza, P., Mutwil, M., 2021. Removing auto-activators from yeast-two-hybrid assays by conditional negative selection. Sci. Rep. 11, 5477. https://doi.org/10.1038/s41598-021-84608-9

Smits, A.H., Vermeulen, M., 2016. Characterizing Protein–Protein Interactions Using Mass Spectrometry: Challenges and Opportunities. Trends Biotechnol. 34, 825–834. https://doi.org/10.1016/j.tibtech.2016.02.014

Thomas, F., Boyle, A.L., Burton, A.J., Woolfson, D.N., 2013. A Set of *de Novo* Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. J. Am. Chem. Soc. 135, 5161–5166. https://doi.org/10.1021/ja312310g

Thompson, K.E., Bashor, C.J., Lim, W.A., Keating, A.E., 2012. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. ACS Synth. Biol. 1, 118–129. https://doi.org/10.1021/sb200015u

Trigg, S.A., Garza, R.M., MacWilliams, A., Nery, J.R., Bartlett, A., Castanon, R., Goubil, A., Feeney, J., O'Malley, R., Huang, S.-S.C., Zhang, Z.Z., Galli, M., Ecker, J.R., 2017. CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. Nat. Methods 14, 819–825. https://doi.org/10.1038/nmeth.4343

Velásquez-Zapata, V., Elmore, J.M., Banerjee, S., Dorman, K.S., Wise, R.P., 2021. Next-generation yeast-two-hybrid analysis with Y2H-SCORES identifies novel interactors of the MLA immune receptor. PLOS Comput. Biol. 17, e1008890. https://doi.org/10.1371/journal.pcbi.1008890

Weile, J., Sun, S., Cote, A.G., Knapp, J., Verby, M., Mellor, J.C., Wu, Y., Pons, C., Wong, C., Lieshout, N., Yang, F., Tasan, M., Tan, G., Yang, S., Fowler, D.M., Nussbaum, R., Bloom, J.D., Vidal, M., Hill, D.E., Aloy, P., Roth, F.P., 2017. A framework for exhaustively mapping functional missense variants. Mol. Syst. Biol. 13, 957. https://doi.org/10.15252/msb.20177908

Weimann, M., Grossmann, A., Woodsmith, J., Özkan, Z., Birth, P., Meierhofer, D., Benlasfer, N., Valovka, T., Timmermann, B., Wanker, E.E., Sauer, S., Stelzl, U., 2013. A Y2H-seq approach defines the human protein methyltransferase interactome. Nat. Methods 10, 339–342. https://doi.org/10.1038/nmeth.2397

Yachie, N., Petsalaki, E., Mellor, J.C., Weile, J., Jacob, Y., Verby, M., Ozturk, S.B., Li, S., Cote, A.G., Mosca, R., Knapp, J.J., Ko, M., Yu, A., Gebbia, M., Sahni, N., Yi, S., Tyagi, T., Sheykhkarimli, D., Roth, J.F., Wong, C., Musa, L., Snider, J., Liu, Y.-C., Yu, H., Braun, P., Stagljar, I., Hao, T., Calderwood, M.A., Pelletier, L., Aloy, P., Hill, D.E., Vidal, M., Roth, F.P., 2016. Pooled-matrix protein interaction screens using Barcode Fusion Genetics. Mol. Syst. Biol. 12, 863. https://doi.org/10.15252/msb.20156660

Yang, F., Lei, Y., Zhou, M., Yao, Q., Han, Y., Wu, X., Zhong, W., Zhu, C., Xu, W., Tao, R., Chen, X., Lin, D., Rahman, K., Tyagi, R., Habib, Z., Xiao, S., Wang, D., Yu, Y., Chen, H., Fu, Z., Cao, G., 2018. Development and application of a recombination-based library versus library high- throughput yeast two-hybrid (RLL-Y2H) screening system. Nucleic Acids Res. 46, e17. https://doi.org/10.1093/nar/gkx1173

Yang, J.-S., Garriga-Canut, M., Link, N., Carolis, C., Broadbent, K., Beltran-Sastre, V., Serrano, L., Maurer, S.P., 2018. rec-YnH enables simultaneous many-by-many detection of direct protein–protein and protein–RNA interactions. Nat. Commun. 9, 3747. https://doi.org/10.1038/s41467-018-06128-x

Younger, D., Berger, S., Baker, D., Klavins, E., 2017. High-throughput characterization of protein-protein interactions by reprogramming yeast mating. Proc. Natl. Acad. Sci. U. S. A. 114, 12166–12171. https://doi.org/10.1073/pnas.1705867114

Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X., Sahalie, J., Salehi-Ashtiani, K., Hao, T., Cusick, M.E., Hill, D.E., Roth, F.P., Braun, P., Vidal, M., 2011. Next-generation sequencing to generate interactome datasets. Nat. Methods 8, 478–480. https://doi.org/10.1038/nmeth.1597

Zhao, L., Liu, Z., Levy, S.F., Wu, S., 2018. Bartender: a fast and accurate clustering algorithm to count barcode reads. Bioinformatics 34, 739–747. https://doi.org/10.1093/bioinformatics/btx655

## Methods

### Experimental workflow

**Plasmids and fragments**

All protein binder coding sequences were ordered as double-stranded gene fragments from Twist Biosciences. In the gene fragment, the binder's sequence is preceded by the coding sequence for a peptide linker, either the poly-Glycine-Serine (G4S) linker or the Glycine-Glycine Poly-Serine (GGS4) linker for the DBD and AD fusion proteins, respectively. This sequence also serves as a primer handle for initial fragment amplification and a homology sequence during plasmid assembly in yeast. The binder's sequence is followed by a stop codon and a synthetic terminator from (Curran et al., 2015). Tsynth23 and Tsynth27 are used to terminate the transcription of DBD and AD fusion proteins, respectively. The terminators are followed by 21nt PCR handles for barcode amplification and predetermined 20nt long DNA barcode sequences. Barcodes were followed by 22nt long insulation sequences. This insulation sequence serves as a homology sequence during plasmid assembly in yeast.

The plasmid vector contains 30nt end homologies with the G4S peptide linker and GGS4 peptide linkers in the binder fragments for assembly in yeast. The vector also contains the AD (VP16) and the DBD (zinc finger DNA-binding domain of mouse transcription factor Zif268). An ColE1 origin and beta lactamase expression for antibiotic resistance are also included to enable cloning in E. coli for troubleshooting. The vector uses a CEN/ARS sequence for yeast replication, expressing the plasmid genes on a stable, independent mini chromosome. TRP1 expression is included for tryptophan selection.

**Competent Yeast Cell Preparation**

The yeast transformation protocol from (Benatuil et al., 2010) was adapted for this workflow. *S. cerevisiae* cells (EBY100) were grown overnight to stationary phase (OD600 around 3.0) in YPD media (10 g/L yeast nitrogen base, 20 g/L Peptone, and 20 g/L D-(+)-Glucose) on a platform shaker at 225 rpm and 30 °C. The following morning, we inoculated 100 mL of YPD media with the overnight culture at an initiate 0.3 OD600. The inoculated cells were grown on a platform shaker at 30 °C and 225 rpm until OD600 was approximately 1.6 (or for about 5 hours). Yeast cells were collected by centrifugation at 3000 rpm for 3 minutes and washed twice with 50 mL ice-cold water and once with 50 mL of ice-cold electroporation buffer (1 M Sorbitol / 1 mM CaCl2). Cells were resuspended in 20 mL 0.1 M LiAc/10 mM DTT in a culture flask and incubated for 30 mins at 30 °C and 225 rpm. We conditioned the cells by collecting them with centrifugation, washing once with 50 mL ice-cold electroporation buffer, and resuspending in 100 to 200 µL electroporation buffer to reach a final volume of 1 mL. This corresponds to approximately $1.6 \times 10^9$ cells/mL and is sufficient for two electroporation reactions of 400 µL each. The cell culture and preparation can be proportionally scaled up if more electrocompetent cells are needed to make larger or more libraries.

**Library Transformation**

Prepared electrocompetent yeast cells were kept on ice until electroporation. We prepared the DNA by combining 4 µg of digested vector backbone and 12 µg of each DNA insert for each 400 µL

electroporation reaction. The DNA mixture was reduced via precipitation and resuspension as necessary to reach a volume of less than 50 µL. We combined 400 µL of electrocompetent cells and the DNA mixture and transferred it to a pre-chilled BioRad GenePulser cuvette (0.2 cm electrode gap), then put it on ice for 5 minutes. Cells were electroporated at 2.5 kV and 25 µF. Typical time constant ranges ranged from 3.0 to 4.5 milliseconds. Electroporated cells were transferred from each cuvette into 8 mL of 1:1 mix of 1 M sorbitol: YPD media in a culture flask and incubated on a platform shaker at 225 rpm and 30°C for 1 hour. Cells were collected by centrifugation and resuspended in synthetic complete media lacking tryptophan amino acid, SC-TRP media, (20 g/L glucose, 6.7 g/L yeast nitrogen base without amino acids, 5.4 g/L $Na_2HPO_4$, 8.6 g/L $NaH_2PO_4 \cdot H_2O$ and 5 g/L casamino acids [CSM-TRP]). 250 mL SC-TRP media were used for every 400 µL transformation reaction. The optical density value OD600 of resuspended cells varied between 0.3 and 0.4 depending on the experiment. A small aliquot of the resulting cell suspension was diluted $10^{-3}$, $10^{-4}$, $10^{-5}$, and $10^{-6}$. We plated 100 µL of each dilution onto SC-TRP agar plates and incubated at 30 °C in a static incubator. Library size was determined from the colony counts on the agar plates after three days of incubation. The remaining transformed cells in SC-TRP media were grown in a baffled flask overnight at 225 rpm and 30 °C. The resulting transformed library was glycerol stocked the following day after about 20 hours of incubation at OD600 of 1.0 by combining 750µL of culture with 750µL of 50% v/v sterile glycerol in a cryovial. Glycerol stock was stored at -80°C.

**Library Selection**

After the 20-hour incubation period above, the passage process was started by diluting 20-hour transformed library cell culture in a clean autoclaved baffled flask with fresh 250 mL of SC-TRP media at a 1:400 ratio and incubating at 30 °C and 225 rpm. Alternatively, the transformed library can be inoculated from a glycerol stock at the same dilution and final volume. The cells were grown to approximately 1.0 OD600 value at which point the dilution procedure was repeated as needed with the same parameters until 36 hours had passed.. This passage process allows yeast cells to drop redundant plasmid copies and achieve an average of a single plasmid per cell. At the end of the passage process, a fraction of cells was inoculated in a baffled flask with SC-HIS-TRP media (20 g/L glucose, 6.7 g/L yeast nitrogen base without amino acids, 5.4 g/L $Na_2HPO_4$, 8.6 g/L $NaH_2PO_4 \cdot H_2O$ and 5 g/L casamino acids [CSM-TRP-HIS]) at a 1:400 ratio to start selecting for interactions. The remaining SC-TRP cells were collected by centrifugation at 5000rpm (3000xg) for 5 minutes and resuspended in 200 µLof Zymo Research Solution 1 Digestion Buffer per 50 mL of cells and stored in -80°C freezer until plasmid extraction. At least 100-fold over library size number of cells were collected. The selection for interaction step usually lasted between 24 and 36 hours depending on the library and concluded when cells in the SC-HIS-TRP flask reached OD600 value of 1.0. At that point, cells were collected by centrifugation, resuspended in Solution 1 buffer, and frozen at -80°C.

**Yeast Plasmid Extraction Protocol**

Cells after TRP selection and HIS selection were collected by centrifugation as described in the library selection protocol. Cells were resuspended in 200 µL Zymo Research Solution 1 Digestion Buffer per 50 mL of cell culture. Resuspended cells were frozen at -80°C. We then thawed the cells and added 20 µL of Zymolyase per each 200 µL volume of cells. Cells were then incubated at 37°C for 4 hours, mixing once

per hour by inverting the tubes ~ 20 times each. Cells were again frozen at -80°C and thawed. 200 µL of Zymo Research Solution 2 Lysis Buffer was added to each sample, inverted to mix, and incubated at room temperature for 3-5 minutes. We then added 400 µL Zymo Research Solution 3 Neutralizing Buffer to each sample and inverted to mix. Lysed cell were centrifuged at max speed (about 17,000 xg) for 5 minutes. Supernatant was transferred to a new microcentrifuge tube and spun again at 17,000 xg for 5 minutes. Supernatant was transferred to Zymo-Spin I Columns in a collection tube and centrifuged at 13000 xg for 30 seconds. Flowthrough was discarded. Columns were then washed by adding 550 µL of Zymo Research DNA Wash Buffer onto the column and centrifuged for 1-2 minutes at 13,000 xg. Flowthrough was discarded and wash was repeated. Empty column was spun for 1 min at 13,000 xg to dry. DNA was then eluted by adding 20 µL of molecular grade water to the column, allowing it to incubate for 5 minutes at room temperature, and spinning for 1 min at 13000 xg. Eluate was then reloaded onto the column and the spin was repeated.

**Quantitative PCR (qPCR) and Next Generation Sequencing Preparation**

Using plasmid as the template DNA, the first round of qPCR amplified the barcode region and added a 20nt long unique molecular identifier (UMI) and Illumina Read1 and Read2 primer sites. KAPA HiFi HotStart ReadyMix and suggested reagent and PCR cycle parameters were used. The annealing temperature was 67°C. The reaction was allowed to proceed until the qPCR curve reached an inflection point, and the resulting fragments were purified using KAPA Pure Beads. The resulting product was used as the template for the second stage of qPCR to add sample indexes and P5/P7 flow cell adaptor sites. This reaction was allowed to proceed for 6-7 cycles and used the same parameters as the previous qPCR reaction. An Agilent real-time PCR machine was used for all qPCR steps.

## Analysis Workflow

**Barcode counting**

Fastq files were obtained from Illumina NextSeq 550 or MiSeq sequencers using bcl2fastq with the default parameters. Cutadapt (Martin, 2011) was then used to trim out fastq barcode regions, and barcodes were clustered using Bartender (Zhao et al., 2018) with d=2.

**Autoactivator screening**

We defined autoactivators as proteins for which the trimmed interquartile mean of the library normalized enrichments is an extreme positive outlier (interquartile range > 3) with respect to the other assayed proteins. Autoactivator screening was carried out for both possible domain fusions.

**Autotune autoactivator correction**

Following autoactivator detection, the non-autoactivating alternate fusion ordering data was used to infill the pre-His selection read count according to Equation 1 and the post-His barcode read count according to Equation 2.

$$His^{+}_{new} = \left\lceil \frac{His^{+}_{FNAi}}{\Sigma_{FNA} His^{+}_{FNA}} \Sigma_{FA} His^{+}_{FA} \right\rceil \tag{1}$$

$$His^{-}_{new} = \left\lceil \left( \frac{\Sigma_{FA} His^{-}}{\Sigma_{FA} His^{+}} \right) E_{FNAi} His^{+}_{new} \right\rceil \tag{2}$$

Where FA values are for the fusion order values which has autoactivation, FNA are for the fusion order values which does not have autoactivation, and *i* is the PPI in question. The summation of the His counts for selection are done pre-update of all new barcode count corrections. Note that since there is only set of values for homodimers there can be corrections made.

For Autotune of homodimers, Equation 3 is used to infill post-His selection values for both pseudoreplicates, using the same starting pre-His selection barcode count.

$$His^{-}_{F} = \left\lceil \left( \frac{\Sigma_{F} His^{-}}{\Sigma_{F} His^{+}} \right) E_{Fi} His^{+}_{i} \right\rceil \tag{3}$$

Where *F* is the fusion order being used to construct the pseudoreplicate, and *i* is the homodimer PPI.

**Undetected PPI filtering.**

For all experiments, if a PPI had a barcode count of 0 in the pre-His and post-His selection stages, it was considered undetected. Otherwise, if it was only undetected in one stage, the other was infilled using the corresponding minimum (missing pre-His reads were infilled using the minimum of detected pre-HIs reads, etc.). For any missing PPIs, the largest possible set of replicates were used during data processing for DESeq2 analysis.

**DeSeq2 Settings.**

DESeqDataSet objects were prepared with raw pre and post-His read counts, without infilling. A contrast using the pre and post-His columns as the conditions was used, and ashr lfcShrink was used to get final Mp3-seq LFC values. For pseudoreplicate MP3-seq, pre and post-His conditions were used along with a fusion order batch effect for a contrast of ~cond + batch for DESeqDataSet creation. P-LFCs were obtained by ashr lfcShrink.

**DUET.**

A Pre-DUET weighted undirected graph *G* was created from MP3-seq values with $p_{adj} \leq 0.01$ and P-LFC > 0, and vertices with degree 0 removed. Then, until $|G|=0$ or $\Delta(G)=1$, all $G \setminus \{v\}$ were scored for $v \ni V$ with Equation 4, using $c=4$ and the highest scoring vertex removal was used to reduce the graph. The greedy reduction algorithm pseudocode can be seen in **Figure 6**.

(4)

$$score(G,e,c) = \begin{cases} w_{v_1,v_2}, & \text{if } deg(v_1) = deg(v_2) = 1 \\ w_{v_1,v_2}, & \text{if } deg(v_1) = 1 \text{ and } deg(v_2) \neq 1 \\ & \text{ and } max(v_i \ni N(v_2) \setminus v_1 : w_{v_i,v_2}) \leq cw_{v_1,v_2} \\ -w_{v_1,v_2}, & \text{if } deg(v_1) = 1 \text{ and } deg(v_2) \neq 1 \\ & \text{ and } max(v_i \ni N(v_2) \setminus v_1 : w_{v_i,v_2}) > cw_{v_1,v_2} \\ w_{v_1,v_2}, & \text{if } deg(v_2) = 1 \text{ and } deg(v_1) \neq 1 \\ & \text{ and } max(v_i \ni N(v_1) \setminus v_2 : w_{v_i,v_1}) \leq cw_{v_1,v_2} \\ -w_{v_1,v_2}, & \text{if } deg(v_2) = 1 \text{ and } deg(v_1) \neq 1 \\ & \text{ and } max(v_i \ni N(v_1) \setminus v_2 : w_{v_i,v_1}) > cw_{v_1,v_2} \\ w_{v_1,v_2}, & \text{if } deg(v_1) \neq 1 \text{ and } deg(v_2) \neq 1 \\ & \text{ and } max(v_i \ni N(v_2) \setminus v_1 : w_{v_i,v_2}) \leq cw_{v_1,v_2} \\ & \text{ and } max(v_i \ni N(v_1) \setminus v_2 : w_{v_i,v_1}) \leq cw_{v_1,v_2} \\ -w_{v_1,v_2}, & \text{if } deg(v_1) \neq 1 \text{ and } deg(v_2) \neq 1 \\ & \text{ and } max(v_i \ni N(v_2) \setminus v_1 : w_{v_i,v_2}) > cw_{v_1,v_2} \\ & \text{ or } max(v_i \ni N(v_1) \setminus v_2 : w_{v_i,v_1}) > cw_{v_1,v_2} \\ 0 & \text{else} \end{cases} \quad \text{where } v_1, v_2 \ni e$$

$G = G \setminus \{v \ni V(G) : deg(v) = 0\}$
while $|G| \neq 0$ and $\Delta(G) > 1$ :
$\quad v_{rem} = max_{\{v \ni V(G)\}} \left(\Sigma_{e \ni E(G \setminus \{v\})} score(G \setminus \{v\}, e, c)\right)$
$\quad G = G \setminus \{v_{rem}\}$
$\quad G = G \setminus \{v \ni V(G) : deg(v) = 0\}$

**Figure 6. DUET pseudocode.**

**Orthogonality gap reduction.**

To reduce the DUET pairs, any P-LFC>0 interaction between the set of proteins involved in the DUET pairs was considered a real interaction. Non DUET-pair interactions were sorted by P-LFC, and the largest value interaction was selected to remove a DUET-pair. Whichever protein was involved with the smaller P-LFC DUET-pair and its DUET partner was removed. This was done one protein pair at a time until only one pair remained.

## Computational protein design

**Design of helical bundles**

Helical bundles were designed using methods developed in (Chen et al., 2019). Briefly, systematic sampling of Crick coiled-coil parameters was used to independently sample all helices of the

heterodimers supercoiled around the same axis. The Rosetta HB-Net protocol (Boyken et al., 2016) was used to design buried HB-Nets in the core of the designs.

**Hydrogen bond network removal**

The design model of mALB8 was used as a starting point. The two hydrogen bond networks were manually annotated (R1 has residues 23+40+84+96, R2 has residues 16+43+103), see **Figure S5**. Two alternative protocols were used for each redesign. Either a FastRepack with a fixed static backbone, which results in replacement with smaller amino acid residues (T, A), or a FastDesign protocol, where the backbone was allowed to move, which resulted in replacement with larger amino acid residues (V, L, F).

The residues of the HB-net and all residues within a 6 A shell were redesigned, and a shell of 8 was repacked. The beta_nov2016 score function was used. The charge was fixed to -5 for each chain. Five sequences were made for each design. The sequences were filtered by score_per_res>-3.5, shape complementarity >0.7, sbuns<2, and ranked by ddg (the lowest ddg was chosen for the final design). Rosetta version 2021-07 (c04b624d89b45b55f573bbf2f6685ffbf2e4fcaf) was used for the calculations.

**Structure visualization**

**Visualization of 3D structures.** All structural images were generated using PyMOL version 2.5.2.

**Structure prediction.** For non-DHD proteins without PDB codes noted, structures for visualization purposes were generated with ESMFold (Lin et al., 2022).

## Data and code availability

Barcode count data and scripts necessary to recreate the MP3-seq pipelines and analysis can be found at https://github.com/Seeliglab/MP3-DUET-AUTOTUNE

## Supplementary Figures

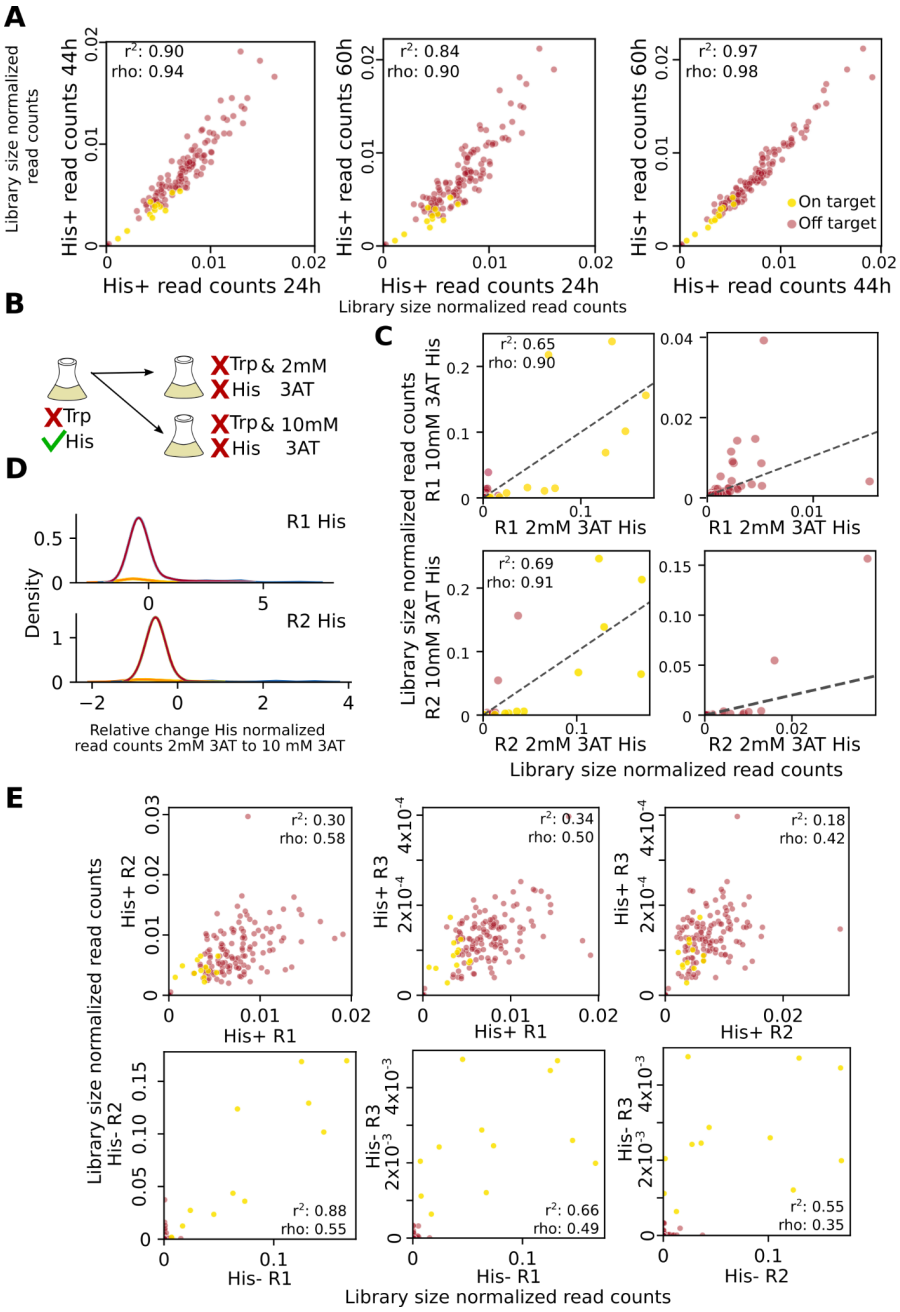



**Figure S1. Confirming pre-His and post-His selections proceeded as expected. A.** Pre-His selection barcode counts hold steady over three samples taken from a P-series all by all screen at 24, 44, and 60 h. **B.** Two concentrations of 3-AT were tested during selection for two P-series all-by-all replicates. **C.** Higher concentration 3-AT post-His selection barcode counts are lower when normalized to library size compared to lower 3-AT selection, as is expected. **D.** Relative changes between the two 3-AT conditions for both replicates. **E.** Correlation of normalized barcode counts for pre and post-His selection conditions. Note the drastic increase in on-target P-series normalized barcode counts between conditions, and that overall, correlation is weaker pre-His than post-His selection.

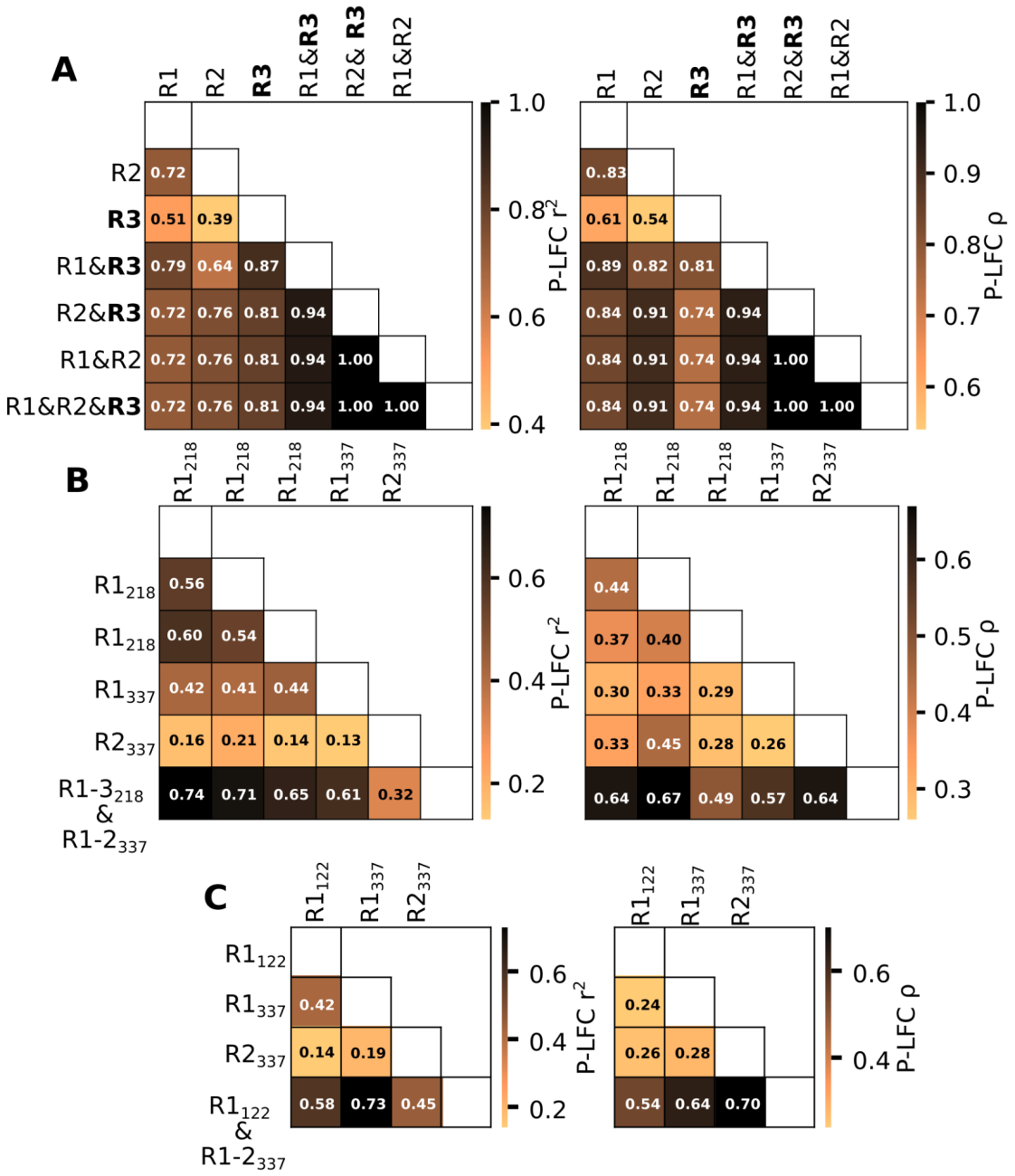



**Figure S2. Correlations of P-LFCs for individually processed and merged replicates. A.** Correlations of BCL-2 and inhibitor P-LFC for replicates and combined replicates. R3 was collected with the final assay version, while R1 and R2 were collected with the earlier. **B.** Correlations between replicates and combined replicates for the DHD0/DHD1 proteins. Subscripts denote size of the library for all the all-by-all screen. **C.** Correlations between replicates and combined replicates for the DHD0/DHD2/mALb proteins. Subscripts denote size of the library for all the all-by-all screen.

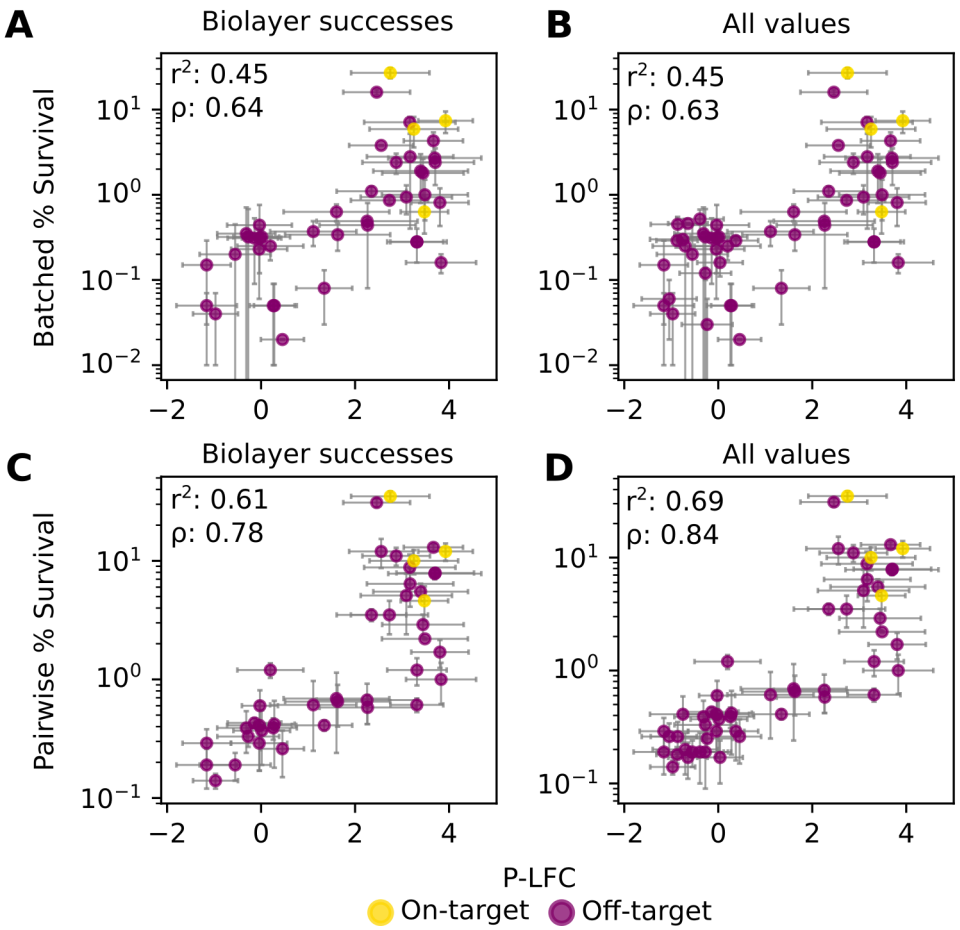



**Figure S3. Alpha-Seq correlations with MP3-seq**. **A.** MP3-Seq P-LFC versus batched survival for only interactions which succeeded biolayer interferometry **B**. MP3-Seq P-LFC versus batched survival for all BCL-2 homolog and inhibitor interactions **C.** MP3-Seq P-LFC versus pairwise survival for only interactions which succeeded biolayer interferometry **D.** MP3-Seq P-LFC versus pairwise survival for all BCL-2 homolog and inhibitor interactions

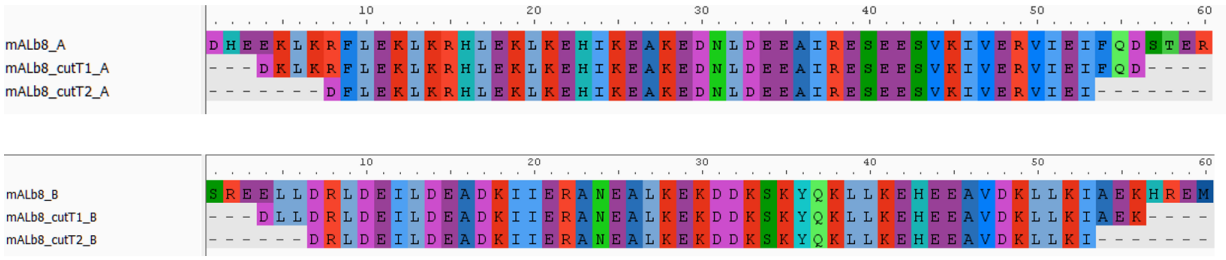
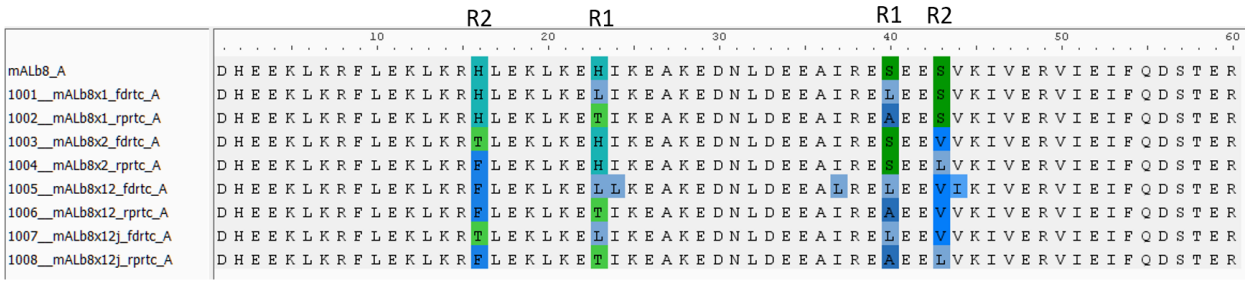



**Figure S4: Alignments of shorter mALb8 variants.** Original mALb8 protomers at top of rows, followed by T1 and T2 variants



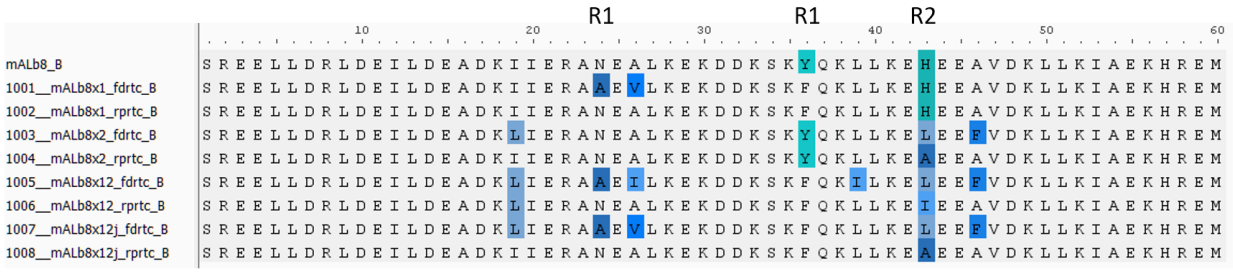




**Figure S5: Sequences of mALB8 with systematically removed hydrogen networks.** Original mALb8s are at tops of the rows. Names with x1 correspond to R1, x2 to R2, and x12 to R12. The small replacement protocol is rprtc and large replacement protocol is fdrtc.