Article

Massively parallel measurement of protein–protein interactions by sequencing using MP3-seq

Alexandr Baryshev[1,8], Alyssa La Fleur[2,8], Benjamin Groves[1], Cirstyn Michel[3], David Baker [3,4,5,6], Ajasja Ljubetič [4,5,7] ✉ & Georg Seelig [1,2] ✉

Protein–protein interactions (PPIs) regulate many cellular processes and engineered PPIs have cell and gene therapy applications. Here, we introduce massively parallel PPI measurement by sequencing (MP3-seq), an easy-to-use and highly scalable yeast two-hybrid approach for measuring PPIs. In MP3-seq, DNA barcodes are associated with specific protein pairs and barcode enrichment can be read by sequencing to provide a direct measure of interaction strength. We show that MP3-seq is highly quantitative and scales to over 100,000 interactions. We apply MP3-seq to characterize interactions between families of rationally designed heterodimers and to investigate elements conferring specificity to coiled-coil interactions. Lastly, we predict coiled heterodimer structures using AlphaFold-Multimer (AF-M) and train linear models on physics-based energy terms to predict MP3-seq values. We find that AF-M-based models could be valuable for prescreening interactions but experimentally measuring interactions remains necessary to rank their strengths quantitatively.

Synthetic protein binders mediating interactions between other proteins or cells can potentially revolutionize fields ranging from cell therapy[1] to synthetic biology[2–5] and material science[6–8]. Early work often relied on natural interaction domains such as SH3 and PDZ (ref. 9). However, such components provide a poor starting point for rationally designing large-scale assemblies because of crosstalk. Consequently, synthetic protein circuits remain much smaller and simpler than biological protein–protein interaction (PPI) networks. To scale up synthetic protein-based circuits, we need large-scale libraries of modular interaction domains. Ideally, these interaction domains would be orthogonal, whereby each domain interacts only with its designated binding partner.

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 (1×1s); 1×1 coil binding is primarily determined by complementary electrostatic and hydrophobic interactions at specific heptad positions. Because the biophysical rules guiding these interactions are relatively well understood and can be captured by predictive models[10,11], orthogonal sets of up to six 1×1 heterodimers have been generated[12,13]. Recently, a set of orthogonal 1×1s (primarily homodimers) was identified in a high-throughput screen[14]. However, the restricted geometry required for 1×1 coil interactions limits the number of possible orthogonal interactions. The Baker lab expanded the coiled-coil toolbox by introducing helical bundle heterodimers, with protomers consisting of two helices connected by a hairpin loop, where interactions are determined by designed hydrogen-bond networks (HB-nets) between the bundles (2×2s)[15,16]. This multihelix bundle approach could pave the way to generating larger orthogonal sets. However, reliably

[1]Department of Electrical & Computer Engineering, University of Washington, Seattle, WA, USA. [2]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [3]Department of Bioengineering, University of Washington, Seattle, WA, USA. [4]Department of Biochemistry, University of Washington, Seattle, WA, USA. [5]Institute for Protein Design, University of Washington, Seattle, WA, USA. [6]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [7]Department for Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia. [8]These authors contributed equally: Alexandr Baryshev, Alyssa La Fleur. ✉e-mail: ajasja.ljubetic@ki.si; gseelig@uw.edu

minimizing off-target interactions (for example, because of proteins associating in an unexpected register or orientation) remains challenging as these are not captured in the biophysical design objective to produce 2×2s. A practical alternative is to prepare diverse de novo 2×2 protomer libraries and measure an all-by-all matrix of interactions. Afterward, an orthogonal set can be extracted[17].

Many PPI measuring methods have been developed that could be used to run all-by-all PPI screens with varying throughput levels. Mass-spectrometry-based and protein-array-based methods can measure PPIs but require laborious protein purification steps[18,19]. Yeast-display or phage-display methods use next-generation sequencing technologies to increase throughput but are limited to 'several-versus-many' screening[20]. Alpha-seq, a recent high-throughput method that takes advantage of the yeast mating pathway, overcomes this limitation and allows library-on-library screening[21]. However, throughput is still limited and not all proteins correctly fold when displayed on the yeast surface.

Yeast two-hybrid (Y2H) methods are a powerful alternative to surface display for characterizing PPIs[22]. In Y2H, one protein is fused to a DNA-binding domain (DBD) and the second 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[23,24] 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[25–34]. Concomitantly with experimental methods, custom workflows for analyzing HT-Y2H data were developed[35,36]. Most of the experimental methods require library construction in *Escherichia coli* or rely on yeast mating and, thus, need separate protein libraries to transform MATα and MATa yeast. High-throughput bacterial two-hybrid assays have been developed, which provide a noneukaryotic alternative for screening PPIs that avoid yeast mating and library transfer[14]. However, when testing protein interactions, it can be desirable to have an environment similar to their native context. Therefore, using Y2H-based assays to test interactions of naturally intracellular proteins with eukaryotic origins can be beneficial. Specifically, expressing such proteins in yeast allows for more accurate folding, the introduction of post-translational modifications and increased solubility when compared to bacterial expression[35–37].

Deep learning models such as AlphaFold2 (AF2)[38] and RoseTTAFold[39] have been shown to predict protein structures with near-experimental accuracy. Structure prediction for proteins with multiple chains is possible, either with the original AF2 model or with specialized models such as AF-Multimer (AF-M)[38]. Given the speed at which these models operate compared to running Y2H assays, determining how these models can be used to prescreen PPIs or supplement PPI assays is attractive. Tools have been developed to run AF-M in an all-by-all manner to assist with PPI prescreening[40] and AF-M error metrics have been demonstrated to be state-of-the-art protein–peptide interaction predictors[41]. However, it remains unclear how well AF-M can predict protein binding orthogonality.

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 pair abundance before and after a selection experiment serves as a proxy for interaction strength. Plasmids are assembled in yeast through homologous recombination encoding the protein pairs of interest, their associated barcodes and all other elements required for Y2H experiments. Our workflow bypasses the need for plasmid cloning in *E. coli* or yeast mating and proteins fold and interact inside the cell instead of on the surface. In contrast to bacterial assays, MP3-seq also works with glycosylated proteins. We validate MP3-seq using well-characterized coiled-coil heterodimer interactions and synthetic

binders for different human B cell lymphoma 2 (Bcl-2) protein family members. Then, we apply MP3-seq to characterize interactions between rationally designed 2×2 and 1×2 heterodimers, demonstrating that MP3-seq can measure over 100,000 PPIs in a single experiment. We identified successful designs and used a greedy algorithm to find potentially orthogonal subsets. We delved into the elements expected to confer 2×2 interaction specificity by screening variations of a successful 2×2 pair. Lastly, we predict complexes for coiled-coil dimers with AF2 and AF-M and assess the ability of these models to predict interactions and orthogonality. We then use AF-M error values and physics-based structure energy terms from Rosetta to train simple models to predict MP3-seq results to investigate how complex predictors could supplement high-throughput measurements.

Results

MP3-seq workflow

In MP3-seq, all molecular components required for measuring interactions between two specific proteins are encoded on a single plasmid (Fig. 1a). A plasmid library is constructed directly through homologous recombination in yeast to measure all possible interactions for a set of proteins. Haploid MATa-type yeast is first transformed with a mixture of DNA fragments. One fragment is a backbone carrying a centromere (CEN) sequence, the selection marker, a DBD and an AD. We use the Cys(2) His(2) zinc-finger domain of the mouse transcription factor Zif268 (ref. 42) 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 contain one protein of interest and its associated barcode separated by a terminator sequence. Because of their short length and distinct sequences, we used Tsynth23 and Tsynth27 as terminators in most experiments[43]. A simplified map of the plasmid can be seen in Supplementary Fig. 1.

After transformation, we perform an extended outgrowth and selection step in medium without tryptophan to ensure plasmid maintenance. At this point, the CEN sequence ensures each yeast cell (and any subsequent daughter cells) will contain approximately one library plasmid, which is critical to link the growth of transfected yeast to barcode counts[44]. An aliquot of the cells is frozen, while the remainder undergo selection in medium lacking histidine. Plasmid DNA is extracted from His preselection and postselection cells and the barcode-containing regions are amplified (Fig. 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 (Fig. 1b). The expression or folding of some proteins may be impacted by fusion with either the AD or the DBD; therefore, we test all proteins in two fusion orders (DBD fused to P1 and AD fused to P2; DBD fused to P2 and AD fused to P1).

MP3-seq analysis pipeline

First, we calculate the enrichment between the His preselection and postselection stages for PPI fusion order (Fig. 1b). Next, we detect autoactivators and replace their His postselection values with values calculated from the nonautoactivating fusion order (see Autotune in Methods); autoactivation is an error mode in Y2H experiments where high enrichment is observed for a protein for all interaction partners, suggesting nonspecific activation of selection marker expression[45]. Typically, this behavior is observed fused to either the AD or the DBD but not in both orientations. If an autoactivator is found, its homodimer is not recoverable with Autotune, as there is no reverse fusion order to use for correction.

Following autoactivator removal, enrichment values from replicate experiments 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 the log₂ fold change (LFC) using the DESeq2 package[46]. DESeq2 calculates differential enrichment across multiple replicates
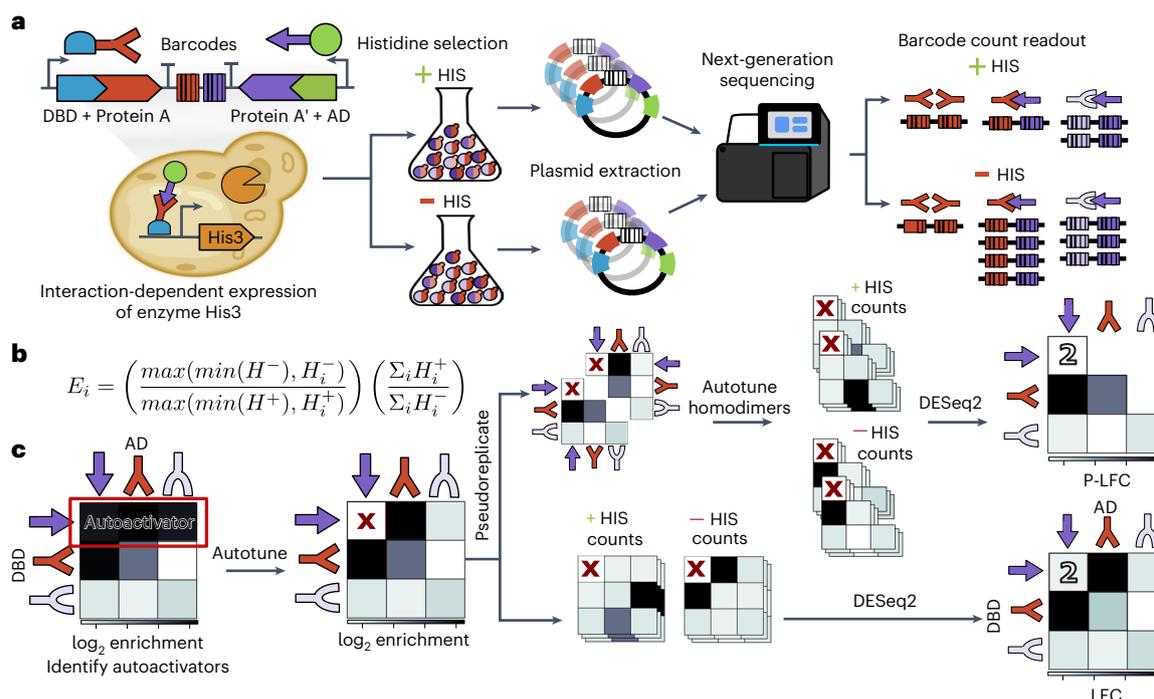
Fig. 1 | MP3-seq workflow. a, DNA-barcoded proteins A and A′ are transformed into yeast fused to an AD or DBD. The cells are incubated first in histidine-rich media and then used to inoculate histidine-poor media for growth selection. If the hybrid proteins interact, the DBD and AD form a transcription factor that drives *His3* expression and increases growth in histidine-poor media. Therefore, interacting pairs should be enriched in the population after selection. Plasmids are extracted from both cultures and barcode-containing fragments are amplified and sequenced. **b**, Enrichment (*E*) 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 (see Methods). Replicate preselection and postselection barcode counts are merged directly with DESeq2 or split into pseudoreplicates and merged to obtain the LFC or P-LFC.

and provides Hochberg-adjusted Wald test *P* values (P_{adj}), identifying PPIs with LFCs significantly different from an LFC of zero.

In some cases, combining MP3-seq measurements for both fusion orders is desirable (for example, if comparing MP3-seq LFCs to K_d values collected for the pair of interest). We treat each fusion order as an independent set of measurements, turning them into two pseudoreplicates. These pseudoreplicates are combined with DESeq2 to calculate LFCs (Fig. 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 1×1s in the National Institute of Chemistry Peptides (NICP) set[13] (Fig. 2a). A pool of 24 DNA fragments with each of the 12 proteins fused to either the AD or the DBD was ordered. All interaction pairs were assembled in a pooled experiment and interaction strengths were quantified with MP3-seq. As expected, His preselection barcode counts remained steady over time (Supplementary Fig. 2a) and applying different levels of 3-amino-1,2,4 triazole (3-AT), a competitive inhibitor of His3, affected His postselection counts (Supplementary Fig. 2b–d). LFCs for both orientations were calculated from five experimental replicates, with interactions occurring almost exclusively between designed (on-target) partners (that is, P1:P2, P3:P4, etc.; Fig. 2b). Homodimer plasmid assembly was less efficient in MP3-seq and generally resulted in lower input barcode counts but coverage was sufficient for inclusion in the analysis (Supplementary Fig. 2e). This phenomenon is likely because of increased sequence homology interfering with plasmid assembly. For a more quantitative validation, we correlated MP3-seq LFCs with luciferase expression assay interaction scores in HEK293T cells[13] and found good agreement ($r^2 = 0.74$; Fig. 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 interactions weighted by P-LFC.

In a separate all-by-all experiment with 28 proteins, we screened the NICP 1×1s and two sets of 1×1s derived from them (N1, N2, N5–N8 and P5A–P8A) but with increased thermodynamic stability of their on-target interactions[47]. In this experiment, we expect related coils to have similar interaction patterns (for example, P5A can bind to P6A, N6 and P6), which agrees with the interaction graph created from significant MP3-seq P-LFCs in Fig. 2e. The full interaction graph (Supplementary Fig. 3a) showed minimal off-target crosstalk between NCIP proteins and their variants. We found low correlation with melting temperatures collected for the interactions but good correlations with split luciferase and split transcription factor interaction assays[47] (Supplementary Fig. 3b–d).

To assess whether MP3-seq values provide quantitative information about interaction strength, we performed three all-by-all replicates for interactions between varying-length 1×1s designed to span a wide range of K_d values[48] (Fig. 2f). These interactions are not expected to be orthogonal as weaker interactions were achieved by truncating two parent four-heptad binders (A_N^4, B_N^4) by a half or full heptad[48] (Fig. 2g). The MP3-seq P-LFCs correlated very well with previously measured dissociation constants over approximately three orders of magnitude ($r^2 = 0.94$; Fig. 2h).

We explored using minimum read count filtering at different thresholds and found that filtering brought minimal or no changes to correlation with the reported K_d values. However, for the NICP series, filtering did improve correlation (particularly for Spearman's rho) (Supplementary Fig. 4a,b). Lastly, to examine the performance for weaker interactions, we compared the correlation of enrichment and P-LFC values with the K_d values for the five weakest interactions[48]. Enrichments had a slightly better ability to order weak interactions than P-LFCs, suggesting that a non-DESeq2 analysis may be better suited for the high-throughput study of weak PPIs (Supplementary Fig. 4c).
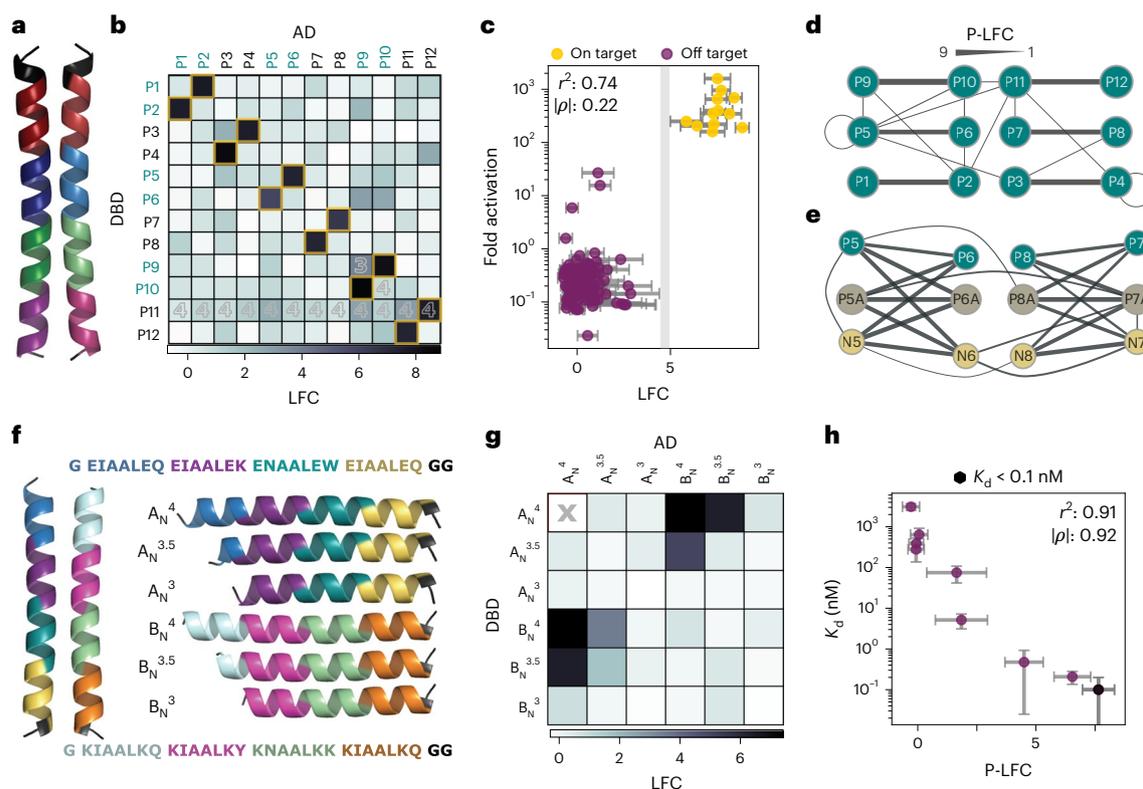
Fig. 2 | Validation of MP3-seq with coiled-coil heterodimers. a, NICP series 1×1s[13] with complementary heptads designed to interact shown in like colors. **b,** MP3-seq LFC of the NICP series interactions. All MP3-seq values were calculated from five biological replicates except for those labeled, where labels indicate the number of replicates available. Yellow outlines denote designed on-target interactions. **c,** Correlation of the on-target and off-target NICP series MP3-seq LFCs with average fold activation fluorescence values[13] ($n = 141$ PPIs; three homodimers had insufficient reads or were autotuned and were omitted). The gray bar is the gap separating on-target and off-target interactions. LFCs are

presented as the LFC ± standard error (SE). **d,e,** Filtering NICP series interactions (**d**) and P, PA and N series designed coil interactions[47] (**e**) to include only those with $P_{adj} \leq 0.05$. Line weights correspond to MP3-seq P-LFCs. **f,** A designed 1×1 and its truncations from a previous study[48]. **g,** MP3-seq enrichment for the A_N and B_N coils and their truncations from three biological replicates. A gray '×' indicates a missing interaction. **h,** Correlation of $n = 9$ A_N and B_N truncation PPI MP3-seq P-LFCs with K_d values from Thomas et al.[48]. P-LFCs are presented as the P-LFC ± SE and dissociation constants are presented as the mean ± s.d.

MP3-seq benchmarking with Bcl-2 family binders

To validate MP3-seq outside of 1×1 interactions, we tested a set of proteins previously characterized by biolayer interferometry[49] and Alpha-seq[21] composed of six homologous proteins from the Bcl-2 family (Bcl-2, Bcl-X_L, Bcl-w, Mcl-1, Bfl-1 and Bcl-B) and nine de novo designed inhibitors of said homologs. A crystal structure of a synthetic binder bound to its target is shown in Fig. 3a, while Fig. 3b shows the domain organization of the six human proteins. All inhibitors were designed to interact with the BH3 domain-binding pocket. To better compare with the surface-display-based HT-Y2H method Alpha-seq, a truncated version of Mcl-1 was used.

A two-replicate MP3-seq screen of all Bcl-2 homologs against all inhibitors is shown in Fig. 3c. Version 1 of MP3-seq was used for the first replicate of this experiment. Unlike version 2, barcodes were inserted into the 3′ untranslated regions of the binders upstream of the terminators and could, thus, differentially impact mRNA stability and protein levels. However, the inter-replicate correlations suggested consistent interactions between versions (Supplementary Fig. 5a). Only binders beginning with α corresponded to the final, specific designs, while the others were intermediate or failed designs. This can be seen in Fig. 3c, with a divide between the largely orthogonal rightmost four columns of the heat map of MP3-seq P-LFCs and the leftmost less specific columns.

Our data agree well with dissociation constants obtained from biolayer interferometry[49] ($r^2 = 0.61$; Fig. 3d). We found good agreement with Alpha-seq percentage survival for their low-throughput pairwise and high-throughput batched assays on the same interactions (batched

$r^2 = 0.45$, paired $r^2 = 0.61$, $n = 43$; Supplementary Fig. 5b–e). MP3-seq interactions are measured with proteins expressed in yeast, while biolayer interferometry uses purified proteins and Alpha-seq displays proteins on the yeast surface, partially explaining the variation between our results and those published earlier.

Some Bcl-2 inhibitors failed to produce K_d values when measured with biolayer interferometry, likely because the interactions were below detection limits. We examined the undetected ($n = 11$) versus detected interactions ($n = 43$) and found that the mean MP3-seq value of PPIs that biolayer interferometry detected was significantly greater than those that were undetected (one-tailed independent t-test, H₁: $\mu_{detected} > \mu_{undetected}$, $t = 4.51$, $P = 1.858 \times 10^{-5}$; Fig. 3e). Pairwise Alpha-seq also had a significantly greater mean of detected interactions than undetected ones (H₁: $\mu_{detected} > \mu_{undetected}$, $t = 1.91$, $P = 0.0307$). However, high-throughput batched Alpha-seq did not exhibit this behavior (H₁: $\mu_{detected} > \mu_{undetected}$, $t = 1.36$, $P = 0.0904$). MP3-seq correlation improved for batched but not paired Alpha-seq when including all data points (batched $r^2 = 0.45$, paired $r^2 = 0.69$, $n = 54$; Supplementary Fig. 5b–e) (Fig. 3f–g). 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.

Lastly, we screened a set of intrinsically disordered p53 upregulated modulator of apoptosis (PUMA) peptides that bind to Mcl-1 at the BH3 pocket (Fig. 3b). Mcl-1 binding results in the disordered PUMA BH3-binding motif transitioning to an ordered helix[50]. Peptides of BH3 motif mutants were used to investigate the effects of helicity degree on
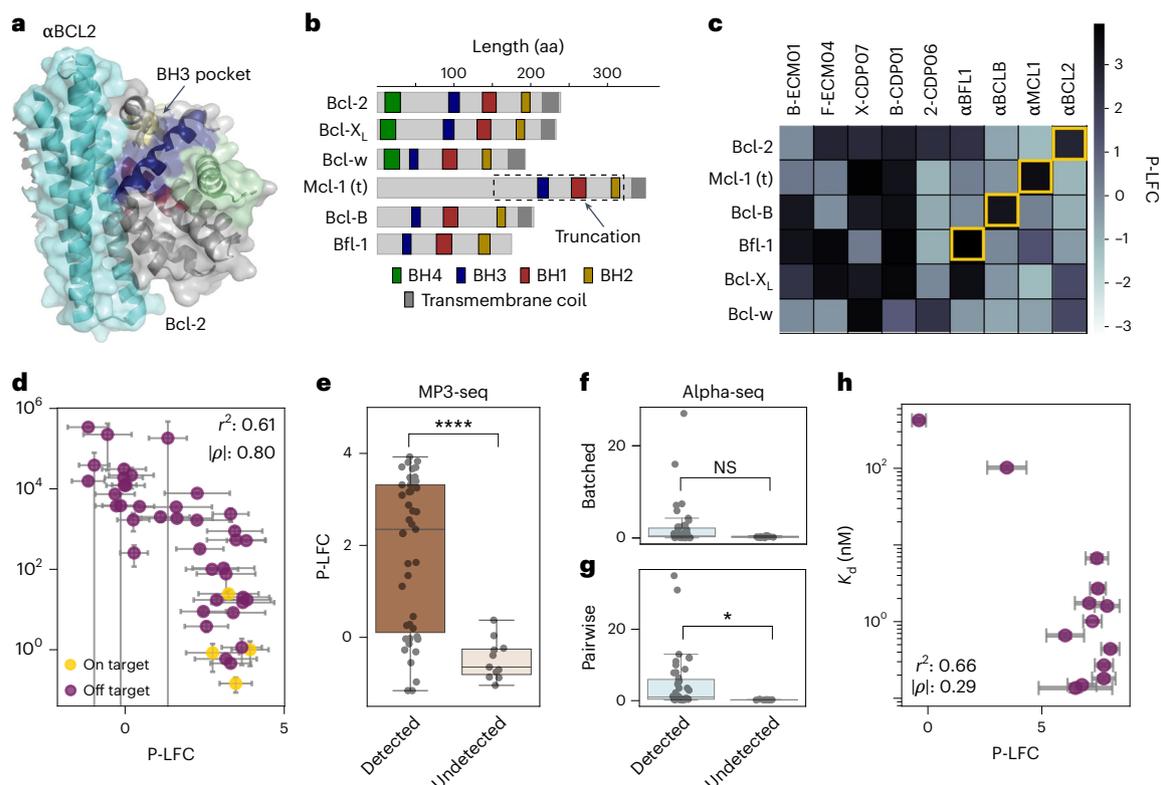
Fig. 3 | Validation of MP3-seq with Bcl-2 proteins. a, Colored crystal structure of Bcl-2 and its designed BH3-binding inhibitor[49] (Protein Data Bank 5JSN). The binder is blue and Bcl-2 is colored according to the domains. **b**, BH3-binding domain annotations of the six human Bcl-2 homologs measured[21,49], with the experimental truncation used by Alpha-seq and that in this work shown. **c**, MP3-seq P-LFCs for inhibitor and Bcl-2 interactions calculated from two biological replicates. Yellow boxes highlight intended on-target interactions. **d**, Correlations with K_d measurements from biolayer interferometry. P-LFCs are presented as the P-LFC ± SE and K_d values are presented as the mean ± s.d. Only PPIs within biolayer interferometry detection limits were used ($n = 43$). **e**, MP3-seq P-LFC distributions for interactions that were detected ($n = 43$) and undetected ($n = 11$) by biolayer interferometry because of instrument detection

limits ($K_d \geq 25$ mM). **f**, Pairwise Alpha-seq distributions. **g**, Batched Alpha-seq distributions. For all box plots in panels **e**–**g**, the whiskers are at ±1.5 times the interquartile range, while the boxes have border lines showing the first quartile, median and third quartile, with all data points for both sets shown in a swarm overlay. One-tailed independent t-tests (H1: $\mu_{detected} > \mu_{undetected}$) were used to compare the detected and undetected measurements. ****$P < 0.0001$ and *$P < 0.05$; NS, not significant ($P \geq 0.05$). Both MP3-seq P-LFC and pairwise Alpha-seq were significant for $\mu_{detected} > \mu_{undetected}$ ($P = 1.858 \times 10^{-5}$ and $P = 0.0307$, respectively), while batched Alpha-seq was not significant ($P = 0.0904$). **h**, Correlation of PUMA peptide and Mcl-1 (t) PPI ($n = 13$) P-LFCs from two biological replicates with PUMA peptide and full Mcl-1 average K_d measurements[50]. P-LFCs are presented as the P-LFC ± SE and K_d values are presented as the mean ± s.d.

this intrinsically disordered protein-dependent interaction. We found a good correlation between MP3-seq and earlier stopped-flow fluorescence measurements of the K_d values for PUMA peptides interacting with the full Mcl-1 protein ($r^2 = 0.66$; Fig. 3h).

Large-scale assay of designed heterodimers (DHDs)
To explore the feasibility of using MP3-seq for large-scale screening, we prepared a library of designed heterodimers (DHDs)[16] (Supplementary Methods). On-target pairs were designed as three-helix or four-helix bundles with buried HB-nets connecting all helices and then split into a helix–turn–helix hairpin and a single helix (1×2s) or 2×2s. Initially, we chose 100 designs for testing (DHD1; Fig. 4a), adding 52 more designs in a second round of experiments (DHD2; Fig. 4a). As controls, we selected nine previously tested 2×2s[16]. We also included 35 pairs of binders derived from these controls by modifying the hairpin loops or truncating the helices (DHD0; Fig. 4a). Finally, we added a series derived from a common parent 2×2 through truncating heptads (mALb; Fig. 4a). Sequences for DHD0–DHD2 and mALb truncations can be found in Supplementary Table 1. For each on-target pair, one protomer is designated 'A' and the other is designated '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, DHD1, DHD2 and mALb proteins, as well as the Bcl-2

homologs and their binders, resulting in a matrix of 337 × 337 = 113,569 interactions. We performed three replicates of a screen using only DHD1 with a subset of DHD0 and one with DHD0, DHD2 and mALb. MP3-seq version 1 was used for most experiments in this section. Nevertheless, the experiments highlight the large scales achievable with MP3-seq and the inter-replicate correlations suggested that the data were internally consistent (Supplementary Fig. 6).

We applied autotune corrections and calculated 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 $P_{adj} \leq 0.01$ and P-LFC ≥ 4. The number of successes out of the total possible on-target PPIs in each design set can be seen in Fig. 4b. Every design set had an approximately 20% success rate in the largest screen, consistent with past 22% success rates for α-helical bundles[51]. The top five P-LFC on-target interactions for each design category are shown in Fig. 4c.

Finding orthogonal subsets
For all four sets of designs (DHD0–DHD2 and mALb), we detected a large fraction of strong off-target interactions (for example, A1 and B2 of DHD1 instead of A1 and B1). For example, in the all-by-all screen with all of DHD0–DHD2 and mALb, only 33 of 914 total PPIs with $P_{adj} \leq 0.01$ and P-LFC ≥ 4 were on target.
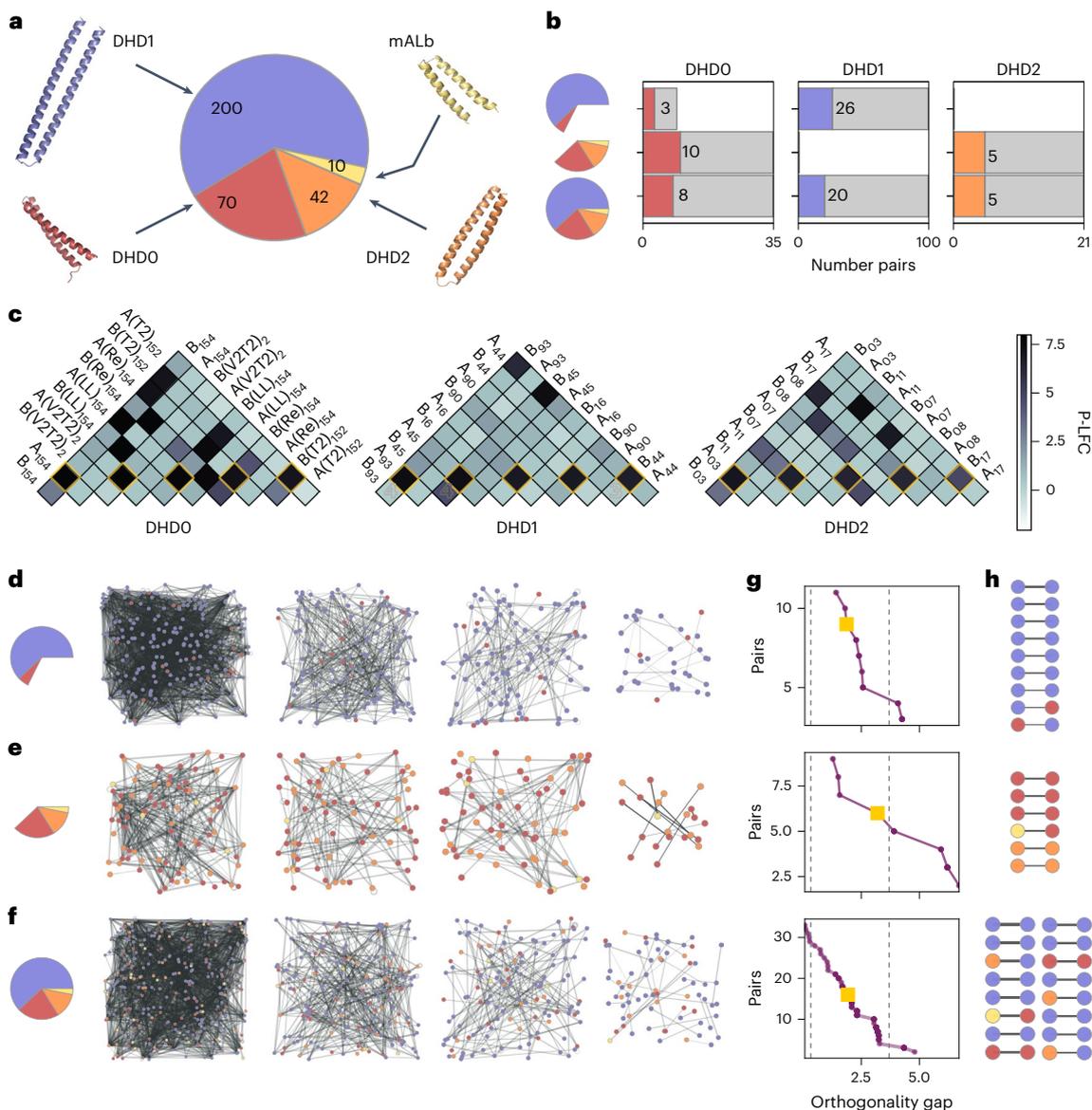
Fig. 4 | High-throughput screening of DHDs. a, Summary of the synthetic binder pair types. **b**, Successful designs for the three main sets. Gray bars are possible pairs for the set and colored bars indicate the number of successes. **c**, Top five successful designs from each set by P-LFC. Yellow boxes show on-target interactions. The DHD1 and DHD0 subset had three biological replicates, the DHD0, DHD2 and mALb subset had one biological replicate and the full set had two replicates. Biological replicates for overlapping all-by-all PPI measurements were combined using the data pipeline. **d**, From left to right, pre-DUET positive P-LFC $P_{adj} \leq 0.01$ PPI network for the DHD1 and DHD0 subset and the network

at DUET iterations 50, 75 and 113 (final). **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 (final). **g**, Orthogonality gaps of the DUET final networks without significance filtering. Left and right dashed lines show the MP3-seq orthogonality gaps for Bcl-2 and final inhibitors and the NCIP series, respectively. Yellow squares correspond to half of the starting networks remaining. **h**, DUET networks reduced to half their final iteration size.

We asked whether we could extract potential orthogonal subsets. We used significant (α = 0.01) positive P-LFC values to construct a weighted undirected graph as in Fig. 2d,e. This approach allowed us to rephrase the problem as finding a graph of degree one vertices without self-edges, which maximizes the sum of the remaining edges. To find a graph satisfying our constraints, we developed a simple scoring function that rewards graphs on the basis of existing orthogonal edges or those over a desired orthogonality gap and punishes graphs for non-orthogonal edges. The orthogonality gap is defined as the difference between the weakest on-target interaction and the strongest off-target interaction[14]. A larger gap generally results in smaller sets. 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 one-regular graph remains (Supplementary Methods) (the number of degree one vertices and score of total graphs per algorithm step can be seen in Supplementary Fig 7a,b). For the DHD1 and DHD0 subset, we went from 2,001 edges between 202 vertices to 18 DUET pairs (Fig. 4d); for the DHD0, DHD2 and mALb data, we went from 279 edges between 85 vertices to 11 DUET pairs (Fig. 4e); for all designs (DHD0–DHD2 and mALb), we went from 1,562 edges between 270 vertices to 36 DUET pairs (Fig. 4f). Of these, two, two and four 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 noninteracting. 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 protomers 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 (Supplementary Methods and Fig. 4g). When reduced to half the starting DUET pairs (squares in Fig. 4g), the orthogonal sets shown in Fig. 4h were left, which all had orthogonality gaps comparable to the NICP set in Fig. 2 and the Bcl-2 inhibitor pairs in Fig. 3. An additional undesirable behavior for these potentially orthogonal sets would be if DUET is biased toward selected proteins with missing interaction data (and, therefore, fewer edges in the graph). To determine whether this was the case, we ran permutation tests where we counted the number of missing interactions between the proteins in the initial DUET results and randomly sampled protein sets of the same size. We did not find that DUET results had a significantly higher number of missing interactions (Supplementary Fig 7c).

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. We used MP3-seq to understand binding specificity better in designed 2×2 pairs by testing variants of different lengths and with a different number of buried HB-nets. Our target pair, mALb8 (the common parent of the mALb set), is a high-affinity 2×2 heterodimer with three HB-nets between its A and B protomers (Fig. 5a). Two mALb protomer truncations were designed by removing one or two turns from the end of each α-helix (Fig. 5b and Supplementary Fig. 8a). The original binders and the truncations were screened with MP3-seq and LFCs were calculated from four replicates. The relationship between length and binding affinity was highly nonlinear, as truncating the binders by two turns eliminated binding (Fig. 5c). We could use these data to infer the shortest possible length (3.5 heptads) for 2×2 binders containing two buried HB-nets. This pattern was similar to that seen for the A_N and B_N single-helix-truncation experiments in Fig. 1d,e.

We wanted to find the minimum number of buried HB-nets needed for orthogonal binding. Therefore, we removed the first (R1), second (R2) or both (R12) buried HB-nets from the mALb8 dimer and replaced them with either large (L) or small (S) hydrophobic residues (Supplementary Methods). The original HB-nets are shown in Fig. 5d and the replacement sequences are shown in Supplementary Fig. 8b. We observed that two HB-net mismatches were needed to prevent binding (Fig. 5e, boxed areas). For example, $A_{R1(S)}$ and $B_{R2(s)}$ had two HB-net mismatches (half of the second HB-net on protomer A and half of the first HB-net on protomer B) and showed no binding. $A_{R12(S)}$ and $B_{R2(s)}$ had one HB-net mismatch (first HB-net on chain B) and still bound. $B_{R12(s)}$ weakly bound to both $A_{R1(S)}$ and $A_{R2(s)}$ with one HB-net mismatch. Additionally, hydrophobic residue size alone was not sufficient to confer orthogonality. For example, $B_{R1(S)}$ and $B_{R1(L)}$ bound to both $A_{R1(L)}$ and $A_{R1(S)}$. Using these simple rules and MP3-seq, we created two new orthogonal pairs ($A_{R1}:B_{R1}$ and $A_{R2}:B_{R2}$). We also note that the designs largely did not homodimerize (Fig. 5e, unboxed areas). The all-by-all screen of original, truncated and HB-net-removed mALb proteins can be found in Supplementary Fig. 8c, which showed that mALb proteins with one truncation had similar binding patterns with HB-net variants to the originals but that the second truncation eliminated binding.

Predicting orthogonal binding with linear models

To assess AF's ability to predict orthogonal interactions, we used AF2 and three published versions of AF-M (v1–v3) to predict complexes for all six 1×1 NICP pairs in Fig. 2b (Supplementary Methods). We compared computed error metrics (predicted local distance difference test, predicted aligned error, etc.) with MP3-seq LFC values and on-target and off-target classification. We noted that AF-M v2 and v3 performed better than AF2 and AF-M v1 (Supplementary Fig. 9a,b). The best-performing metric was the interface predicted TM (iPTM) score averaged across our predicted complexes (Fig. 6a). We wanted to see

how AF-M metrics compare to a state-of-the-art specialized model for 1×1 binding prediction, such as iCipa[14] (Supplementary Fig. 9c). We found that iCipa correlated better with LFC values (Fig. 6b), particularly in the classification task (Fig. 6c). iCipa also better predicted orthogonality (Supplementary Fig. 9d,e).

Next, we wanted to test the generalizability of AF2 on more complex protein architectures; thus, we predicted structures for the mALb8 (2×2) complexes. There was no apparent increase in correlation between AF-M v2 and v3 (Supplementary Fig. 9f,g). The AF-M v3 average iPTM for the mALb8 interactions is shown in Fig. 6d. Heterodimeric interactions had higher confidence than homodimeric complexes but the detailed effect of the HB-net mismatches was only partially captured.

Encouraged by the generalization abilities of AF-M, we set out to determine whether a combination of AF-M error metrics and structural metrics can be used to train a better predictor of LFC values. Rosetta was used to collect physics-based metrics (energy of interaction, surface of the interface, shape complementarity, etc.) for each simulated dimer complex. Agglomerative hierarchical clustering was used to reduce the number of multicollinear features between the collected energy terms and AF error metrics (Supplementary Fig. 10). Linear least square ridge regression models and logistic ridge regression classifiers were trained on feature sets of decreasing sizes (Methods). We used two train–test approaches. For the first, we ranked all data points by their P_{adj} values and partitioned the dataset into high-quality test set interactions and a mix of high-quality and low-quality training set interactions (see Supplementary Methods for modeling details). This approach was chosen to assess the ability of models using AF-M complex features to fill in missing interactions. In the other approach, a subset of proteins was selected and all interactions involving those proteins were assigned to the test set. The protein-based split was used to evaluate how AF-M complex-based models could predict interactions involving new proteins instead of filling in missing interactions between known proteins. Multiple train–test sets were created for this split by holding out different protein sets to get a distribution of test performances (Methods). Examples of the two test sets can be seen in Supplementary Fig. 11a.

The regression models performed better than using only iPTM on both protein architectures (Fig. 6b) and reached performances similar to iCipa on the 1×1 set. As expected, models using more Rosetta features yielded better results (Supplementary Figs. 11c–g and 12a,b). All regression models dropped drastically in performance in the held-out protein task. All models did well on the held-out interaction test set but the iPTM area under the precision recall curve (AUPRC) dropped drastically when considering the full dataset, while iCipa performed at the same level (Fig. 6c and Supplementary Fig. 11e). The classification task's performance stayed relatively high between the two tasks even when the training set size was reduced (Supplementary Fig. 13a,b).

Similarly, models trained on the more complicated mALb8 interactions had drops in performance between the held-out interaction and held-out protein tasks (Supplementary Figs. 11c–g and 14). Although there was a smaller gain in performance when more features were included in modeling compared to the NICP predictors, the model with the most features still outperformed using iPTM alone in LFC prediction (Fig. 6e). Interestingly, reducing the training set size led to gains in model performance, likely because of the held-out interaction test set initially consisting of only high LFC interactions (Supplementary Fig. 14).

Lastly, to evaluate whether simple models trained on AF-M features to predict MP3-seq LFC values could generalize to new protein families, we used models trained with only AF error metrics to try to predict across protein families. That is, we attempted to predict the mALb8 interactions with models trained on NCIP proteins and vice versa (Supplementary Figs. 12c–f and 14b). We found in both cases that the models performed worse than iPTM when applied to protein families on which they were not trained.
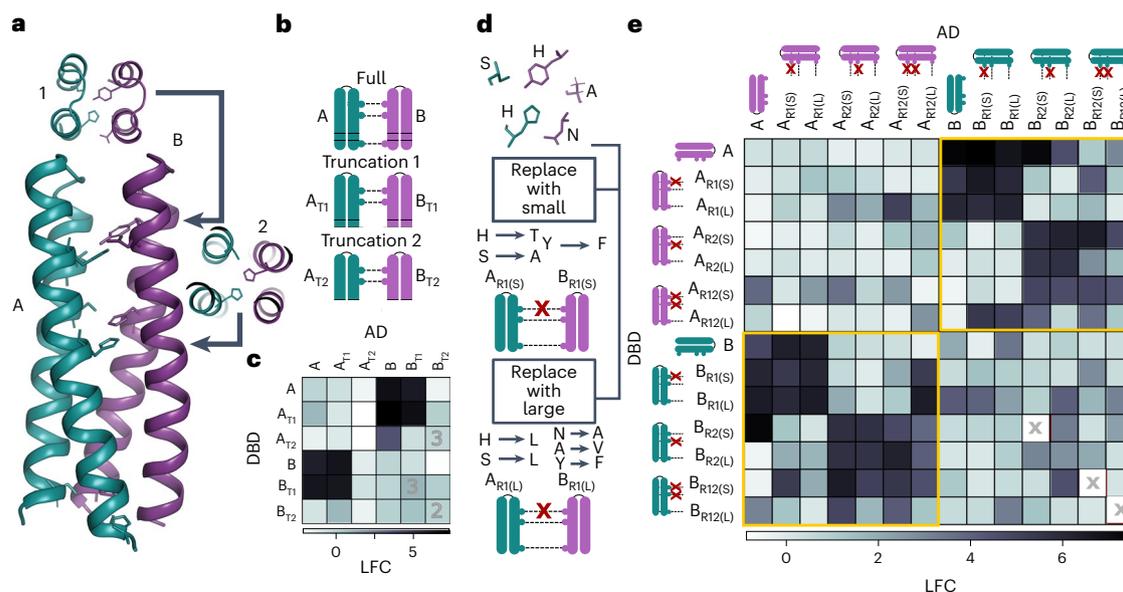
Fig. 5 | Systematic exploration of length and HB-net on helix–loop–helix binding. a, The mALb8 helix–loop–helix heterodimer with HB-nets 1 and 2 shown. **b**, mALb8 and the two shorter variants (T1, one helical turn shorter; T2, two helical turns shorter). **c**, LFC of the truncated mALb8 interactions. We observed a markedly nonlinear response to the length of the binder. T2 variants do not bind. All PPIs had four biological replicates, except those marked with their replicate number. **d**, An example of HB-net removal using the small and

large hydrophobic replacement protocols for the first HB-net. **e**, LFCs of mALb8 with and without removed HB-nets. S and L designate the replacement protocol used, while R_1, R_2 and R_{12} denote whether the first, second or both HB-nets were replaced. A gray '×' indicates insufficient reads. The yellow boxed regions show that two HB-net mismatches are enough to specify orthogonality. Hydrophobic residues alone cannot confer specificity.
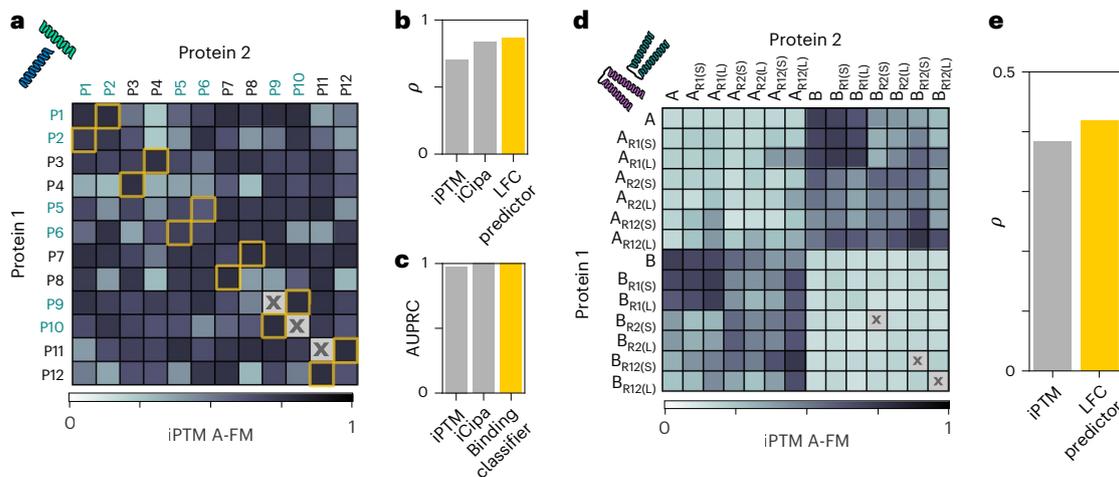
Fig. 6 | Predicted complex error for coiled dimers and simple model performance. a, Average iPTM per complex from AF-M for the NICP 1×1s. On-target interactions are marked with yellow squares. **b**, Test set performance of an LFC predictor with AF-M complex features compared to iPTM values only. **c**, AUPRC for classifying intended versus unintended test set interactions from a

binding classifier using AF-M complex features compared to only iPTM. The 1×1 binding predictor iCipa[14] is shown for comparison. Test set, $n = 14$. **d**, Average iPTM per complex with AF-M for a subset of the mALb8 interactions. **e**, Test set LFC predictor performance with AF-M complex features compared to iPTM only. Test set, $n = 19$.

Discussion

Here, we introduce MP3-seq, an easy-to-use HT-Y2H method that can measure pairwise PPIs in a single yeast strain without surface display. We also developed a data analysis workflow based on DESeq2 that makes it easy to merge replicates, remove autoactivators and identify statistically significant interactions. The MP3-seq workflow could be further generalized in future work by adjusting the selection scheme to (experimentally) eliminate autoactivators[45] or by integrating it with protein stability and expression assays such as Stable-seq[52] or high-throughput protease assays[53] to confirm that the tested proteins

are folded correctly and that an apparent negative interaction measurement is not because of reduced protein levels.

We validated MP3-seq using several sets of proteins for which interactions were previously characterized: a family of human Bcl-2 proteins and their de novo designed inhibitors, a set of peptides binding to Mcl-1 and three sets of coiled-coil peptides. We found quantitative agreement between our results and those reported previously. We then applied our method to characterize interactions in a pool of de novo 2×2 heterodimers containing buried HB-net and showed that it could scale to measure over 100,000 interactions in a single experiment.

Our computational workflow enabled us to identify potential orthogonal subsets at various orthogonality gaps from these data.

By screening interactions between protomers with truncations and modified HB-nets, we probed design rules for 2×2 binders. We showed that the minimum length for strong binding is 3.5 heptads, with binding affinity dropping sharply for shorter designs. Moreover, we found that at least two HB-net mismatches are needed for orthogonality, thereby setting minimum requirements for future sets of orthogonal 2×2 binders with buried HB-nets.

Lastly, we assessed the ability of AF2 and AF-M to predict orthogonal binding interactions. On the set of 1×1 dimers, we found that a sequence-based predictor (iCipa) correlated better with experimental LFC values than any single AF metric. However, using a combination of AF metrics and Rosetta physics-based metrics, accurate regression models (LFC predictor) could be trained on minimal data, capable of matching the best available family-specific models such as iCipa. While iCipa was trained on over 8,000 pairs and uses energy terms refined by hand, the LFC predictor can be generated automatically for virtually any protein family with one hundredfold fewer experimental data points.

While the LFC predictors excelled at test–train splits designed to mimic filling in missing interactions in an all-by-all screen, they struggled when test sets were selected to mirror the task of predicting interactions of 'new' proteins not present in training sets. This examination of AF-M error metrics and models trained on AF-M metrics and Rosetta simulation values showcases that, while complex predictors are a powerful tool, high-throughput experimental assays remain necessary for tasks such as orthogonality confirmation and determining mutation effects on interactions.

Looking forward, we believe the increased scale and streamlined workflow of MP3-seq will further accelerate the adoption of HT-Y2H methods. These benefits will facilitate use in applications ranging from training predictive models of PPIs for generative protein design to characterizing interactions between human protein variants at high throughput.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-024-01718-x.

References

1. Cho, J. H., Collins, J. J. & Wong, W. W. Universal chimeric antigen receptors for multiplexed and logical control of T cell responses. *Cell* **173**, 1426–1438 (2018).
2. Chen, Z. et al. De novo design of protein logic gates. *Science* **368**, 78–84 (2020).
3. Chen, Z. & Elowitz, M. B. Programmable protein circuit design. *Cell* **184**, 2284–2301 (2021).
4. Gao, X. J., Chong, L. S., Kim, M. S. & Elowitz, M. B. Programmable protein circuits in living cells. *Science* **361**, 1252–1258 (2018).
5. Groves, B., Khakhar, A., Nadel, C. M., Gardner, R. G. & Seelig, G. Rewiring MAP kinases in *Saccharomyces cerevisiae* to regulate novel targets through ubiquitination. *eLife* **5**, e15200 (2016).
6. Ljubetič, A. et al. Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat. Biotechnol.* **35**, 1094–1101 (2017).
7. Ben-Sasson, A. J. et al. Design of biologically active binary protein 2D materials. *Nature* **589**, 468–473 (2021).
8. Gonen, S., DiMaio, F., Gonen, T. & Baker, D. Design of ordered two-dimensional arrays mediated by noncovalent protein–protein interfaces. *Science* **348**, 1365–1368 (2015).
9. Kuriyan, J. & Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288 (1997).
10. Fong, J. H., Keating, A. E. & Singh, M. Predicting specificity in bZIP coiled-coil interactions. *Genome Biol.* **5**, R11 (2004).
11. Potapov, V., Kaplan, J. B. & Keating, A. E. Data-driven prediction and design of bZIP coiled-coil interactions. *PLoS Comput. Biol.* **11**, e1004046 (2015).
12. Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth. Biol.* **1**, 118–129 (2012).
13. Lebar, T., Lainšček, D., Merljak, E., Aupič, J. & Jerala, R. A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat. Chem. Biol.* **16**, 513–519 (2020).
14. Boldridge, W. C. et al. A multiplexed bacterial two-hybrid for rapid characterization of protein–protein interactions and iterative protein design. *Nat. Commun.* **14**, 4636 (2023).
15. Boyken, S. E. et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
16. Chen, Z. et al. Programmable design of orthogonal protein heterodimers. *Nature* **565**, 106–111 (2019).
17. Brodnik, A., Jovičić, V., Palangetić, M. & Silađi, D. Construction of orthogonal CC-sets. *Informatica* **43**, 19–22 (2019).
18. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. S. Protein–protein interaction detection: methods and analysis. *Int. J. Proteom.* https://doi.org/10.1155/2014/147648 (2014).
19. Smits, A. H. & Vermeulen, M. Characterizing protein–protein interactions using mass spectrometry: challenges and opportunities. *Trends Biotechnol.* **34**, 825–834 (2016).
20. Rouet, R., Jackson, K. J. L., Langley, D. B. & Christ, D. Next-generation sequencing of antibody display repertoires. *Front. Immunol.* **9**, 118 (2018).
21. Younger, D., Berger, S., Baker, D. & Klavins, E. High-throughput characterization of protein–protein interactions by reprogramming yeast mating. *Proc. Natl Acad. Sci. USA* **114**, 12166–12171 (2017).
22. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246 (1989).
23. Weile, J. et al. A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.* **13**, 957 (2017).
24. Luck, K. et al. A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
25. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
26. Erffelinck, M.-L. et al. A user-friendly platform for yeast two-hybrid library screening using next generation sequencing. *PLoS ONE* **13**, e0201270 (2018).
27. Jin, F., Avramova, L., Huang, J. & Hazbun, T. A yeast two-hybrid smart-pool-array system for protein-interaction mapping. *Nat. Methods* **4**, 405–407 (2007).
28. Rajagopala, S. V. & Uetz, P. Analysis of protein–protein interactions using array-based yeast two-hybrid screens. In *Yeast Functional Genomics and Proteomics* (ed. Stagljar, I.) Vol. 548 (Humana Press, 2009).
29. Trigg, S. A. et al. CrY2H-seq: a massively multiplexed assay for deep-coverage interactome mapping. *Nat. Methods* **14**, 819–825 (2017).
30. Weimann, M. et al. A Y2H-seq approach defines the human protein methyltransferase interactome. *Nat. Methods* **10**, 339–342 (2013).
31. Yachie, N. et al. Pooled-matrix protein interaction screens using barcode fusion genetics. *Mol. Syst. Biol.* **12**, 863 (2016).

32. Yang, F. et al. Development and application of a recombination-based library versus library high-throughput yeast two-hybrid (RLL-Y2H) screening system. *Nucleic Acids Res.* **46**, e17 (2018).

33. Yang, J.-S. et al. rec-YnH enables simultaneous many-by-many detection of direct protein–protein and protein–RNA interactions. *Nat. Commun.* **9**, 3747 (2018).

34. Yu, H. et al. Next-generation sequencing to generate interactome datasets. *Nat. Methods* **8**, 478–480 (2011).

35. Banerjee, S., Velásquez-Zapata, V., Fuerst, G., Elmore, J. M. & Wise, R. P. NGPINT: a next-generation protein–protein interaction software. *Brief. Bioinform.* **22**, bbaa351 (2021).

36. Velásquez-Zapata, V., Elmore, J. M., Banerjee, S., Dorman, K. S. & Wise, R. P. Next-generation yeast-two-hybrid analysis with Y2H-SCORES identifies novel interactors of the MLA immune receptor. *PLoS Comput. Biol.* **17**, e1008890 (2021).

37. Quartley, E. et al. Heterologous expression of *L. major* proteins in *S. cerevisiae*: a test of solubility, purity and gene recoding. *J. Struct. Funct. Genomics* **10**, 233–247 (2009).

38. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

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

40. Yu, D., Chojnowski, G., Rosenthal, M. & Kosinski, J. AlphaPulldown—a Python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**, btac749 (2023).

41. Johansson-Åkhe, I. & Wallner, B. Improving peptide–protein docking with AlphaFold-Multimer using forced sampling. *Front. Bioinform.* **2**, 959160 (2022).

42. McIsaac, R. S. et al. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41**, e57 (2013).

43. Curran, K. A. et al. Short synthetic terminators for improved heterologous gene expression in yeast. *ACS Synth. Biol.* **4**, 824–832 (2015).

44. Scanlon, T. C., Gray, E. C. & Griswold, K. E. Quantifying and resolving multiple vector transformants in *S. cerevisiae* plasmid libraries. *BMC Biotechnol.* **9**, 95 (2009).

45. Shivhare, D., Musialak-Lange, M., Julca, I., Gluza, P. & Mutwil, M. Removing auto-activators from yeast-two-hybrid assays by conditional negative selection. *Sci. Rep.* **11**, 5477 (2021).

46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

47. Plaper, T. et al. Coiled-coil heterodimers with increased stability for cellular regulation and sensing SARS-CoV-2 spike protein-mediated cell fusion. *Sci. Rep.* **11**, 9136 (2021).

48. Thomas, F., Boyle, A. L., Burton, A. J. & Woolfson, D. N. 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 (2013).

49. Berger, S. et al. Computationally designed high specificity inhibitors delineate the roles of Bcl2 family proteins in cancer. *eLife* **5**, e20352 (2016).

50. Rogers, J. M., Wong, C. T. & Clarke, J. Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J. Am. Chem. Soc.* **136**, 5197–5200 (2014).

51. Ljubetič, A., Gradišar, H. & Jerala, R. Advances in design of protein folds and assemblies. *Curr. Opin. Chem. Biol.* **40**, 65–71 (2017).

52. Kim, I., Miller, C. R., Young, D. L. & Fields, S. High-throughput analysis of in vivo protein stability. *Mol. Cell. Proteom.* **12**, 3370–3378 (2013).

53. Rocklin, G. J. et al. Global analysis of protein folding using massively parallel design, synthesis and testing. *Science* **357**, 168–175 (2017).

Methods

Experimental workflow

Gene fragments were ordered from Twist Biosciences for the DBD and AD fusions of protein binders of interest. Each fragment consisted of a linker coding sequence (CDS), the binder CDS, a stop codon, a synthetic terminator, a PCR handle, a barcode and an insulation sequence. The linker and insulation sequences serve as homology sequences during plasmid assembly in yeast. The plasmid vector contained linker homologies, the AD, the DBD, a CEN autonomously replicating sequence for yeast replication and TRP1 for tryptophan selection. A plasmid can be seen in Supplementary Fig. 1. These fragments and vectors are then transformed into electrocompetent Y777 yeast cells for combinatorial assembly through homologous recombination. Transformed cells were resuspended in synthetic complete medium lacking tryptophan, SC-TRP medium, and grown to allow redundant plasmid dropping. Postpassage cells were inoculated in SC-HIS-TRP medium for His selection or reserved and frozen for the His[+] sample. After selection in His[−] medium, cells were frozen for the His[−] sample. Samples were thawed, plasmids were extracted and barcode regions were prepped for sequencing with qPCR. Detailed information about the experimental workflow can be found in the Supplementary Methods.

Analysis workflow

FASTQ files were obtained using bcl2fastq, cutadapt[54] was used to trim barcode regions and Starcode[55] was used to cluster and count barcode–barcode pairs for the His[+] and His[−] files. At this point, experiments were screened for autoactivators and Autotune was used to infill His[−] barcode counts as needed (Supplementary Equations 1 and 2). Proteins classified as autoactivators can be found in Supplementary Table 2. LFC and P-LFC values were then calculated with DESeq2 (ref. [46]). Detailed information about the analysis workflow can be found in the Supplementary Methods.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of this work are available in this article and the Supplementary Information. Barcode count data (both raw barcode counts and processed barcode counts) and de novo designed sequences are available on GitHub (https://github.com/Seeliglab/MP3-DUET-AUTOTUNE). Raw sequencing data are available from the Gene Expression Omnibus under accession code GSE271790. Source data are provided with this paper.

Code availability

All scripts necessary to recreate the MP3-seq pipelines, analysis and models can be found on GitHub (https://github.com/Seeliglab/MP3-DUET-AUTOTUNE).

References

54. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).
55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).

Author contributions

A.B. optimized the MP3-seq workflow, performed most experiments and analyzed the data. A.L.F. designed the analytical pipeline and handled writing the paper, figure creation and modeling. B.G. conceptualized the MP3-seq approach and performed proof-of-concept experiments. C.M. performed experiments, created figures and wrote the paper. D.B. proposed conceptual ideas and provided project advice. A.L. designed the mALb8 and other de novo proteins, performed simulations, wrote the paper and provided funding. G.S. supervised experiments, provided funding and wrote the paper. All authors read and contributed to the final paper.

Competing interests

G.S. is a cofounder and shareholder of Parse Biosciences. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-024-01718-x.

Correspondence and requests for materials should be addressed to Ajasja Ljubetič or Georg Seelig.

Peer review information *Nature Chemical Biology* thanks Gabriel Rocklin and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s): Ajasja Ljubetič, Georg Seelig

Last updated by author(s): Jul 9, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	pyrosetta 2022.04+release.0c9f888, alphafold-colab 2.3.4, colabfold 1.5.2, cutadapt 4.7, starcode-v1.4, pymol 2.5.0
Data analysis	The following languages and packages were used to clean and analyze data and for modeling. Python 3.11.3: pandas 1.5.3, numpy 1.24.3, seaborn 0.12.2, matplotlib 3.7.1, scipy 1.10.1, scikit-learn 1.2.2, networkx 3.2.1, graph-tool 2.59. R 4.4.4: DESeq2 1.38.3. Code for the paper is available at https://github.com/Seeliglab/MP3-DUET-AUTOTUNE.git, which has been added to the main text availability statement.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Processed barcode counts, differential expression data, and designed protei sequences are available at https://github.com/Seeliglab/MP3-DUET-AUTOTUNE.git, and raw data files are available on GEO at GSE271790.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender (identity/presentation), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Not applicable to this study.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable to this study.
Population characteristics	Not applicable to this study.
Recruitment	Not applicable to this study.
Ethics oversight	Not applicable to this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](#)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed,. The size of the screened assays depended on the number of proteins present in the benchmarking set for which previous measurements of interaction existed, or on the number of designs generated and ordered for screening which was constrained by cost.
Data exclusions	Interactions where a protein was determined to be an autoactivator were excluded. It is worth noting that autoactivation is a known issue with Y2H assays of this type, and we were interested in developing a way to codify them and attempt to correct them as part of this work. We defined autoactivators using the distribution of enrichments in each assay, and replaced auto-activating interaction readcount values from alternate interactions in the assay (see methods). Additionally, some of the measured proteins in large assays were excluded from analysis and discussion due to the protein designers' expressed wishes, lack of interest, and/or lack of new discussion points for the paper. These proteins are still available in the raw data and code repository. We note that the NICP coils in Fig 2 were included in later experiments as controls, but had decreased correlation with earlier NICP experiments due to what we believe was decreasing stock quality.
Replication	Experiments were biologically replicated at least two times, with key benchmarking sets being biologically replicated up to five times. Technical replicates were conducted and analyzed for the NICP protein set. All attempts at replication, barring those with errors in the experimental pipeline and experiments with nonfunctional assembly strategies, were successful.
Randomization	Randomization was not applicable for this study, as high throughput sequencing studies such as this do not typically use randomization. Some reasons for this are that the targeted nature of such assays require specific mappings of barcodes to sequences of interest which make randomization difficult to implement, and that sequencing platforms already introduce a error rate and randomization could further increase errors in barcode mapping.
Blinding	Blinding was not relevant to this study, as it is not usually performed with high throughput sequencing experiments due to their complex natures. Additionally, our output was objective numerical data of barcode counts so blinding would not significantly affect our data interpretation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Plants

Seed stocks	Not applicable to this study.
Novel plant genotypes	Not applicable to this study.
Authentication	Not applicable to this study.