A phage-assisted continuous selection approach for deep mutational scanning of protein-protein interactions

Julia Zinkus-Boltz[1], Craig DeValk[2], Bryan C. Dickinson[1]*

[1]Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA

[2]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA

*Dickinson@uchicago.edu

Abstract

Protein-protein interactions (PPIs) are critical for organizing molecules in a cell and mediating signaling pathways. Dysregulation of PPIs are often key drivers of disease. To better understand the biophysical basis of such disease processes – and to potentially target them - it is critical to understand the molecular determinants of PPIs. Deep mutational scanning (DMS) facilitates the acquisition of large amounts of biochemical data by coupling selection with high throughput sequencing (HTS). The challenging and labor-intensive design and optimization of a relevant selection platform for DMS, however, limits the use of powerful directed evolution and selection approaches. To address this limitation, we designed a versatile new phage assisted continuous selection (PACS) system using our proximity-dependent split RNA polymerase (RNAP) biosensors with the aim of greatly simplifying and streamlining the design of a new selection platform for PPIs. After characterization and validation using the model KRAS/RAF PPI, we generated a library of RAF variants and subjected them to PACS and DMS. Our HTS data revealed that amino acid (aa) positions 66, 84, and 89 on RAF, key residues in the KRAS/RAF PPI, are intolerant to mutations. We also identified a subset of residues with broad aa substitution tolerance, aa positions 52, 55, 76, and 79. Due to the plug and play nature of RNAP biosensors, this method can easily be extended to other PPIs. More broadly, this, and other methods under development, supports the application of evolutionary and high-throughput approaches to bear on biochemical problems, moving towards a more comprehensive understanding of sequence-function relationships in proteins.

Protein-protein interactions (PPIs) are critical to cellular signaling, metabolism, and cellular organization.[1] Dysregulated PPIs are often key drivers of disease[2–5] and are therefore compelling targets for therapeutic development.[6–8] PPI dysregulation can emerge from either mis-regulated protein synthesis or post-translational modifications, or mutations in one of the binding partners, resulting in aberrant, disease-contributing interactions.[9] In either case, understanding the molecular determinants of PPIs is crucial to establishing the biophysical basis of disease and assisting in the creation of therapeutic PPI modulators.

Traditionally, a reductionist approach is deployed to understand PPIs, which generally involves alanine scanning to identify key sites for affinity and specificity of the interface between binding partners.[10,11] While these methods have generated a wealth of information, such as the identification of PPI hot spots,[12] they are labor intensive, requiring functional biochemical characterization of each variant, which often limits the number of variants characterized. The advent of display technologies, such as phage[13] and yeast display,[14] coupled with advances in two-hybrid approaches for the detection of PPIs,[15] has allowed the gathering of large amounts of information about the genotype-phenotype relationships of biomolecules[16] and PPIs.[17] These and other evolutionary and selection-based approaches, and their resultant large data sets, paint a more nuanced picture of how individual mutations correlate with altered function.

The application of high-throughput sequencing methods to understand sequence-function relationships in biomolecules is referred to as deep mutational scanning (DMS).[18,19] DMS generates libraries of protein variants whose individual functional consequences are assayed through selection experiments. Based upon the relative enrichment of each variant in the selection, as monitored by high throughput sequencing (HTS), the functional consequence of each mutation can be deduced. A key component of every DMS experiment is the design and characterization of a selection platform for the desired protein and its relevant properties.[20] Display-based platforms,[21–23] and other assays that directly link protein function to an observable phenotype, such as cell viability[24] or growth,[25] have been used to screen protein libraries.

Continuous evolution methods,[26] which link defined biochemical fitness to viral life cycles, allow rapid and large-scale evolution and selection. However, the *in vivo* nature of viral selection systems requires robust genetic selection systems for a biochemical activity of interest. Phage-assisted continuous evolution (PACE)[27] and phage-assisted non-continuous evolution (PANCE),[28,29] are especially powerful technologies in which large populations of gene-encoding protein variants can be evolved or selected for over time, based on defined biochemical properties. For example, PACE has been used to evolve T7 RNAP promoter specificity,[27,30] protease specificity,[31] drug resistance,[32] Cas proteins,[33] and several other biochemical machines.[26,34] PACE has also been used to evolve PPIs using a bacterial 2-hybrid-based selection system.[35] While this evolutionary system is powerful, it is limited in terms of requiring optimization for each new PPI target, necessitating re-design for each new target of interest.

The design and optimization of a relevant selection platform is often the most challenging and labor-intensive step in DMS experiments and, we reasoned, is one of the key reasons why powerful selection approaches such as PACE and PANCE have not found more widespread use. To address this limitation, we designed a versatile new phage-assisted continuous selection (PACS) system to interrogate the binding interface between interacting proteins. We developed PACS in the context of probing the KRAS/RAF PPI, due to its high value as a target for cancer therapies; both KRAS and RAF are known oncogenes, with roughly 20% of all cancers harboring mutations in KRAS.[36,37] Practically, we chose to use the KRAS4b isoform due to its high mutation

rate across different cancers,[38] and the RAS binding domain (RBD) of RAF, which is the only part of the protein that binds RAS isoforms.[39,40]

Our new system relies on proximity-dependent split RNA polymerase (RNAP) biosensors,[41] greatly simplifying and streamlining the genetic sensing of PPIs *in vivo* compared with more traditional 2-hybrid approaches. Using these split RNAP biosensors, we designed a PACS system that links replicating bacteriophages to RAF, contained within the page genome, binding KRAS. After validating the system, we generated libraries of phage-encoded RAF variants and subjected them to PACS. Analysis by HTS revealed whether each site in RAF is tolerant to mutation, and if so, which mutations are tolerated. Aside from providing new insights into the KRAS/RAF interface from the RAF perspective, our new method is broadly applicable, facilitating rapid understanding of the mutational landscape of PPIs using DMS.

Results

Engineering split RNAPs to detect the KRAS-RAF PPI

In PACE, the viral life cycle of M13 bacteriophage is linked to a target activity through the production of gIII, a required phage gene that is moved from the phage genome into host *Escherichia coli* cells.[27] The phage must therefore infect the host cell and produce a phage-encoded protein variant capable of activating gIII production in the host cell in order to replicate. Therefore, the key to developing a system for a desired target activity is linking the target activity of interest, in our case RAF and KRAS binding, to the production of gIII. To do this, we deployed our previously-reported split RNAP based biosensor platform, based on an evolved variant of split T7 RNAP, that only forms a functional enzyme and produces a defined RNA output when proteins fused to each half of the split enzyme bind one another.[41] Our split RNAP biosensor technology is quite versatile and able to monitor different PPIs without requiring re-optimization, as evidenced by its successful deployment to detect various PPIs, including PPIs from the BCL-2 family of proteins[42] and small molecules.[43]

First, we assayed whether our split RNAP biosensor can detect the PPI between KRAS and RAF in *E. coli* using a modified version of our previously reported *in vivo* luciferase assay, in which interactions between KRAS and RAF should drive luciferase production.[41] We designed the *in vivo* luciferase assay with three plasmids: 1) an N-terminal expression plasmid containing the N-terminal fragment of the split RNAP ($RNAP_N$) with a C-terminal fusion of either RAF or no interaction partner, 2) a C-terminal expression plasmid that constitutively expresses the C-terminal RNAP fragment ($RNAP_C$) with a N-terminal fusion of either KRAS or an off-target protein interaction partner, and 3) a reporter vector with the bacterial luciferase gene (luxAB)[44] under control of the T7 RNAP promoter (**Figure 1a**). Under these conditions, only interaction of the two fusion proteins should result in assembly of the split T7 RNAP biosensor, expression of luciferase, and generation of luminescence (**Figure 1b**).

After cloning and optimizing the system components, we then measured luciferase output in reporter *E. coli* cells expressing either on-target (KRAS/RAF) or control (ZB/RAF; ZB is a leucine zipper peptide that does not bind RAF)[41,45] PPI fusions. As expected, the KRAS/RAF combination yielded robust luciferase signal (**Figure 1c**). Absence of either KRAS or RAF on either half of the biosensor resulted in ~20-fold less signal, confirming KRAS/RAF PPI-dependent gene expression. To further validate the system, we introduced a known, interaction-disrupting mutation

in RAF (R89L).[46] Again, as expected, $RNAP_N$ fused to RAF R89L produced levels of luciferase activity comparable with that of off-target controls (**Figure 1c**). After validating the use of our split biosensor and our *in vivo* luciferase assay for the KRAS/RAF PPI, we next sought to deploy the biosensors in a new PACS system.

Designing a PACS system for PPIs

To adopt the biosensors into a PACS selection system, we cloned *E. coli* selection plasmids and modified phage to encode the appropriate system components (**Figure 2a**). We modified a previously-optimized M13 bacteriophage genome[41] by removing gIII and replacing it with our split $RNAP_N$ biosensor C-terminally fused to RAF or an off-target protein as a control. The selection plasmids contained in the *E. coli* selection cells were used to modulate the expression of gIII through the assembly of our split RNAP biosensors, including: 1) a "positive selection plasmid" that contains the CGG $RNAP_C$ half of our RNAP biosensor, which is a known variant of the T7 RNAP that recognizes an orthogonal CGG RNAP promoter,[47] N-terminally fused to KRAS, and gIII under the control of the CGG RNAP promoter; and 2) a "negative selection plasmid" that contains the T7 $RNAP_C$ half of our RNAP biosensor N terminally-fused to an off target leucine zipper peptide, ZBneg,[41,45] and contains a dominant negative form of gIII, gIIIneg,[48] controlled by the cognate T7 RNAP promoter. With these selection plasmids, only a KRAS-binding $RNAP_N$-fusion should drive reassembly of the split CGG RNAP biosensor and gIII expression, generating infectious phage progeny. Non-PPI-driven assemblies would activate both the positive and negative selection plasmids, resulting in the production of gIIIneg and non-infectious phage. In other words, the negative selection plasmid keeps the phage from "cheating" the selection by maintaining proximity-dependency in $RNAP_N$.

After constructing the selection cells and phage, we validated the ability of the system to discriminate between phage encoding the known binding partner of KRAS, RAF, and an off-target protein, an isoleucine zipper peptide termed iZA,[41,45] using activity-dependent plaque assays. Because this assay informs upon the ability of a given phage to replicate and infect neighboring cells by forming "plaques," we were able to directly monitor if the proteins fused to the $RNAP_N$ biosensor contained in the phage activated the positive selection plasmid in the selection cells, producing gIII. Indeed, only the phage containing $RNAP_N$ fused to RAF were able to form distinct plaques on the selection cells, validating the system components (**Figure 2b**).

Next, we sought to validate the selection system in a mock PACS experiment using spike-in samples to see if we can enrich an active phage variant from an excess of inactive variants (**Figure 2c**). To do this, we seeded a phage vessel with 1,000-fold excess iZA phage to RAF phage. Fresh *E. coli* selection cells containing the selection plasmids were continuously flowed in from a chemostat, allowing constant diluting inflow and waste outflow. We monitored the selection over the course of 24 h, sampling the phage at various time points. We assayed the relative distribution of the inactive iZA phage versus the active RAF phage by PCR amplification of the phage samples, using primers that selectively amplify the phage-encoded gene product. Indeed, although there was no detectable RAF in the initial phage population used to seed the experiment, the population of RAF-encoding phage increased over time, while the inactive iZA phage were diluted out, with complete washout and enrichment after 6 h (**Figure 2d**).

Validating PACS system for PPIs

After validating the system components in a mock selection, we next sought to confirm the system was capable of evolving PPIs from non-interacting variants. In addition to the aforementioned R89L variant, we cloned a series of RAF mutants that we predicted would disrupt the KRAS/RAF PPI, informed by either literature or intuition from the crystal structure.[49] These variants included: K84A, K84E, Q66A, and Q66K. We first assayed the mutants in the luciferase reporter assay, which revealed that alanine mutations at positions 66 and 84 had lower activity compared to wild-type (wt) RAF, but higher activity than the R89L null mutation (**Figure 3a**). The K84E and Q66K mutations, with more dramatic charge changes to the protein surface, had activity similar to R89L, suggesting abrogation of the PPI. We cloned each mutant into RAF phage for activity-dependent plaque assays. As expected, based on the reporter assay, K84A and Q66A showed slightly smaller, "weaker" plaques than wt RAF, indicating weaker affinity for KRAS. In contrast, R89L phage produced no plaques, indicating no detectable PPI with KRAS, consistent with its known loss of interaction affinity (**Figure 3b**).[49] The K84E and Q66K phage generated extremely small, "diminished" plaques compared to the alanine mutants at those positions, indicating further reduction in affinity to KRAS. Although only qualitative, these plaque assay experiments suggest both the RNAP biosensors and phage selection are capable of detecting RAF variants with weak but measurable binding affinity and are capable of discriminating between the relative binding affinities.

We next aimed to evolve a non-interacting RAF variant to regain its KRAS-binding capabilities. To do this, we coupled the PACS selection cells (**Figure 2a**) with the previously reported *in vivo* mutagenesis plasmid MP6,[50] to develop a new PPI PACE system. We chose to initiate the experiment with phage encoding the R89L, Q66K, or K84E RAF variants, which started with either undetectable or diminished activity, to test if we could rescue wt-like binding phenotypes. We seeded three separate phage vessels, each with one population of either R89L, Q66K, or K84E RAF phage, then ran PACE for 24 h under KRAS PPI selection. At the end of the experiment, all three lagoons still contained robust populations of phage, which we collected and Sanger-sequenced to assess the mutations in the bulk library population (**Figure 3c**). The phage populations that began with R89L and K84E both reverted back to wt through evolution. Unsurprisingly for the R89L-starting population, only one codon was observed due to only one codon, "CTT," requiring a single nucleotide base substitution to encode arginine, while the other five degenerate codons for arginine required two or three nucleotide base substitutions. Similarly, for the K84E-starting population, it was not surprising that only one codon was observed because of the two degenerate codons encoding lysine, only one, "GAA," required a single nucleotide base substitution to revert back to lysine. In contrast, more than one codon was observed in the bulk sequencing library for the Q66K-starting population. This was surprising, considering that only a single nucleotide base substitution would change the mutant codon "AAA" to one of the two degenerate codons encoding for glutamine. Based upon the bulk sequencing data, there are six possible codons, encoding for six different amino acids, that could be present in the final population: K (the starting mutant amino acid), Q (the wt amino acid), N, H, D, and E. Of these possible amino acids, asparagine was the only polar amino acid, containing the same amide functional group and differing only in the length of the carbon chain in the R group, as compared with the wt glutamine. We therefore reasoned that this would be one of the species observed upon further assessment of individual variants from the bulk population.

We then subcloned individual variants from each of the populations to test KRAS binding in the luciferase reporter assay. As expected, the initial R89L and K84E populations contained mostly wt phage, with some additional silent mutations that performed similarly to wt in the reporter assay (**Figure 3d**). The Q66K phage population contained multiple variants, including the Q66N mutation predicted from the bulk sequencing results. The Q66N variants had luciferase activity levels greater than the starting Q66K mutant, but slightly less than wt, indicating partial recovery of wt-like KRAS binding affinity. Taken together, these experiments confirm that the selection vectors can be used to either select for or evolve RAF variants that bind KRAS for either PACS or PACE, respectively. Next, we aimed to test whether the system could be used in a high-throughput DMS experiment using a phage-encoded RAF library.

Deep mutational scanning of the KRAS/RAF binding interface

To demonstrate how our PACS system could easily be integrated into the DMS workflow,[20] we generated a library of RAF variants introducing mutations along the interface of the KRAS/RAF interaction based upon crystal structures of these proteins, focusing on amino acid residues 52-90. We generated the library utilizing error prone PCR (epPCR) mutagenesis methods,[51] with the level of mutagenesis tuned such that double nucleotide mutations were predominantly observed in the final library (**Supplementary Methods 1**). To ensure that the generated library still contained active variants, a small subset of the library was cloned into the N-terminal luciferase assay plasmid (**Figure 1a**) and tested (**Supplemental Figure 1**). The assayed variants showed varying degrees of activity compared to the wt RAF, ranging from variants with activity similar to the R89L null variant to activity comparable to wt RAF.

Next, we subjected three replicate populations of the RAF library to our PACS platform (**Figure 4a**) to enrich for RAF variants with the ability to bind KRAS. Phage library samples were collected at multiple time points throughout the course of the 72 h selection experiment. We cloned the collected phage samples from the end of the selection at 72 h into the N-terminal expression plasmid (**Figure 1a**) and tested them in the luciferase assay to confirm that the selection experiment had enriched for RAF variants with the ability to bind KRAS (**Supplemental Figure 2**). The assayed RAF variants all had activity above that of the RAF R89L null variant, demonstrating that the selection had functioned as anticipated, i.e., active variants were enriched over the course of PACS.

We subjected the collected phage samples from all three replicate experiments to HTS analysis to perform DMS, with time-dependent information, of the mutated binding interface. Practically, only the region that was subjected to mutagenesis was generated into sequencing amplicons to allow for paired end reads, improving the overall quality of the data by reducing error associated with sequencing. In our data set, we were able to observe 1279 unique protein variants over the course of the experiment. Functional scores were calculated for each observed variant in the HTS dataset based upon the observed counts and were plotted to show the distribution of functional scores overtime for the different variants observed in the data set (**Figure 4b, Supplemental Figure 5**).

To test if the calculated functional scores from PACS correspond to the relative binding affinity, we compared literature-reported K_D values[49] for various RAF mutants observed in our HTS dataset against our calculated functional scores (**Figure 4c, Supplemental Figure 6**). There is a clear correlation between the functional scores and the relative binding affinities of the different

RAF mutants (R^2 = 0.96; p = 0.0035). We also verified that our *in vivo* luciferase assay could approximate binding affinity by plotting each variant's luminescence against the corresponding calculated functional score from the HTS datasets (**Supplemental Figure 7**). Taken together, these data confirm that PACS enriched active RAF variants capable of binding KRAS, resulting in datasets delineating relative fitness across a range of RAF mutants.

Functional consequence of amino acid substitutions along the KRAS/RAF binding interface

From our DMS dataset, we were able to examine the relative functional consequence of different amino acid substitutions along the region that we mutagenized (**Supplemental Figure 8**) by calculating the enrichment values for each amino acid observed along the binding interface relative to the wild type amino acid at the 72 h time point (**Figure 5a, Supplemental Figure 9)**. For the 39 amino acid positions explored in our mutant library, many positions favored positively charged side chains, with either the wt amino acid (aa) being a positively charged residue or positions enriching for positive residues (16 total positive aa: 9 positions with a wt positive aa; 10 positions enriching for a positive aa, 3 of these positions having a wt positive aa), and very few favored negatively charged side chains (3 total aa: 1 position with a wt negative aa, 2 positions enriching for a negative aa), which corresponds with the overall structure of RAF having a large number of basic residues (K and R) along the binding interface with KRAS.[52,53] Also, many of the positions along the binding interface favored the wt amino acid across experimental replicates: 59, 62, 63, 65, 66, 69, 77, 78, 80, 82, 84, 89, and 90. As expected, the amino acids at positions 89, 84, and 66 that are reported in the literature as being key in the binding interface of KRAS and RAF, showed the least tolerance to substitutions with only the wt sequence being favored and all other observed amino acids de-enriching at these positions.[46,49,52]

Other positions that tolerated substitutions with similar aa properties groups to wt were: 56, 58, 60, 61, 67, 68, 70, 72, 85, and 86. Residues 56 and 68 favored polar residues, similar to the wt amino acid asparagine or threonine, respectively; residue 68 was unique in that it only favored the polar residues with alcohol moieties on the R groups (wt S, or T). In the case of residues 56 and 68, they are directly involved in stabilizing the β-sheet structure. Residue 61 favored bulky aromatic side chains, favoring either phenylalanine or tyrosine. Positions 58, 60, 70, 72, and 86 all favored non-polar residues; only favoring a subset of the residues of similar sizes. For example, position 70 only seemed to favor nonpolar aliphatic residues, but not small nonpolar side chains like glycine or alanine, possibly due to the need to reduce flexibility in the β-sheet structure and increase the overall surface area of binding between KRAS and RAF. Position 85 was unique in that it enriched for both glycine and proline, but not other non-polar residues. It is interesting that these two particular amino acids were enriched because they have very different effects on structure; glycine is highly flexible while proline is the most rigid in terms of conformation. Residue 85 is located in a turn of an α-helix requiring the backbone to change direction by either introducing a proline to introduce such a change, or the flexible glycine to allow for such a change.

There were a few amino acid positions that tolerated amino acid substitutions with properties that differed from the wt amino acid: 53, 54, 57, 64, 73, 74, and 88. For example, position 53 enriched for the nonpolar amino acids isoleucine and methionine, but in the wt protein it is lysine. Other amino acid positions tolerated a diverse set of amino acid substitutions, in particular positions 52, 55, 71, 75, 76, 79, 81, 83, 87. Positions 52, 55, 76, and 79 all enriched for multiple types of amino

acids, including those from nonpolar, polar, and charged categories. Position 76 was unique in that it had the greatest number of amino acid substitutions among all the replicates (**supplemental figure 9**) that were tolerated without a clear categorical preference: alanine, valine, leucine, isoleucine, threonine, lysine, and arginine. It is interesting to note that this particular residue, as well as residue 79, is located on the loop region between an α-helix and a β-sheet, possibly accounting for the varied amino acids tolerated at these positions (**Figure 5b**).

Discussion

While not every amino acid at each position was sampled, potentially due to either bias present in error prone mutagenesis methods,[51,54] or from the limited sequencing coverage of each time point examined over the course of the experiment, we did observe trends associated with certain positions. Known key residues in the KRAS/RAF interaction showed intolerance to amino acid substitutions, while residues not directly involved in the interface, located in loop regions between an α-helix and a β-sheet, had more variability in the amino acids tolerated at those positions. Deeper sequencing or improved mutagenesis methods[55] would likely improve the depth of the datasets and number of unique variants assayed. Nevertheless, these experiments confirm the new PACS platform is able to rapidly select for RAF variants capable of biding KRAS, recapitulating known invariable positions identified in RAF. We were also able to show how the functional scores calculated correlated to binding affinity of different RAF variants based on reported K_D values[49] and observed luciferase activity *in vivo*, facilitating rapid characterization of variants obtained through the selection experiment.

Due to the plug and play nature of the RNAP based biosensor, other PPIs can be similarly examined in a high-throughput manner. The only requirements for a given target are that the proteins are expressible in *E. coli* and can tolerate fusion to the split RNAP biosensor. Notably, this latter limitation has recently been mitigated through the evolution of the C-terminal half of the RNAP biosensor to tolerate C-terminal fusions.[56] This method could also be utilized to facilitate the interrogation of very similar proteins, such as members of a protein family, to determine differences in highly similar protein variants and to identify key residues and properties that are unique to individual members. Another key benefit of this method is the ease with which it can be converted into a directed evolution campaign by the addition of *in vivo* mutagenesis to convert the PACS system into PACE. Because of this feature, not only can information about the binding interface be determined, but a *de novo* protein binder could also be evolved in response to given selection parameters. More broadly, this and other methods under development[57,58] will continue to apply evolutionary and high-throughput approaches to interrogate biomolecular interactions, moving away from traditional reductionist approaches toward a more comprehensive understanding of the sequence-function relationships in protein machines.

Acknowledgements

We thank Rama Ranganathan for providing the Illumina MiSeq instrument. This work was supported by the University of Chicago, the National Cancer Institute (R21CA21775) of the National Institutes of Health and a CAREER award from the National Science Foundation (1749364). J.Z.-B. was supported by a Chemical Biology Training Grant from the National Institutes of Health (T32 GM008720).

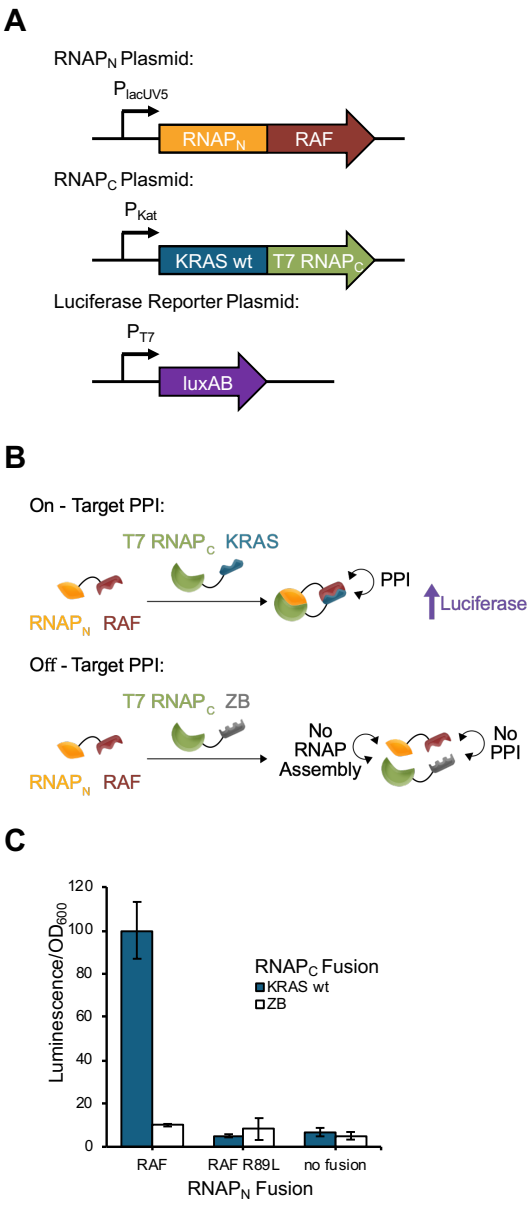

Figure 1. Detecting the KRAS-RAF PPI with split RNAP biosensors. (A) Plasmids for the *in vivo* luciferase assay to monitor the KRAS/RAF PPI. (B) Schematic of on-target and off-target PPI partners and there expected output in the *in vivo* luciferase assay system; only pendent protein fusions that interact with each other can drive the reassembly of the split RNAP biosensor to produce the luciferase enzyme and luminescence. (C) Luminescence signal for the on-target and off-target PPIs as monitored by the *in vivo* luciferase assay (error: std. dev., n= 4).

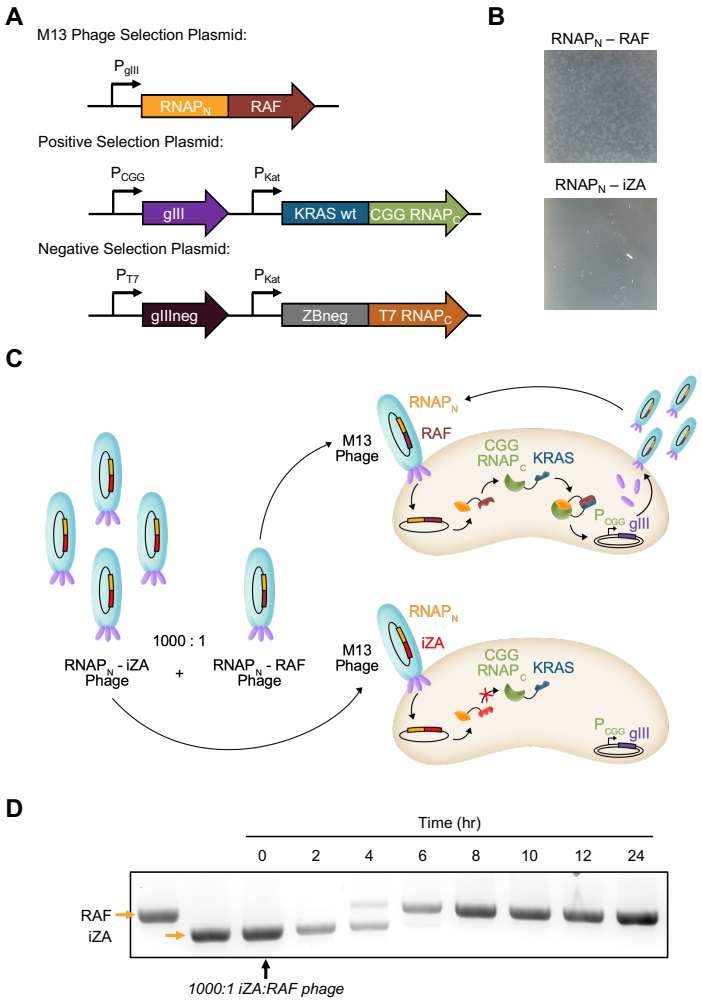

Figure 2. Design of a phage-assisted continuous selection (PACS) system for PPIs. (A) Plasmids for the *in vivo* PACS system to monitor the KRAS/RAF PPI. (B) Plaque assays of on-target and off-target PPI fusions on the PACS selection cells. (C) Schematic of the mock selection with and the on-target RAF phage and off-target iZA phage; only the RAF phage will be able to replicate on the selection cells and produce gIII to generate infectious phage progeny during the course of the selection. (D) DNA agarose gel depicting the relative fraction iZA and RAF phage over the course of the mock selection experiment.

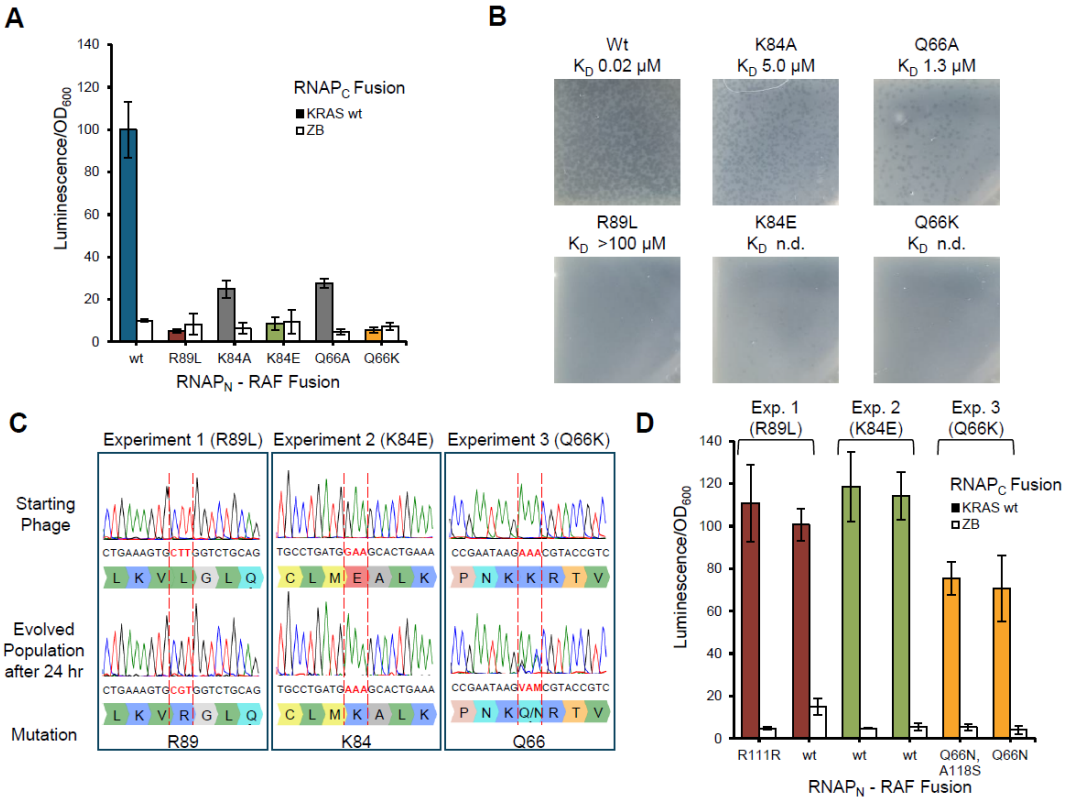

Figure 3. Validation of PACS PPI selection system by evolving inactive RAF variants to bind KRAS. (A) Luminescence signal for wt RAF and RAF variants with known or suspected lower binding affinity to wt KRAS on the wt KRAS or off-target PPIs (error: std. dev., n= 4). The RAF mutants used for additional testing were color coded: red: R89L; green: K84E; yellow: Q66K. (B) Plaque assays of the different RAF variants on the PACS selection cells. (C) Bulk Sanger sequencing of the phage populations before and after 24 h of PACE. (D) Luminescence signal for two evolved RAF variants from each experiment on the wt KRAS or off-target PPIs (error: std. dev., n= 4).

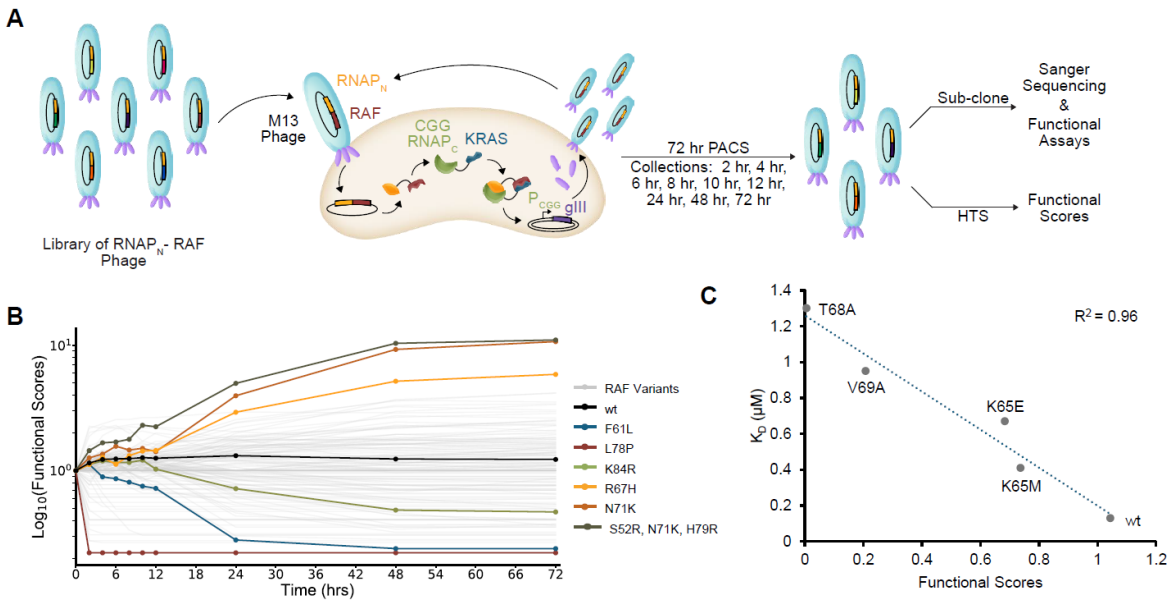

Figure 4. Deploying PACS system for DMS of RAF. (A) Schematic of the work-flow for the PACS DMS experiment for an error prone RAF mutant library. (B) Plot of time-dependent functional scores for all observed RAF variants in a single replicate lagoon from the PACS DMS experiment. (C) Plot of the correlation between literature reported K_D values and their corresponding calculated functional scores observed from a single replicate lagoon from the PACS DMS experiment (R^2 = 0.96, p value = 0.0035).

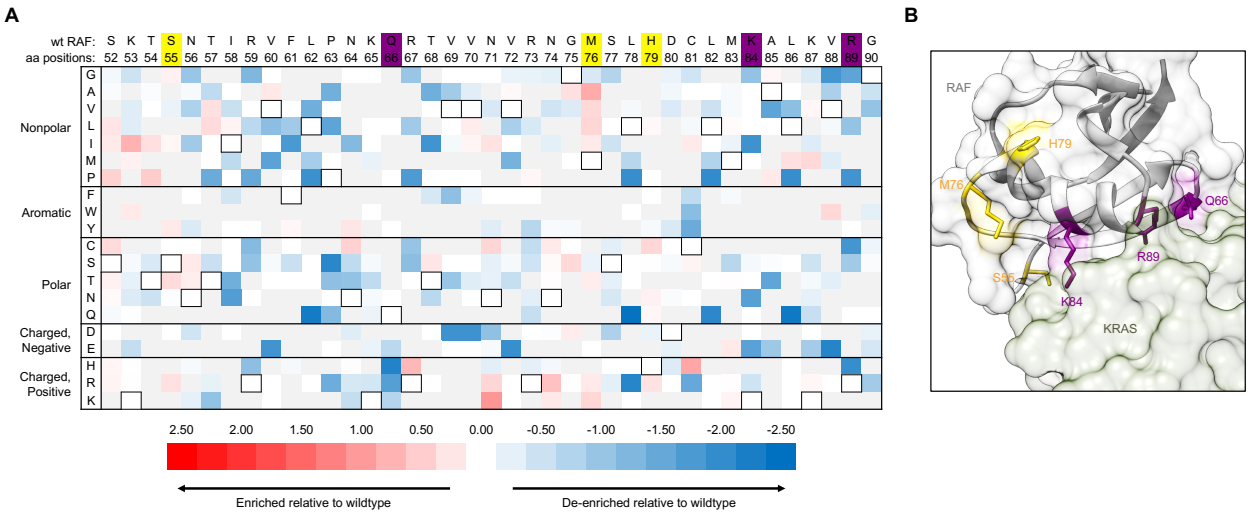

Figure 5. Analysis of DMS data for RAF. (A) Heat map of relative enrichment values at 72 h for each amino acid compared to the wild type amino acid. The wt aa is indicated with a think boarder, and any amino acid not observed at either 0 h or 72 h in the HYS dataset was colored gray. (B) Crystal structure of the KRAS/RAF PPI (PDB: 4G0N). Positions 66, 84, and 89, which are key residues in the KRAS/RAF PPI, showed intolerance to amino acid substitutions in the HTS dataset, and are indicated in purple. Positions 55, 76, and 79 showed a high degree of amino acid substitution tolerance in the HTS data, and are indicated in yellow.

Methods

Cloning

All plasmids used in this study were generated using Gibson assembly of PCR products generated with NEB Q5 high fidelity polymerase. All constructs were validated via Sanger sequencing performed by the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping facility. All constructs are listed in **supplemental table 1 - 5** with links to online vector maps. The wt KRAS4b gene was obtained from the Frederick National Laboratory for Cancer Research.

RAF Phage Library Generation

The insert was prepared by introducing mutations into the RAF gene using error prone PCR (epPCR).[51] Briefly, 50 µL epPCR reactions were set up with the following final concentrations: 1 ng/50 µL 14-70 template plasmid, 0.2 µM forward primer, 0.2 µM reverse primer, 1x standard NEB Taq buffer, 0.2 mM dNTPs, 750 µM $MnCl_2$, 0.025 U/µL NEB Taq polymerase. The thermocycler conditions utilized an initial denaturation step (95°C, 30 sec) followed by 30 cycles of denaturation (95°C, 30 sec), annealing (58°C, 30 sec) and extension (72°C, 1 min), and a final extension (72°C, 2 min). The PCR product was digested with 1 µL NEB DpnI enzyme in the PCR reaction buffer for 1 h at 37°C, and then purified using a Zymo clean and concentrator kit.

The $RNAP_N$-fused wt RAF phage vector (14-70) backbone was amplified overnight in 1059 cells[59], which replicate phage lacking gIII, and purified using a Qiagene plasmid purification kit.

Inserts were subcloned into the 14-70 vector using the XhoI and PstI sites with NEB XhoI enzyme and NEB PstI-HF enzyme following the standard NEB double digest protocol. The T4-mediated ligation reactions were performed overnight at 16°C using NEB T4 DNA ligase following the standard NEB ligation protocol, and purified with a Zymo DNA clean and concentrator kit. All resulting plasmids were suspended in 30ul water.

The phage library was prepared by electroporating (1800 V, 5 ms, 0.2 cm cuvettes, 50 µL cells, 143 ng DNA, 7 transformations total) 1 µg ligated vector product into 350 µL electrocompetent 1059 cells,[59] which were recovered in 25 mL of SOC media for 1 h. Afterwards, the recovered cells were used to inoculate 250 mL super broth to allow for phage library growth overnight. The phage containing supernatant was collected by centrifugation and sterile filtration to generate a RAF phage library.

Plaque assays

To quantify phage titers, 1059 activity-independent cells[59] were used; while S1030 cells[59] transformed with the positive selection plasmid (31-69) and the negative selection plasmid (20-6) were used to quantify the activity of different phage containing $RNAP_N$ protein fusions on KRAS. For all experiments, E. coli were grown to an optical density (OD_{600}) of approximately 0.6.

Phage dilutions (1/50, 1/12500, 1/125000, 1/6250000) were made by serially diluting a phage sample in water. Plaque assays were performed by mixing 50 µL of each phage dilution with 50 µL of cells, suspending these solutions in 55°C warmed soft agar (7 g agar/1 L LB), and plating

this mixture onto quartered solid agar plates without antibiotics (15 g agar/1 L LB). These were then grown in a 37°C incubator overnight, and phage plaques were counted and imaged to determine the relative plaque forming units (PFU).

Luciferase assays

Three separate vectors were constructed: 1) a N-terminal expression plasmid containing the $RNAP_N$ fragment C-terminally fused to either RAF or no protein partner upstream of the UV5 IPTG inducible promoter,[60] 2) a C-terminal expression plasmid containing the T7 $RNAP_C$ fragment N terminally fused to either KRAS or an off target protein partner upstream of the constitutive P_{Kat} promoter, and 3) a luciferase reporter plasmid containing the bacterial luxAB gene[44] upstream of the T7 RNAP promoter. The N-terminal, C terminal, and the luciferase reporter plasmids were electroporated into S1030 cells,[59] and plated onto agar (15 g/L in LB) plates containing 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, 33 μg/mL chloramphenicol, and 7.5 μg/mL tetracycline. Individual colonies were then picked and grown overnight to saturation with shaking at 37°C in 500 μL of 2xYT media containing 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, 33 μg/mL chloramphenicol, 7.5 μg/mL tetracycline in a 96, deep-well plate. A new culture was then prepared using 60 μL of the overnight culture to inoculate 540 μL of fresh 2xYT media containing 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, 33 μg/mL chloramphenicol, 7.5 μg/mL tetracycline and 5 μM IPTG. After 3 h of growth at 37°C with shaking, a 150 μL aliquot was transferred to a black-walled, clear-bottomed, 96-well plate and read on a Biotek Synergy Neo2 plate reader, monitoring both the optical density at 600 nm (OD_{600}) and the luminescence signal.

Luciferase activity was reported as normalized luminescence signal/OD_{600} for each well. Values are reported as the arithmetic average of replicate wells (n = 4) with error reported as the standard deviation of the replicates. In order to compare different days, the normalized luminescence signal/OD_{600} for a given sample was divided by the normalized luminescence signal for the wt KRAS/RAF interaction partners, which was given an arbitrary value of 100.

PACS & PACE

PACS was performed utilizing a modified version of the previously described PACE method.[41] S1030 cells were transformed by electroporation with the positive selection plasmid (31-69) and the negative selection plasmid (20-6). A 5 mL starter culture was grown overnight to saturation in LB media supplemented with 50 μg/mL carbenicillin, 40 μg/mL kanamycin, and 7.5 μg/mL tetracycline. Chemostats (250 mL sterile bottles) containing 150 mL of Davis rich medium[41] were inoculated with 5 mL starter culture and grown at 37 °C with magnetic stir-bar agitation. When the OD_{600} reached ~0.6, fresh Davis rich medium was pumped in at 120-150 mL per h, with a waste needle set at 150 mL. Selection phage containing $RNAP_N$ protein fusions were used to seed fresh lagoons (25 mL flask with a rubber septum) in triplicate. For the mock PACS experiment, 40 μL of a 1000:1 mixture of $RNAP_N$ – iZA to $RNAP_N$ – RAF phage were used to inoculate the lagoons; for the DMS PACS experiment, 1 mL of the $RNAP_N$ – RAF error prone phage library was used to inoculate the lagoons. Waste needles were set to maintain the lagoon volume at 20-25 mL, and host cell cultures were flowed in at 20-25 mL per h. Lagoon samples were taken from the waste withdrawal line, centrifuged, sterile filtered, and the supernatant was stored at 4 °C. For the mock PACS experiment, lagoon samples were collected at 2 h, 4 h, 6 h, 8 h, 10 h, 12 h, and 24 h. For

the DMS PACS experiment, lagoon samples were collected at 2 h, 4 h, 6 h, 8 h, 10 h, 12 h, 24 h, 48 h, and 72 h.

PACE was performed in a similar fashion to PACS experiments, with some differences. S1030 cells were transformed by electroporation with the positive selection plasmid (31-69), the negative selection plasmid (20-6), and the additional *in vivo* mutagenesis plasmid MP6.[50] A 5 mL starter culture was grown to saturation overnight In LB media supplemented with 50 μg/mL carbenicillin, 33 μg/mL chloramphenicol, 40 μg/mL kanamycin, 7.5 μg/mL tetracycline, and 20 mM glucose. For the PACE reversion experiment, 40 μL of single clonal phage populations of each $RNAP_N$ – RAF R89L, K84E, or Q66K phage were used to inoculate sperate fresh lagoons. Mutagenesis was initiated by the direct addition of arabinose (20% w/v in water) into each lagoon at a rate of 0.5 mL per h. Lagoon samples were taken at 24 h.

Preparing PACS library Amplicons for HTS

Amplicons for HTS on the Illumina MiSeq platform were prepared by amplifying the gene of interest from the phage supernatant of the initial RAF phage library and each time point collected during the DMS experiment using PCR. For the first PCR (PCR 1), additional diversity was added with a variable sequence of 6-9 nucleotides, and complementary regions to the i7 and i5 Illumina TruSeq adaptors were added using the indicated primers below. Briefly, 50 μL PCR reactions set up with the following final concentrations: 1 μL phage supernatant, 0.5 μM forward primer, 0.5 μM reverse primer, 1x standard NEB Q5 HF buffer, 0.2 mM dNTPs, 0.02 U/μL NEB Q5 HF polymerase. The thermocycler conditions utilized an initial denaturation step (98°C, 1 min) followed by 5 cycles of denaturation (98°C, 10 sec), annealing (60°C, 15 sec) and extension (72°C, 10 sec), and a final extension (72°C, 2 min). The PCR product was purified using a Zymo clean and concentrator kit, and eluted in 20 μL of water.

Primers for PCR 1:

The italic sequence is the region complementary to the immediate upstream and downstream region of the mutagenized RAF gene. The bold region is the random sequence of either 6 (pictured) to 9 nucleotides. The underlined region in the forward and reverse primers are the complementary regions for the Illumina TruSeq i7 adaptor and the i5 adaptor, respectively.

Forward primer:

5’ – <u>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT</u> – **NNNNNN** – *TCTGGCTCTGGCTCGAGC* – 3’

Reverse primer:

5’ – <u>ACACTCTTTCCCTACACGACGCTCTTCCGATCT</u> – **NNNNNN** – *ACAACACTCCGGCTGCAG* – 3’

The i7 and i5 Illumina TruSeq adaptors were added using PCR to give each time point from each replicate lagoon a unique i7/i5 barcode combination to allow for sample multiplexing. Briefly, 50 μL PCR reactions set up with the following final concentrations: 1.25 μL purified PCR 1 product, 0.5 μM forward primer, 0.5 μM reverse primer, 1x standard NEB Q5 HF buffer, 0.2 mM dNTPs, 0.02 U/μL NEB Q5 HF polymerase. The thermocycler conditions utilized an initial denaturation

step (98°C, 1 min) followed by 10 cycles of denaturation (98°C, 10 sec), annealing (58°C, 15 sec) and extension (72°C, 10 sec), and a final extension (72°C, 2 min). The PCR product was visualized using DNA agarose gel to verify the generation of the correctly sized product, and was purified using a Zymo clean and concentrator kit, and eluted in 20 µL of water.

The dsDNA concentration was determined using a QuBit dsDNA high sensitivity (HS) assay kit (Invitrogen, Ref Q32854). The purified PCR products were then pooled into a single DNA library by combining equal amounts of each PCR product into a single sample. The combined library was then purified using a Zymo clean and concentrator kit to remove additional salts as measured by a nanodrop (260/230 ratio: 2.0; 260/280 ratio: 1.8), then the dsDNA concentration was determined using the QuBit dsDNA HS assay kit (Invitrogen, Ref Q32854). The library was then diluted to a final concentration of 4 nM with water, which was stored in -20°C.

The amplicon library was prepared the day-of for HTS on the Illumina MiSeq platform using the Illumina MiSeq v2 reagent kit (300 cycles, MS-102-2002) and the associated "MiSeq: Denature and Dilute Libraries Guide" from Illumina. The procedures outlining how to prepare a final library concentration of 10 pM for the Illumina MiSeq V2 kit with a 25% PhiX spike in control were used. This mixture was kept on ice until run on the Illumina MiSeq instrument following the procedures outlined in the "Sequencing Analysis Viewer v2.4" guide for paired end reads.

HTS Data Processing

Initial data processing was performed on the Illumina MiSeq instrument using the Sequencing Analysis Viewer v2.4 software to demultiplex pooled samples based on their i7 and i5 indexes.

The demultiplexed fastq.gz files were downloaded, unzipped, and subjected to a standard data processing pipeline using a Python script (available upon reasonable request). Briefly, paired end reads were merged using the FLASH program[61] using the default parameters. Afterwards, all reads were trimmed to the open reading frame of the RAF protein sequenced (aa 52-90). Next the length of the trimmed reads and quality scores were determined. Those below a length of 100 nucleotides (trimmed sequence length is 117) or with a greater than 5% chance of not containing a sequencing error were removed from the data set. The nucleotide sequences were then translated into amino acid sequences, and each amino acid sequence with their total observed count was stored as our final data set used for further analysis.

Functional Score Calculation

The final data set was used for functional score calculation, which were calculated by first determining the frequency of each sequence for a given time point. To avoid issues with future data processing of variants that were unobserved in the initial library, sequence counts for each variant *x* at time *t* had an arbitrary value of 1 added.

$$Frequency\ of\ variant\ x\ at\ time\ t = \frac{(sequnce\ counts\ for\ variant\ x\ at\ time\ t) + 1}{total\ sequence\ counts\ at\ time\ t}$$

Functional scores were then determined by dividing the frequency of variant *x* at time *t*, by the frequency of variant *x* at time 0 (the observed counts for the initial library).

$$Functional\ score\ of\ variant\ x\ at\ time\ t = \frac{frequncy\ of\ variant\ x\ at\ time\ t}{frequncy\ of\ variant\ x\ at\ time\ 0}$$

Generation of Global Functional Score Plots

Global plots of functional scores over time were generated using a Python script. All sequences that appeared at any time over the course of the selection were given an arbitrary functional score of 1 at time 0 h to account for variants that appeared at later time points that were unobserved at the initial time point. Frequencies of each variant were calculated using the modified equation below. Any sequence below a calculated frequency of 0.001 was given a value of 0.001; this value was used because it was determined to be the minimum threshold frequency above the stochastic noise observed in the dataset. Functional scores were calculated as above, and plotted as the $\log_{10}$ of the functional score.

$$Modified\ frequency\ of\ variant\ x\ at\ time\ t = \frac{sequnce\ counts\ for\ variant\ x\ at\ time\ t}{total\ counts\ at\ time\ t}$$

Generation of Amino Acid Distribution Heat Maps

Amino acid distributions at each position for a given time point were determined using a Python script. An additional data processing step was performed on the final data set to remove any truncated proteins that had a premature stop codon; this was done by removing any sequences without an exact sequence length of 39 amino acids. The fraction of each amino acid present at each position was determined for each time point. These values were exported into excel where heat maps where generated.

The relative enrichment values for a particular amino acid *z* at a given position were calculated using the 72 h time point from the selection. This was done by taking the ratio of the percentage of amino acid *z* at 72 h over the percentage of amino acid *z* at 0 h in reference to the ratio of the wt amino acid (see equation below). The relative enrichment values were utilized to construct color coded heat maps showing the functional consequence of these amino acid substitutions. Any amino acid that wasn't observed in the data set for either the 0 h or 72 h time points, was colored gray.

$$Relative\ enrichment\ value = \log_{10}\left(\frac{\%\ aa\ at\ position\ x\ at\ 72\ h}{\%\ aa\ at\ position\ x\ at\ 0\ h}\right) - \log_{10}\left(\frac{\%\ wt\ aa\ at\ position\ x\ at\ 72\ h}{\%\ wt\ aa\ at\ position\ x\ at\ 0\ h}\right)$$

REFERENCES

(1) Braun, P.; Gingras, A. C. History of Protein-Protein Interactions: From Egg-White to Complex Networks. *Proteomics* **2012**, *12* (10), 1478–1498.

(2) Vidal, M.; Cusick, M. E.; Barabási, A. L. Interactome Networks and Human Disease. *Cell* **2011**, *144* (6), 986–998.

(3) Sahni, N.; Yi, S.; Taipale, M.; Fuxman Bass, J. I.; Coulombe-Huntington, J.; Yang, F.; Peng, J.; Weile, J.; Karras, G. I.; Wang, Y.; et al. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **2015**, *161* (3), 647–660.

(4) Jubb, H. C.; Pandurangan, A. P.; Turner, M. A.; Ochoa-Montaño, B.; Blundell, T. L.; Ascher, D. B. Mutations at Protein-Protein Interfaces: Small Changes over Big Surfaces Have Large Impacts on Human Health. *Prog. Biophys. Mol. Biol.* **2017**, *128*, 3–13.

(5) Laddach, A.; Ng, J. C. F.; Chung, S. S.; Fraternali, F. Genetic Variants and Protein–Protein Interactions: A Multidimensional Network-Centric View. *Curr. Opin. Struct. Biol.* **2018**, *50*, 82–90.

(6) Petta, I.; Lievens, S.; Libert, C.; Tavernier, J.; De Bosscher, K. Modulation of Protein-Protein Interactions for the Development of Novel Therapeutics. *Mol. Ther.* **2016**, *24* (4), 707–718.

(7) Valeur, E.; Jimonet, P. New Modalities, Technologies, and Partnerships in Probe and Lead Generation: Enabling a Mode-of-Action Centric Paradigm. *J. Med. Chem.* **2018**, *61* (20), 9004–9029.

(8) Valeur, E.; Narjes, F.; Ottmann, C.; Plowright, A. Emerging Modes-of-Action in Drug Discovery. *Medchemcomm* **2019**.

(9) Taipale, M. Disruption of Protein Function by Pathogenic Mutations: Common and Uncommon Mechanisms. *Biochem. Cell Biol.* **2018**, *97* (1), 46–57.

(10) Cunningham, B.; Wells, J. High-Resolution Epitope Mapping of HGH-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science* **1989**, *244* (4908), 1081–1085.

(11) Weiss, G. A.; Watanabe, C. K.; Zhong, A.; Goddard, A.; Sidhu, S. S. Rapid Mapping of Protein Functional Epitopes by Combinatorial Alanine Scanning [In Process Citation]. *Proc.Natl.Acad.Sci.U.S.A* **2000**, *97* (16), 8950–8954.

(12) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot Spots-A Review of the Protein-Protein Interface Determinant Amino-Acid Residues. *Proteins Struct. Funct. Bioinforma.* **2007**, *68* (4), 803–812.

(13) Goodyear, C. S.; Silverman, G. J. Phage-Display Methodology for the Study of Protein-Protein Interactions: Overview. *Cold Spring Harb. Protoc.* **2008**, *2008* (9), pdb.top48-pdb.top48.

(14) Younger, D.; Berger, S.; Baker, D.; Klavins, E. High-Throughput Characterization of Protein–Protein Interactions by Reprogramming Yeast Mating. *Proc. Natl. Acad. Sci.* **2017**, *114* (46), 12166–12171.

(15) Lin, J.-S.; Lai, E.-M. Protein–Protein Interactions: Yeast Two-Hybrid System; 2017; pp 177–187.

(16) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (12), 866–876.

(17) Erffelinck, M. L.; Ribeiro, B.; Perassolo, M.; Pauwels, L.; Pollier, J.; Storme, V.; Goossens, A. A User-Friendly Platform for Yeast Two-Hybrid Library Screening Using next Generation Sequencing. *PLoS One* **2018**, *13* (12), 1–21.

(18) Fowler, D. M.; Fields, S. Deep Mutational Scanning: A New Style of Protein Science. *Nat. Methods* **2014**, *11* (8), 801–807.

(19) Araya, C. L.; Fowler, D. M. Deep Mutational Scanning: Assessing Protein Function on a Massive Scale. *Trends Biotechnol.* **2011**, *29* (9), 435–442.

(20) Fowler, D. M.; Stephany, J. J.; Fields, S. Measuring the Activity of Protein Variants on a Large Scale Using Deep Mutational Scanning. *Nat. Protoc.* **2014**, *9* (9), 2267–2284.

(21) Starita, L. M.; Young, D. L.; Islam, M.; Kitzman, J. O.; Gullingsrud, J.; Hause, R. J.; Fowler, D. M.; Parvin, J. D.; Shendure, J.; Fields, S. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics* **2015**, *200* (2), 413–422.

(22) Araya, C. L.; Fowler, D. M.; Chen, W.; Muniez, I.; Kelly, J. W.; Fields, S. A Fundamental Protein Property, Thermodynamic Stability, Revealed Solely from Large-Scale Measurements of Protein Function. *Proc. Natl. Acad. Sci.* **2012**, *109* (42), 16858–16863.

(23) Fowler, D. M.; Araya, C. L.; Fleishman, S. J.; Kellogg, E. H.; Stephany, J. J.; Baker, D.; Fields, S. High-Resolution Mapping of Protein Sequence-Function Relationships. *Nat. Methods* **2010**, *7* (9), 741–746.

(24) Stiffler, M. A.; Subramanian, S. K.; Salinas, V. H.; Ranganathan, R. A Protocol for Functional Assessment of Whole-Protein Saturation Mutagenesis Libraries Utilizing High-Throughput Sequencing. *J. Vis. Exp.* **2016**, No. 113, 1–11.

(25) Roscoe, B. P.; Thayer, K. M.; Zeldovich, K. B.; Fushman, D.; Bolon, D. N. A. Analyses of the Effects of All Ubiquitin Point Mutants on Yeast Growth Rate. *J. Mol. Biol.* **2013**, *425* (8), 1363–1377.

(26) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16* (7), 379–394.

(27) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472* (7344), 499–503.

(28) Suzuki, T.; Miller, C.; Guo, L. T.; Ho, J. M. L.; Bryson, D. I.; Wang, Y. S.; Liu, D. R.; Söll, D. Crystal Structures Reveal an Elusive Functional Domain of Pyrrolysyl-TRNA Synthetase. *Nat. Chem. Biol.* **2017**, *13* (12), 1261–1266.

(29) Roth, T. B.; Woolston, B. M.; Stephanopoulos, G.; Liu, D. R. Phage-Assisted Evolution of Bacillus Methanolicus Methanol Dehydrogenase 2. *ACS Synth. Biol.* **2019**, *8* (4), 796–806.

(30) Leconte, A. M.; Dickinson, B. C.; Yang, D. D.; Chen, I. a; Allen, B.; Liu, D. R. A Population-Based Experimental Model for Protein Evolution: Effects of Mutation Rate and Selection Stringency on Evolutionary Outcomes. *Biochemistry* **2013**, *52* (8), 1490–1499.

(31) Packer, M. S.; Rees, H. A.; Liu, D. R. Phage-Assisted Continuous Evolution of Proteases with Altered Substrate Specificity. *Nat. Commun.* **2017**, *8* (1).

(32) Dickinson, B. C.; Packer, M. S.; Badran, A. H.; Liu, D. R. A System for the Continuous Directed Evolution of Proteases Rapidly Reveals Drug-Resistance Mutations. *Nat. Commun.* **2014**, *5*, 1–8.

(33) Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L.; Sun, N.; Zeina, C. M.; Gao, X.; Rees, H. A.; Lin, Z.; et al. Evolved Cas9 Variants with Broad PAM Compatibility and High DNA Specificity. *Nature* **2018**, *556* (7699), 57–63.

(34) Wang, T.; Badran, A. H.; Huang, T. P.; Liu, D. R. Continuous Directed Evolution of Proteins with Improved Soluble Expression. *Nat. Chem. Biol.* **2018**, *14* (10), 972–980.

(35) Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P.; Kain, W.; Nance, A.

M.; Evdokimov, A.; Moshiri, F.; Turner, K. H.; et al. Continuous Evolution of Bacillus Thuringiensis Toxins Overcomes Insect Resistance. *Nature* **2016**, *533* (7601), 58–63.

(36) Prior, I. A.; Lewis, P. D.; Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* **2012**, *72* (10), 2457–2467.

(37) Davies, H.; Bignell, G. R.; Cox, C.; Stephens, P.; Edkins, S.; Clegg, S.; Teague, J.; Woffendin, H.; Garnett, M. J.; Bottomley, W.; et al. Mutations of the BRAF Gene in Human Cancer. *Nature* **2002**, *417* (6892), 949–954.

(38) Hobbs, G. A.; Der, C. J.; Rossman, K. L. RAS Isoforms and Mutations in Cancer at a Glance. *J. Cell Sci.* **2016**, *129* (7), 1287–1292.

(39) Scheffler, J. E.; Waugh, D. S.; Bekesi, E.; Kiefer, S. E.; LoSardo, J. E.; Neri, A.; Prinzo, K. M.; Tsao, K. L.; Wegrzynski, B.; Emerson, S. D.; et al. Characterization of a 78-Residue Fragment of c-Raf-1 That Comprises a Minimal Binding Domain for the Interaction with Ras-GTP. *J. Biol. Chem.* **1994**, *269* (35), 22340–22346.

(40) Vojtek, A. B.; Hollenberg, S. M.; Cooper, J. A. Mammalian Ras Interacts Directly with the Serine/Threonine Kinase Raf. *Cell* **1993**, *74* (1), 205–214.

(41) Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. B. C. Evolution of a Split RNA Polymerase as a Versatile Biosensor Platform. *Nat. Chem. Biol.* **2017**, *13* (4), 432–438.

(42) Pu, J.; Dewey, J. A.; Hadji, A.; Labelle, J. L.; Dickinson, B. C. RNA Polymerase Tags to Monitor Multidimensional Protein-Protein Interactions Reveal Pharmacological Engagement of Bcl-2 Proteins. *J. Am. Chem. Soc.* **2017**, *139* (34), 11964–11972.

(43) Pu, J.; Kentala, K.; Dickinson, B. C. Multidimensional Control of Cas9 by Evolved RNA Polymerase-Based Biosensors. *ACS Chem. Biol.* **2018**, *13* (2), 431–437.

(44) Kirchner, G.; Roberts, J. L.; Gustafson, G. D.; Ingolia, T. D. Active Bacterial Luciferase from a Fused Gene: Expression of a Vibrio Harveyi LuxAB Translational Fusion in Bacteria, Yeast and Plant Cells. *Gene* **1989**, *81* (2), 349–354.

(45) Ghosh, I.; Hamilton, A. D.; Regan, L.; V, Y. U.; V, N. H. Antiparallel Leucine Zipper-Directed Protein Reassembly : Application to the Green Fluorescent Protein Department of Molecular Biophysics and Biochemistry The Dissection and Subsequent Reassembly of a Protein from Peptidic Fragments Provides an Avenue For. *J. Am. Chem. Soc.* **2000**, *122* (11), 5658–5659.

(46) Fabian, J. R.; Vojtekt, A. B.; Coopert, J. A.; Morrison, D. K. A Single Amino Acid Change in Raf-1 Inhibits Ras Binding and Alters Raf-1 Function (Signal Transduction/GTP-Binding Protein/Protein-Serine/Threonine Kinases). *Biochemistry* **1994**, *91* (June), 5982–5986.

(47) Ellefson, J. W.; Meyer, A. J.; Hughes, R. A.; Cannon, J. R.; Brodbelt, J. S.; Ellington, A. D. Directed Evolution of Genetic Parts and Circuits by Compartmentalized Partnered Replication. *Nat. Biotechnol.* **2014**, *32* (1), 97–101.

(48) Bennett, N. J.; Rakonjac, J. Unlocking of the Filamentous Bacteriophage Virion During Infection Is Mediated by the C Domain of PIII. *J. Mol. Biol.* **2006**, *356* (2), 266–273.

(49) Block, C.; Janknecht, R.; Herrmann, C.; Nassar, N.; Wittinghofer, A. Quantitative Structure-Activity Analysis Correlating Ras/Raf Interaction in Vitro to Raf Activation in Vivo. *Nat. Struct. Biol.* **1996**, *3* (3), 244–251.

(50) Badran, A. H.; Liu, D. R. Development of Potent in Vivo Mutagenesis Plasmids with Broad Mutational Spectra. *Nat. Commun.* **2015**, *6*, 1–10.

(51) Cadwell, R. C.; Joyce, G. F. Randomization of Genes by PCR Mutagenesis. *PCR Methods Appl.* **1992**, *2* (1), 28–33.

(52) Marshall, M. Interactions between Ras and Raf: Key Regulatory Proteins in Cellular Transformation. *Mol. Reprod. Dev.* **1995**, *42* (4), 493–499.

(53) Campbell-Valois, F. X.; Tarassov, K.; Michnick, S. W. Massive Sequence Perturbation of the Raf Ras Binding Domain Reveals Relationships between Sequence Conservation, Secondary Structure Propensity, Hydrophobic Core Organization and Stability. *J. Mol. Biol.* **2006**, *362* (1), 151–171.

(54) Vanhercke, T.; Ampe, C.; Tirry, L.; Denolf, P. Reducing Mutational Bias in Random Protein Libraries. *Anal. Biochem.* **2005**, *339* (1), 9–14.

(55) Moore, C. L.; Papa, L. J.; Shoulders, M. D. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions in Vivo. *J. Am. Chem. Soc.* **2018**, *140* (37), 11560–11564.

(56) Pu, J.; Disare, M.; Dickinson, B. C. Evolution of C-Terminal Modification Tolerance in Full-Length and Split T7 RNA Polymerase Biosensors. *ChemBioChem* **2019**, *20* (12), 1547–1553.

(57) English, J. G.; Olsen, R. H. J.; Lansu, K.; Patel, M.; White, K.; Cockrell, A. S.; Singh, D.; Strachan, R. T.; Wacker, D.; Roth, B. L. VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell* **2019**, 1–14.

(58) Berman, C. M.; Papa, L. J.; Hendel, S. J.; Moore, C. L.; Suen, P. H.; Weickhardt, A. F.; Doan, N.-D. D.; Kumar, C. M.; Uil, T. G.; Butty, V. L.; et al. An Adaptable Platform for Directed Evolution in Human Cells. *J. Am. Chem. Soc.* **2018**, *140* (51), 18093–18103.

(59) Carlson, J. C.; Badran, A. H.; Guggiana-Nilo, D. A.; Liu, D. R. Negative Selection and Stringency Modulation in Phage-Assisted Continuous Evolution. *Nat. Chem. Biol.* **2014**, *10* (3), 216–222.

(60) de Boer, H. A.; Comstock, L. J.; Vasser, M. The Tac Promoter: A Functional Hybrid Derived from the Trp and Lac Promoters. *Proc. Natl. Acad. Sci.* **1983**, *80* (1), 21–25.

(61) Magoč, T.; Salzberg, S. L. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* **2011**, *27* (21), 2957–2963.