Multiplex mapping of protein-protein interaction interfaces

Jingxuan He[1]*, Ling-Nan Zou[1]*, Vidhi Pareek[2], Stephen J. Benkovic[1]
[1]Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA
[2]Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA
* These authors contributed equally to this work

Corresponding: Jingxuan He (juh709@psu.edu), Ling-Nan Zou (lxz7@psu.edu)

Abstract

We describe peptide mapping through *Sp*lit *A*ntibiotic *R*esistance *C*omplementation (SpARC-map), a method to identify the probable interface between two interacting proteins. Our method is based on in vivo affinity selection inside a bacterial host, and uses high throughput DNA sequencing results to infer the location of protein-protein interaction (PPI) interfaces. SpARC-map uses only routine microbiology techniques, with no reliance on specialized instrumentation or reconstituting protein complexes in vitro; it can be tuned to detect PPIs over a broad range of affinities; it can be multiplexed to probe multiple PPIs in parallel; its nonspecific background can be precisely measured, enabling the sensitive detection of weak PPIs. Using SpARC-map, we recover the known interface in the (p21-PCNA complex. We also use SpARC-map to probe the purinosome, the weakly bound complex of six purine biosynthetic enzymes, where no PPI interfaces are known. There, we identify interfaces that satisfy structural requirements for substrate channeling; we also identify protein surfaces that participate in multiple distinct interactions, which we validate using site-specific photocrosslinking in live human cells. Finally, we show that SpARC-map results can impose stringent constraints on outputs from machine learning based structure prediction.

Main Text

Introduction

In living cells, protein-protein interactions (PPIs) are vital to biological function. Identifying PPIs, and elucidating the composition/structure of protein complexes, are core tasks in molecular and cellular biology research. The location of the PPI interface can provide a wealth of valuable insight: it can illuminate the structure of the protein complex, reveal the functional consequences of the interaction, and point toward cellular mechanisms that regulate it (1, 2). PPI interfaces are also attractive drug targets, and disrupting disease-relevant PPIs is a promising approach towards precision medicine (3–6). Yet, most PPIs interfaces remain unknown. In large part, this is because structural biology tools such as x-ray protein crystallography and cryo-electron microscopy have significant instrumental/technical demands that limit their accessibility. Furthermore, these tools can only resolve protein complexes that can be reconstituted in vitro, and their throughput is low. Advances in machine learning (ML) based structural prediction (7–10) raises the possibility that where experimental structures are absent, computation models can fill in the gap. However, compared to the large dataset of protein structures available to train ML based modeling, the number of protein complexes with solved structures is far smaller; consequently, the prediction accuracy for protein complexes still requires further evaluation. The need for experimental evidence remains.

Here, we describe peptide-mapping via split-antibiotic resistance complementation (SpARC-map), an indirect method for mapping the probable interface between two interacting proteins. It uses in vivo affinity scanning inside a bacterial host, read out via high throughput DNA sequencing, to infer the location of PPI interfaces; its resolution is comparable to other indirect mapping methods, such as hydrogen-deuterium exchange (11) and protein painting (12), that use mass spectrometry for readout. SpARC-map employs only routine microbiological methods, without reliance on specialized instrumentation. It does not require reconstituting protein complexes in vitro, is tunable to detect PPIs over a broad range of binding affinities, and can be multiplexed to probe multiple PPIs in parallel. The nonspecific background of SpARC-map can be precisely measured and characterized, allowing its users to make a rigorous and quantitative distinction between specific and nonspecific interactions, thereby enabling the detection of weak, but still specific, interactions.

As a proof of concept, we use SpARC-map to recover the known PPI interface of the p21-PCNA complex. As a further test, we use it to map, in multiplex, probable PPI interfaces in the purinosome, the multi-enzyme complex responsible for de novo purine biosynthesis (13). This weakly-bound complex has never been reconstituted in vitro, in part or in toto, and no PPI interfaces are known (14). Using SpARC-map, we identify probable PPI interfaces that satisfy structural demands for substrate channeling (15); multispecific protein surfaces that participate in multiple PPIs, which we validate using site-specific photocrosslinking in live cells; and interfaces decorated with post translational modifications (PTMs) that may regulate purinosome assembly. Our results demonstrate SpARC-map can identify true PPI interfaces in even weakly-bound protein complexes, to reveal biologically significant insights into their structure, function, and regulation.

Results

Peptide mapping through split-antibiotic resistance complementation. SpARC-map is motivated by the observation that many PPIs are mediated by small linear or structural motifs, or are anchored by a small number of hot spots (16–20). The concept is simple: we scan, piecewise, the amino acid sequence of a protein (prey), to identify prey peptide fragments that show affinity for its interaction partner (bait); we expect these peptides will correspond to protein domains containing, in part or in toto, the bait-prey interface (Fig. 1A). While such a prey peptide is unlikely to adopt same structure as the corresponding region on the intact prey protein, so long as the ensemble of peptide conformations overlaps that of the (local) protein structure, the bait-prey peptide interaction can mimic the bait-intact prey interaction (21–23). However, the binding affinity will be reduced due to the entropic penalty of locking in prey peptide conformation upon its binding to the bait.

We begin by constructing a library of bacterial expression vectors (SpARC vectors); these contain a bicistronic cassette expressing both the bait protein and a small prey peptide (Fig. 1C). The prey peptide coding sequences (100-300 bp) are generated by shearing dsDNA containing the full prey protein coding sequence (CDS) (Fig. 1B). To affinity-select bait-binding prey peptides, we engineered host antibiotic resistance conditional on bait-prey peptide binding. We split aminoglycoside phosphotransferase (neo^R), which inactivates kanamycin, into neo^RN (AAs 1-59) and neo^RC (AAs 59-264) fragments. Neither alone can inactivate kanamycin, but when fused to interacting proteins, neo^RN and neo^RC fragments can reconstitute a functional neo^R (24). By fusing the bait and the prey peptide respectively to neo^RN and neo^RC, SpARC vectors expressing bait-binding prey peptides can endow their host with kanamycin resistance (Fig. 1D) — if the binding affinity is sufficiently high.

Selection in SpARC-map is tunable to detect PPIs of varying affinities. In the SpARC vector, bait and prey peptide expressions are controlled by orthogonal inducible promoters: isopropyl-β-D-thiogalactopyranoside (IPTG) inducible for bait, anhydrotetracycline (aTc) inducible for prey peptide (Fig. 1C). High bait and/or prey peptide expression favors complexation, enabling even weak interactions to reconstitute sufficient neo^R activity to result in kanamycin resistance. Conversely, low bait and/or prey peptide expression means only strong bait-prey peptide interactions can lead to kanamycin resistance. By adjusting concentrations of kanamycin and inducers, we estimate we can tune SpARC-map to discriminate bait-prey peptide affinities from K_D = 10 nM - 1 mM (Supplementary Materials). Post selection, we identify probable bait-binding regions on the prey using next-generation sequencing (NGS) to profile SpARC vectors from resistant bacteria (Fig. 1E).

Identifying the non-specific background. Aside from direct structure determination, all PPI assays must contend with distinguishing between specific and non-specific interactions. Starting from the reasonable assumption that specific interactions should be stronger than non-specific ones, most assays use ad hoc methods to this distinction; when the expected specific PPI is weak, this distinction can be difficult to make. Here we note that if the background signal of nonspecific interactions, defined in the case of SpARC-map as the binding between the bait and *random* peptides, can be precisely measured and understood, then given measurement of a bona fide peptide fragment of the prey, we can rigorously evaluate whether its interaction with the bait is non-specific. Prey peptides that are unlikely to be non-specific binders are presumably specific interactors that contain portions of the bait-prey PPI interface.

Constructing the SpARC library automatically supplies a pool of random DNA sequences, encoding random peptides, with which we can directly measure the bait-random peptide interaction background. Because prey peptide coding sequences are generated by shearing dsDNA containing the full prey CDS, only 1/18 of resulting dsDNA fragments will ligate in-sense and translate into a prey peptide fused in frame to neo^{R}C. The remainder will be inverted, and/or frameshifted, and/or contain extra bases. Most such nonsense inserts will disrupt in frame fusion with neo^{R}C on the peptide C-terminus, or contain stop codons; SpARC vectors with such inserts cannot result in kanamycin resistance (Fig. 2A). However, many nonsense inserts will translate through to effectively random peptides fused to neo^{R}C; these can interact with the bait and potentially reconstitute neo^R. We can recognize such random peptides in the NGS results, and for each species i, quantify its abundance n_i. The distribution $P_{NS}(n_i)$ is the non-specific background due to bait-random peptide binding.

For a random peptide to endow its host with sufficient kanamycin resistance to survive selection, its bait-binding affinity must exceed the threshold determined by the selection condition: greater the exceedance, greater the survival. The Pickands–Balkema–De Haan Theorem, a central result of Extreme Value Theory, states the distribution of threshold exceedances, under generous conditions, follows a Generalized Pareto Distribution (GPD) (25). We therefore model the measured non-specific background as $P_{\mathrm{NS}}(n_i) = \mathrm{GPD}(n_i; \mu, \sigma, \xi)$, with parameters μ, σ, and ξ (Fig. 2B) (26). Then, given a NGS species encoding a bona fide prey peptide, we can used the modelled $P_{\mathrm{NS}}(n_i)$ to evaluate, e.g. by computing appropriate p-values, how well its abundance can be explained by bait-random peptide binding (Fig. 2C). Prey peptides whose abundances are consistent with random binding should be excluded from delineating the probable PPI interface.

Mapping the p21-PCNA PPI interface. As a proof-of-concept, we use SpARC-map to probe the known PPI interface between p21 and PCNA. PCNA, a homo-trimer, forms a ring-shaped DNA clamp that is essential for DNA replication and repair. p21, better known as an inhibitor of cell-cycle dependent kinases, can bind PCNA and inhibit its normal functions (27–29). The PCNA-binding region of p21 is its C-terminal PIP-box motif (AAs 140-164), first identified through analyses of p21 truncations (30, 31). Subsequently, the complete PPI interface was revealed by the x-ray crystal structure of PCNA in complex with p21$^{139\text{-}160}$ (32).

We constructed two libraries: one with PCNA as bait and p21 CDS fragments as prey, another with p21 as bait and PCNA CDS fragments as prey (Supplementary Materials). Figure 3 shows the PCNA-p21 atomic contact map (Fig. 3A) juxtaposed with affinity heatmaps generated from these libraries post-selection (Figs. 3B-C); for heatmaps, x-axis indicates prey residue number, and y-axis indicates the probability the prey residue is tiled by a mapped NGS read. For PCNA binding on p21, including all p21 peptides in the analysis resulted in a heatmap covering the entire p21 amino acid sequence. Its main feature is the peak spanning AAs 139-160, corresponding to the true p21 interface with PCNA (Fig. 3B); considering only species that exceed the random peptide background further suppresses heatmap signal outside this region. From the tiling of the interface by distinct p21 fragments, we infer while p21$^{139\text{-}164}$ has the highest PCNA affinity (is the most abundant), truncation to p21$^{143\text{-}158}$ still retains PCNA-binding but at a lower affinity (has lower abundance). This is consistent with known PCNA binding affinities of p21 peptides: K_D = 6.4±2.8 nM for p21$^{140\text{-}163}$, and 67±9 nM for p21$^{143\text{-}157}$ (33).

For the reciprocal PCNA interface that binds p21, including all species results in a heatmap containing two comparably sized features, AAs 104-139 and 152-207, along the PCNA peptide sequence. Once we restrict analysis to species that significantly exceed the random peptide background, the feature corresponding to AAs 152-207 vanishes, leaving only the hotspot on AAs 104-139, with $PCNA^{115-133}$ being the most abundant species (Fig. 3C). This is in excellent agreement with the crystal structure of PCNA in complex with $p21^{139-160}$, where the main PCNA interface with p21, accounting for ~50% of atomic contacts, is on PCNA AAs 118-131 (Fig. 3D).

Mapping PPI interfaces in the purinosome. Compared to p21-PCNA, the purinosome is far more complicated protein complex. It consists of six enzymes — PPAT, GART, PFAS, PAICS, ADSL, and ATIC — responsible for de novo purine biosynthesis (DNPB) (13, 14). Imaging studies have shown these enzymes can colocalize into punctate bodies in purine-starved cells, and multiple molecular studies have found evidence for PPIs (34–39). However, it is unknown if any of these interactions are direct. All evidence indicates the purinosome is a weakly-bound — it has never been reconstituted in vitro, wholly or partially, and no PPI interfaces are known.

To map probable PPI interfaces in the purinosome, we constructed and screened six libraries, one per each purinosome enzyme as bait; all of them have as prey CDS fragments from all six enzymes in multiplex (Supplementary Materials, Table S1). We focus here on probable PPI interfaces between sequential enzymes. Given the clear metabolomic signatures of substrate channeling in the purinosome, we expect at least some enzyme pairs that catalyze sequential reactions should physically interact in such a manner as to bring up- and downstream enzyme domains and/or active sites into proximity, to facilitate direct substrate transfer between them.

PPAT-GART. PRA, product of PPAT in reaction 1 of DNPB, is the substrate of GART's GAR synthase domain for reaction 2. The instability of PRA (half-life <1 minute) was early motivation for suggesting PPAT and GART must physically interact to transfer such a labile intermediate (40, 41). On GART, we found probable PPAT binding sites on AAs 14-55 ($p<0.001$) and 159-232 ($p<0.003$); these flank the GAR synthase active site that receives PRA from PPAT (Fig. 4A), as expected from the structural requirements substrate channeling. On PPAT, the probable GART-binding site is on its C-terminus, AAs 468-504 ($p<0.001$) (Fig. 4B).

GART-PFAS. GART contains three catalytic domains that catalyze reactions 2, 3, 5 of DNPB. PFAS catalyzes reaction 4, receiving its substrate FGAR from GART's GAR transformylase domain, and sending its product FGAM forward to GART's AIR synthase domain. On GART, we found three probable PFAS-binding sites ($p<0.003$): AAs 24-48, AAs 773-801 on the AIR synthase domain that receives FGAM from PFAS, and AAs 1001-1010 on the GAR transformylase domain that supplies FGAR to PFAS (Fig. 4C); the latter two are structurally consistent with substrate channeling from GART to PFAS and back. We could not identify the reciprocal GART-binding interface on PFAS.

GART-PAICS. AIR, the product of GART in reaction 5, is the first substrate of PAICS, which catalyzes reactions 6 and 7. On GART, we found a probable PAICS binding site, AAs 618-669 ($p<10^{-3}$), on the AIR synthase domain that supplies AIR to PAICS (Fig. 4D), structurally consistent with substrate transfer from GART to PAICS. We could not identify the reciprocal GART-binding interface on PAICS.

PAICS-ADSL. SAICAR, the product of PAICS in reaction 7, is the substrate for ADSL for reaction 8. On ADSL, we found a probable PAICS binding site on AAs 451-484 (p<0.001) (Fig. 4E). On PAICS, we identified three potential ADSL-binding interfaces, AAs 251-270, 281-348, and 413-425, with low confidence (p<0.004 – 0.01); two of these, AAs 251-270 and 413-425, are located on PAICS's SAICAR synthase domain that supplies ADSL.

ADSL-ATIC. ADSL supplies its product AICAR as the initial substrate for ATIC, which catalyzes reactions 9 and 10. On ATIC, we found probable ADSL binding site on AAs 174-214 (p<0.002) and 489-529 (p<0.003); the latter lines the active site entrance on ATIC's AICAR transformylase domain that receives AICAR from ADSL (Fig. 4F), and is structurally consistent with substrate channeling from ADSL to GART. We identified two low confidence (p<0.008) ATIC-binding sites on ADSL, AAs 73-130 and 447-484; the latter coincides with the probable PAICS-binding site on ADSL.

In addition to expected PPI interfaces between sequential enzymes, we also found probable PPI interfaces between enzymes that are distal along the DNPB pathway (Supplemental Materials, Table S2). Indeed, we found a dense network of probable PPI interfaces (at least partially), indicative of a direct physical interaction, connecting every enzyme pair. Surprisingly, some enzymes appear to use the same surface regions to participate in multiple PPIs. For example, ADSL appears to use the same C-terminal region (AAs 451-484) to interact with distal enzymes PPAT and PFAS, as well as enzymes that catalyze immediate upstream (PAICS) and downstream (ATIC) reactions; however, the probable GART-binding site on ADSL (AAs 391-402) is distinct.

Validating PPI interfaces via in vivo photocrosslinking. While many of the probable purinosome PPI interfaces (e.g. between sequential enzymes) identified by SpARC-map are expected (consistent with structural requirements of substrate channeling), others (e.g. the multispecific C-terminus region of ADSL) were surprising. We therefore sought to validate SpARC-map identifications using site-specific photocrosslinking in live human cells. To do so, we use genetically encoded unnatural amino acid incorporation to insert *p*-azido-L-phenylalanine (azF) at defined sites on a protein (42–44). Upon UV-activation, the aryl azide group on azF can attack nearby double bonds, primary amines, C-H and N-H groups, etc., to form a covalent bond. Thus azF, when inserted into the surface of a bait protein, can photocrosslink prey proteins that interact with the bait via that surface (Fig. 5A). The small reactive radius (<0.3 nm) and short lifetime (~1 ns post-photoactivation) of azF minimizes nonspecific crosslinking due to random molecular collisions.

We focused on validating $ADSL^{451-484}$ as the multispecific PPI interface between ADSL and other purinosome enzymes. We constructed a mutant ADSL containing the substitution Y466azF, as well as a 2×StepTag epitope tag fused to its N-terminus, and expressed it in HEK293T cells cultured in purine-depleted media. Following brief UV irradiation (for photocrosslinking samples), we lyse the cells and pull-down $ADSL^{Y466azF}$ (plus its crosslinked adducts and co-precipitants) for analysis using protein mass spectrometry (Fig. 5B). Pulled-down proteins should include those that interact specifically with ADSL but not via the C-terminal region, and those that interact with ADSL specifically via $ADSL^{451-484}$. Abundance of the latter, but not the former, should be enhanced by UV-irradiation.

Of the four DNPB enzymes (PPAT, PFAS, PAICS, ATIC) identified by SpARC-map to interact with ADSL via $ADSL^{451-484}$, three (PFAS, PAICS, and ATICS) were low/undetectable in non-irradiated controls, but abundant in UV-irradiated samples. PPAT was only detected upon UV-irradiation, but at low levels

that did not result in statistically significant differences between UV-irradiated versus non-irradiated samples. GART, which SpARC-map indicated interacts with ADSL but not via ADSL$^{451-484}$, was abundantly detected, without statistically significant differences, in both UV-irradiated and non-irradiated samples (Fig. 5C-D). These results support SpARC-map's identification of ADSL$^{451-484}$ as a multispecific interface with multiple purinosome enzymes.

Constraining AlphaFold3 predictions with SpARC-map data. For protein complexes that lack experimentally determined structures, high resolution models from ML based structure prediction may be able to fill in the gap. SpARC-map can provide an experimental grounding for evaluating the plausibility of predicted structures. Using the PPAT-GART complex as an example, we examined whether PPI interfaces identified by SpARC-map can also be found in AlphaFold3 (AF3) predictions. This interaction was long argued to be biochemically necessary, but direct physical evidence had been lacking. We generated 305 randomly seeded AF3 models of the complex comprised of 4 copies of PPAT, which can exist in both homo-dimeric and homo-tetrameric forms, and 2 copies of GART, which is homo-dimeric. For each model, we identify all atomic contacts (separation <0.4 nm) between PPAT and GART monomers. While models did not exhibit a consensus on the location of the PPAT-GART interface (Fig. 6A), a majority identified PPAT contacting GART on GART's AIR synthase (AAs 434-809) and/or GAR transformylase (AAs 809-1010) domains. These configurations are incompatible with structural requirements for substrate channeling between PPAT and GART, and are inconsistent with SpARC-map results. Of the regions identified by SpARC-map as PPI interfaces, PPAT AAs 468-504, contacting GART, was found in ~10% of models, while GART AAs 14-55 and 159-232, contacting PPAT, was found in ~2% of the models; only a handful of models possessed PPI interfaces consistent with SpARC-map results (Fig. 6B). Here, SpARC-map results present strong, experimentally grounded constraints for evaluating plausibility of AF3 structural predictions.

Discussion

Here, we described SpARC-map, a method for identifying the probable PPI interface between two interacting proteins. For the PCNA/p21 complex, SpARC-map recovered the known PPI interface. For the purinosome, SpARC-map identified hereto unknown PPI interfaces that satisfy structural requirements for substrate channeling, as well multispecific interfaces that participate in multiple PPIs. Using site-specific photocrosslinking in vivo, we validated the identification of ADSL$^{451-484}$ as a multispecific interface via which ADSL interacts with several, but not all, purinosome enzymes. We are now in the process of validating other probable PPI interfaces in the purinosome. Despite not being fully validated, our SpARC-map data can already provide tantalizing hints into the still obscure cellular mechanisms that regulate purinosome assembly. Several probable purinosome PPI interfaces are decorated with post-transcriptional modifications, e.g. phosphorylations on PPAT T356, GART S625, ADSL Y466, and ATIC T520 (45); these may modulate the strength of PPIs mediated by the decorated interfaces. If true, identifying the responsible modifying enzymes (e.g. kinases) will establish direct connections between cellular signaling pathways and purinosome assembly.

SpARC-map results also provided us new insights into the structural nature of the purinosome. We have previously found C-terminal tagged ADSL is defective at rescuing ADSL-knockout HeLa (Supplementary Materials). This was puzzling because ADSL's C-terminal region does not contribute to known ligand-binding or active sites (46). Recognizing that this region mediates PPIs between ADSL and multiple purinosome enzymes suggests C-terminal tagging of ADSL disrupts these PPIs,

leading to the loss of substrate channeling and possibly destabilizing the purinosome entirely. Given the density of probable direct interactions in the purinosome revealed by SpARC-map, it is unlikely all six enzymes can arrange into a single definite structure that satisfies all binary contacts. This, and the presence of multispecific binding regions, suggest the purinosome is probably not a classic protein complex with an exact stoichiometry. Intrinsically disordered proteins with polyvalent PPIs are known organizers of "fuzzy" complexes such as liquid-like protein condensates (47–50). Polyvalency in structured proteins with multispecific binding sites, e.g. ADSL, further amplified through homo-oligomerization (ADSL is a homo-tetramer), may play a similar role in organizing the purinosome via so-called bridging attractions (51–53).

Presently, SpARC-map can only map PPIs involving bait proteins that express and properly fold in *E. coli*, although it is adaptable to eukaryotic expression systems. PPI interfaces made up of physically proximal but sequentially distal residues, or require conformation changes (e.g., due to allostery) or post translational modifications, may also escape detection. SpARC-map generally fails to detect homo-oligomeric interfaces, because doing so requires prey peptides to outcompete intact proteins for binding, which suppresses their signal. On the other hand, SpARC-map is suited for mapping PPIs involving disordered proteins where direct structure determination may not be feasible. Despite these limitations, we note the BioGRID database currently curates approximately one million physical interactions in the human proteome (54); if SpARC-map captures the interface from just a fraction of these, it will constitute a massive expansion in our knowledge of protein complexes. Our present perspective is largely informed by tightly bound protein complexes with definite structures and stoichiometries, because these are the most amenable to study. Protein complexes bound by dense networks of weak, possible polyvalent PPIs, and which may lack definite structures and stoichiometries, may be far more typical, but have escaped appreciation because they are difficult to probe. SpARC-map is a tool that enables us to examine this largely unseen world. Proteins inside a cell never act in isolation; it is their concerted action, with each other and with other biomolecules, that gives the cell life. With its accessibility and high throughput, SpARC-map can be a tool of first resort when investigating any biological process where PPIs are vital, and facilitate discovery across molecular and cellular biology.

Materials and Methods

Structure and design of SpARC vectors. The SpARC vectors pSEP4N and pSEP4C harbor low-copy pBR322 origin of replication and ampicillin-resistance gene. Transcriptional suppressors lacI and tet^R (separated by an internal ribosome binding site) are expressed under the *lacIq* promoter. Expression of neo^RN-bait fusion (in pSEP4N) or bait-neo^RN fusion (in pSEP4C) is controlled by IPTG-inducible *trc* promoter/*lac* operator. Expression of the prey peptide-neo^RC fusion is controlled by the aTc-inducible *PletO-1* promoter.

Library construction. Bait CDS is cloned into SpARC vector using the multiple cloning site: p21, PCNA, GART, PFAS, PAICS, ADSL, ATIC were cloned as bait into pSEP4N; PPAT was cloned as bait into pSEP4C. Bait-containing SpARC vector is cut at NotI and SbfI sites flanking the peptide library insertion site, blunted, and dephosphorylated. dsDNA containing prey CDS (without stop codon) is digested using DNA Fragmentase (NEB M0348) into 100-300 bp fragments; for multiplexed experiments, prey CDSs are combined in an equimolar mix and fragmented. Prey CDS fragments are purified, blunted, and ligated into the SpARC vector.

Library pre-selection. The ligated SpARC library is transformed into homemade chemically competent *E. coli* (strain NEBStable) (55) and pre-selected at low levels of kanamycin. For p21 (bait)-PCNA (prey) and PCNA (bait)-p21 (prey) libraries, pre-selection is on LB agar with 100 ng/mL aTc, 100 µg/mL ampicillin, 4 µg/mL kanamycin, with no IPTG, relying on leaky expression from the IPTG-inducible promoter for bait expression. For DNPB enzyme (bait)-DNPB enzymes (prey) libraries, pre-selection is on LB agar with 1 mM IPTG, 100 ng/mL aTc, 100 µg/mL ampicillin, and 5 µg/mL (GART, PAICS, ADSL, and ATIC), 3 µg/mL, (PPAT), or 2 µg/mL (PFAS) kanamycin. Following outgrowth, colonies are collected, pooled, and the hosted SpARC vectors are extracted.

Library final selection. For each pre-selected SpARC library, we measure its kill curve on a series of LB agar plates containing IPTG (no ITPG for p21 and PCNA libraries, 1 mM for DNPB libraries), aTc (100 ng/mL), ampicillin (100 µg/mL), and varying concentrations of kanamycin (5-640 ug/ml). We also measure the kill curve of the bait-only parental vector under identical conditions. We choose as final selection conditions under which the bait-only SpARC vector cannot survive, but where the SpARC library still yields surviving colonies. We then conduct a large-scale transformation of the (pre-selected) SpARC library, with 5-50× redundancy (2.5-5 million transformants), and outgrow on LB agar under final selection conditions.

NGS sequencing. Post final selection, kanamycin-resistant colonies are pooled, collected, and the host SpARC vectors are extracted. The prey peptide coding sequence is amplified using barcoded NGS sequencing primers that anneal on invariant flanking sequences. PCR amplification is for 12-cycles, and the purified PCR product is sequenced using the AmpliconEZ service from Azenta Life Sciences based on Illumina's MiSeq2 platform and yields >50k (guaranteed minimum, >200k typical) 250 bp paired end reads per sample.

Prey peptide identification and mapping. Paired-end NGS reads are trimmed, quality-filtered, and merged; gapped sequences are discarded. For each merged NGS read, we identify the longest open reading frame (ORF) that is fused in-frame with $neo^{R}C$, accepting ATG, TTG, GTG as start codons (56), and map it to prey CDS(s); ORFs that are less than 15 bps in length, or that are concatemers (either of CDS fragments from different preys, or of noncontiguous fragments of the same prey) are discarded. All NGS processing and mapping are done using a custom Julia workflow built using BioSequences.jl and FASTX.jl packages.

Estimating the nonspecific background. Peptide coding sequences that are inverted, and/or frameshifted with respect to the prey CDS to which they are mapped to, are designated as encoding random peptides. We identify all such random species i and quantify their abundances $\{n_i\}$. The results are fitted to a generalized Pareto distribution (GPD) using maximum likelihood estimation using the Extremes.jl package.

In vivo protein photocrosslinking. HEK293T cells cultured in purine-depleted media were co-transfected with a plasmid expressing 2×StrepTag -$ADSL^{Y466TAG}$, and pMAH-POLY, which expresses the unnatural amino acid incorporation system. Transfected cells were supplemented with 4-azido-L-phenylalanine (1.8 mM) for 48 hours. For photocrosslinking, cells are irradiated using a germicidal UV lamp (260 nm peak emission, 3 minutes) and immediately harvested; non-irradiated cells are collected directly. Cleared cell lysate was incubated with anti-StrepTag magnetic beads (IBA

Lifesciences 2-5090-010) and extensively washed. Pulled-down proteins are alklyated and trypsin-digested on-bead. Digested peptides are purified and analyzed via liquid chromatography electrospray ionization tandem mass spectrometry on a Thermo Orbitrap Eclipse. Experiment was repeated with two biological replicates per treatment.

Mass spectrometry analysis. Peptide are identified and quantified (label-free quantification) using Proteome Discoverer 1.4. To identify differentially enriched protein species, we first estimate the inter-replicate distribution for the abundance of protein species *i,* $P_{\text{rep}}(n_i | n_{i,\ \text{avg}})$, given its inter-replicate mean $n_{i,\ \text{avg}}$. We assume $P_{\text{rep}}(n_i | n_{i,\ \text{avg}})$ depends only on $n_{i,\ \text{avg}}$, and is similar for proteins of similar abundance, i.e. if $n_{i,\ \text{avg}} = n_{j,\ \text{avg}} = n_{\text{avg}}$, then $P_{\text{rep}}(n_i | n_{i,\ \text{avg}}) = P_{\text{rep}}(n_j | n_{j,\ \text{avg}}) = P_{\text{rep}}(n | n_{\text{avg}})$. We estimate $P_{\text{rep}}(n | n_{\text{avg}})$ by binning together proteins with similar abundance and using Gaussian kernel density estimation, as implemented by the KernelDensity.jl package, to approximate the result. To evaluate whether the abundance of protein species *i* under UV-irradition is significantly enriched, given its abundance in non-irradiated samples is $n_{i,\ \text{avg, UV-}}$, we use the estimated $P_{\text{rep}}(n | n_{\text{avg}} = n_{i,\ \text{avg, UV-}})$ to compute its *p*-value; a small *p* indicates the it is enriched in UV-irradiated samples.

Acknowledgments

ADSL knock-out HeLa was a kind gift from Marie Zikanova, Charles University and General University Hospital, Prague, Czech Republic. We acknowledge the Proteomics and Mass Spectroscopy Core Facility, Huck Institutes for the Life Sciences, Penn State, for the use of their instrumentation. We thank Tony Pedley, Phil Hanoian, Michelle Spiering, Vivek Kapur, Steve Baker, and Han Nguyen for helpful discussions and comments. This work was supported by the U.S. National Institutes of Health under award R01GM024129.

References

1. T. Pawson, P. Nash, Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science* **300**, 445–452 (2003).
2. S. Wang, A. O. Osgood, A. Chatterjee, Uncovering post-translational modification-associated protein–protein interactions. *Current Opinion in Structural Biology* **74**, 102352 (2022).
3. D. Ryan, J. Matthews, Protein–protein interactions in human disease. *Current Opinion in Structural Biology* **15**, 441–446 (2005).
4. M. Arkin, Protein–protein interactions and cancer: small molecules going in for the kill. *Current Opinion in Chemical Biology* **9**, 317–324 (2005).
5. M. R. Arkin, Y. Tang, J. A. Wells, Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality. *Chemistry & Biology* **21**, 1102–1114 (2014).
6. H. Lu, *et al.*, Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Sig Transduct Target Ther* **5**, 213 (2020).
7. M. Baek, *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
8. R. Krishna, *et al.*, Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).
9. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

10. J. Abramson, *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
11. A. J. Percy, M. Rey, K. M. Burns, D. C. Schriemer, Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry—A review. *Analytica Chimica Acta* **721**, 7–21 (2012).
12. A. Luchini, V. Espina, L. A. Liotta, Protein painting reveals solvent-excluded drug targets hidden within native protein–protein interfaces. *Nat Commun* **5**, 4413 (2014).
13. A. M. Pedley, V. Pareek, S. J. Benkovic, The Purinosome: A Case Study for a Mammalian Metabolon. *Annu. Rev. Biochem.* **91**, 89–106 (2022).
14. Y. Zhang, M. Morar, S. E. Ealick, Structural biology of the purine biosynthetic pathway. *Cell. Mol. Life Sci.* **65**, 3699–3724 (2008).
15. V. Pareek, H. Tian, N. Winograd, S. J. Benkovic, Metabolomics and mass spectrometry imaging reveal channeled de novo purine synthesis in cells. *Science* **368**, 283–290 (2020).
16. L. J. Ball, R. Kühne, J. Schneider-Mergener, H. Oschkinat, Recognition of Proline-Rich Motifs by Protein–Protein-Interaction Domains. *Angew Chem Int Ed* **44**, 2852–2869 (2005).
17. I. S. Moreira, P. A. Fernandes, M. J. Ramos, Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics* **68**, 803–812 (2007).
18. G. Schreiber, A. E. Keating, Protein binding specificity versus promiscuity. *Current Opinion in Structural Biology* **21**, 50–61 (2011).
19. K. Van Roey, *et al.*, Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **114**, 6733–6778 (2014).
20. Y. Ivarsson, P. Jemth, Affinity and specificity of motif-based protein–protein interactions. *Current Opinion in Structural Biology* **54**, 26–33 (2019).
21. P. E. Wright, H. J. Dyson, R. A. Lerner, Conformation of peptide fragments of proteins in aqueous solution: implications for initiation of protein folding. *Biochemistry* **27**, 7167–7175 (1988).
22. D. D. Boehr, R. Nussinov, P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **5**, 789–796 (2009).
23. P. Vanhee, *et al.*, Protein-Peptide Interactions Adopt the Same Structural Motifs as Monomeric Protein Folds. *Structure* **17**, 1128–1136 (2009).
24. D. E. Paschon, Z. S. Patel, M. Ostermeier, Enhanced Catalytic Efficiency of Aminoglycoside Phosphotransferase (3')-IIa Achieved Through Protein Fragmentation and Reassembly. *Journal of Molecular Biology* **353**, 26–37 (2005).
25. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer London, 2001).
26. J. Jalbert, M. Farmer, G. Gobeil, P. Roy, Extremes.jl: Extreme Value Analysis in Julia. *Journal of Statistical Software* **109**, 1–35 (2024).
27. S. Waga, G. J. Hannon, D. Beach, B. Stillman, The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA. *Nature* **369**, 574–578 (1994).
28. C. Cayrol, M. Knibiehler, B. Ducommun, p21 binding to PCNA causes G1 and G2 cell cycle arrest in p53-de®cient cells. *Oncogene* **16**, 311–320 (1998).
29. C. Prives, V. Gottifredi, The p21 and PCNA partnership: A new twist for an old plot. *Cell Cycle* **7**, 3840–3846 (2008).
30. J. Chen, *et al.*, A 39 Amino Acid Fragment of the Cell Cycle Regulator p21 Is Sufficient to Bind PCNA and Partially Inhibit DNA Replication in vivo. *Nucleic Acids Research* **24**, 1727–1733 (1996).

31. E. Warbrick, D. P. Lane, D. M. Glover, L. S. Cox, A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen. *Current Biology* **5**, 275–282 (1995).
32. J. M. Gulbis, Z. Kelman, J. Hurwitz, M. O'Donnell, J. Kuriyan, Structure of the C-Terminal Region of p21WAF1/CIP1 Complexed with Human PCNA. *Cell* **87**, 297–306 (1996).
33. A. Prestel, *et al.*, The PCNA interaction motifs revisited: thinking outside the PIP-box. *Cell. Mol. Life Sci.* **76**, 4923–4943 (2019).
34. S. An, R. Kumar, E. D. Sheets, S. J. Benkovic, Reversible Compartmentalization of de Novo Purine Biosynthetic Complexes in Living Cells. *Science* **320**, 103–106 (2008).
35. J. B. French, *et al.*, Spatial colocalization and functional link of purinosomes with mitochondria. *Science* **351**, 733–737 (2016).
36. C. Wan, *et al.*, Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).
37. C. Doigneaux, *et al.*, Hypoxia drives the assembly of the multienzyme purinosome complex. *Journal of Biological Chemistry* **295**, 9551–9566 (2020).
38. J. He, L.-N. Zou, V. Pareek, S. J. Benkovic, Multienzyme interactions of the de novo purine biosynthetic protein PAICS facilitate purinosome formation and metabolic channeling. *Journal of Biological Chemistry* **298**, 101853 (2022).
39. Z. Sha, S. J. Benkovic, Purinosomes spatially co-localize with mitochondrial transporters. *Journal of Biological Chemistry* **300** (2024).
40. F. J. Schendel, Y. S. Cheng, J. D. Otvos, S. Wehrli, J. Stubbe, Characterization and chemical properties of phosphoribosylamine, an unstable intermediate in the de novo purine biosynthetic pathway. *Biochemistry* **27**, 2614–2623 (1988).
41. J. Rudolph, J. Stubbe, Investigation of the Mechanism of Phosphoribosylamine Transfer from Glutamine Phosphoribosylpyrophosphate Amidotransferase to Glycinamide Ribonucleotide Synthetase. *Biochemistry* **34**, 2241–2250 (1995).
42. J. W. Chin, *et al.*, Addition of *p* -Azido- L -phenylalanine to the Genetic Code of *Escherichia c oli*. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
43. A. Chatterjee, H. Xiao, M. Bollong, H.-W. Ai, P. G. Schultz, Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11803–11808 (2013).
44. Z. Chen, W. Ren, Q. E. Wright, H. Ai, Genetically Encoded Fluorescent Probe for the Selective Detection of Peroxynitrite. *J. Am. Chem. Soc.* **135**, 14940–14943 (2013).
45. P. V. Hornbeck, *et al.*, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research* **43**, D512–D520 (2015).
46. S. P. Ray, *et al.*, Structural and Biochemical Characterization of Human Adenylosuccinate Lyase (ADSL) and the R303C ADSL Deficiency-Associated Mutation. *Biochemistry* **51**, 6701–6713 (2012).
47. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
48. C. J. Oldfield, A. K. Dunker, Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
49. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
50. E. W. Martin, T. Mittag, Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **57**, 2478–2487 (2018).

51. C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, D. Marenduzzo, Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. U.S.A.* **110** (2013).
52. J. Chen, S. R. Kline, Y. Liu, From the depletion attraction to the bridging attraction: The effect of solvent molecules on the effective colloidal interactions. *The Journal of Chemical Physics* **142**, 084904 (2015).
53. G. Yuan, J. Luo, C. C. Han, Y. Liu, Gelation transitions of colloidal systems with bridging attractions. *Phys. Rev. E* **94**, 040601 (2016).
54. R. Oughtred, *et al.*, The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science* **30**, 187–200 (2021).
55. Y. Yang, *et al.*, An Optimized Transformation Protocol for Escherichia coli BW3KD with Supreme DNA Assembly Efficiency. *Microbiology Spectrum* **10**, e02497-22 (2022).
56. A. Hecht, *et al.*, Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Research* **45**, 3615–3626 (2017).

Figures and Tables

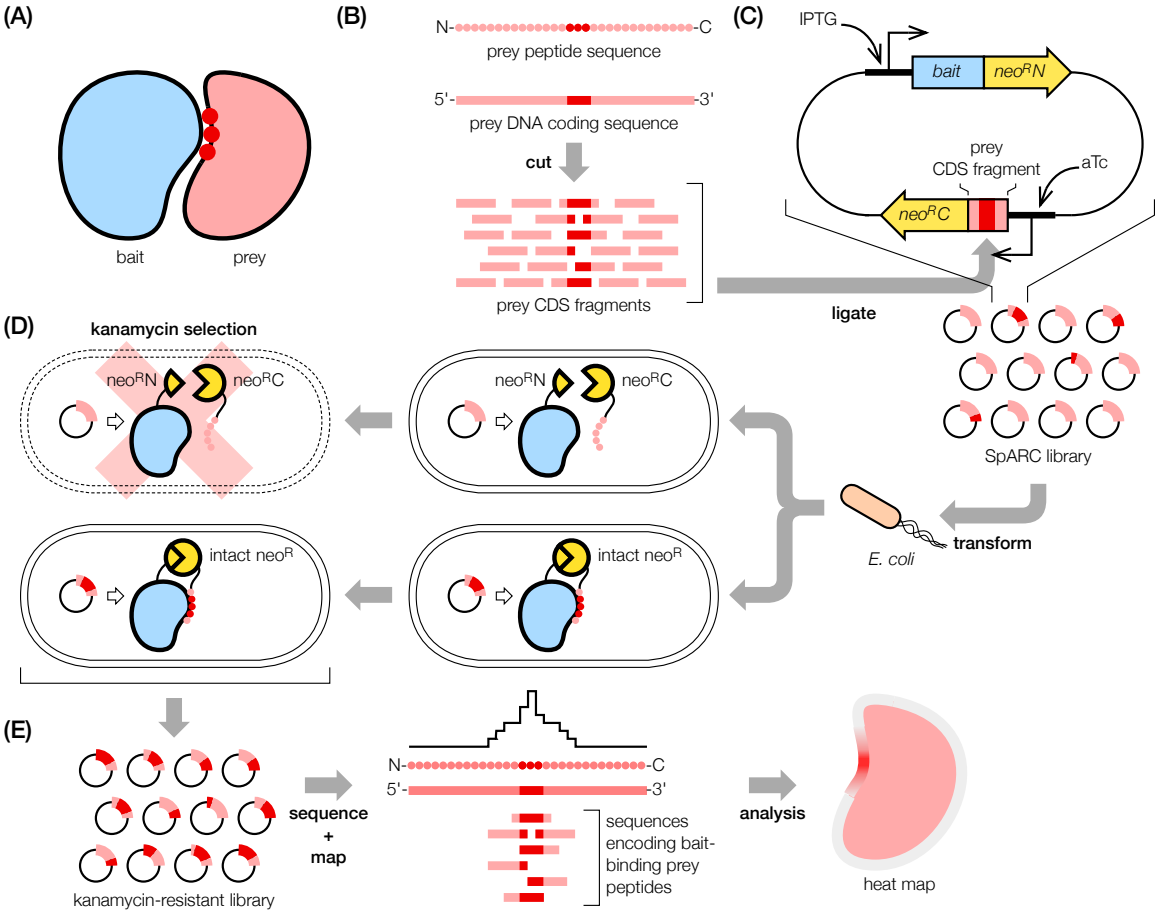

Figure 1. Schematic of SpARC-map. **(A)** Pair of interacting bait and prey proteins; prey residues on the bait-prey PPI interface are highlighted. **(B)** dsDNA containing the prey CDS is sheared into random fragments and **(C)** ligated into the bicistronic SpARC vector, in upstream fusion with the *neo^{R}C*. The intact bait protein is fused to *neo^{R}N*; the fusion terminus depends on the bait. Expression of the bait and prey peptide are under the control of ITPG and aTc inducible promoters. **(D)** The SpARC library is transformed into *E. coli* host and selected in the presence of kanamycin. SpARC vectors that express bait-binding prey peptides can result in kanamycin resistance through the reconstitution of neo^{R}N and neo^{R}C fragments into a functional neo^R. **(E)** SpARC vectors from kanamycin-resistant colonies are sequenced, mapped, and analyzed to extract the PPI interface.

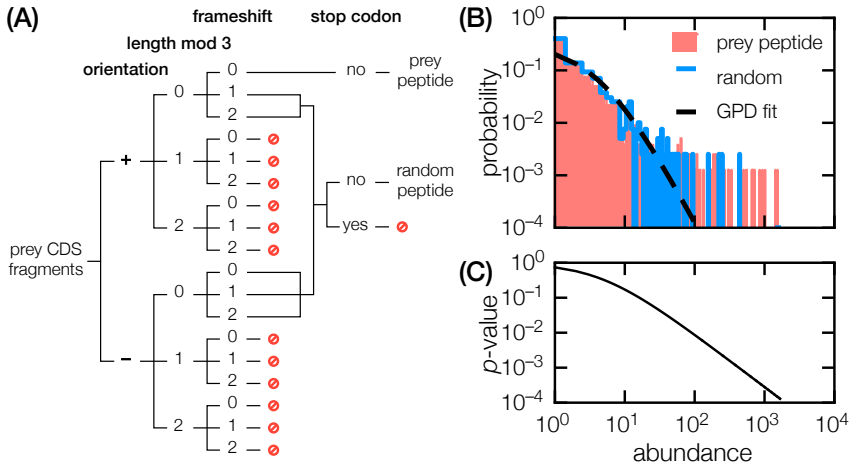

Figure 2. Analysis of NGS species abundance to detect non-specific interactions. **(A)** In addition to in-sense prey CDS fragments (orientation = +, length mod 3 = 0, frameshift = 0), some nonsense fragments, when ligated into the SpARC vector upstream of *neo^{R}C*, can translate into a random peptide fused in frame with neo^{R}C fragment. **(B)** From a representative SpARC-map dataset: measured abundance of NGS species encoding random peptides (blue line) and the resulting GPD fit (black dashed line), compared against NGS species that encode true prey peptides (red shaded). **(C)** From the GPD fitted to the random peptide abundance, we can compute a *p*-value for any true prey peptide given its abundance; those with small *p* are unlikely to be random binders of the bait.

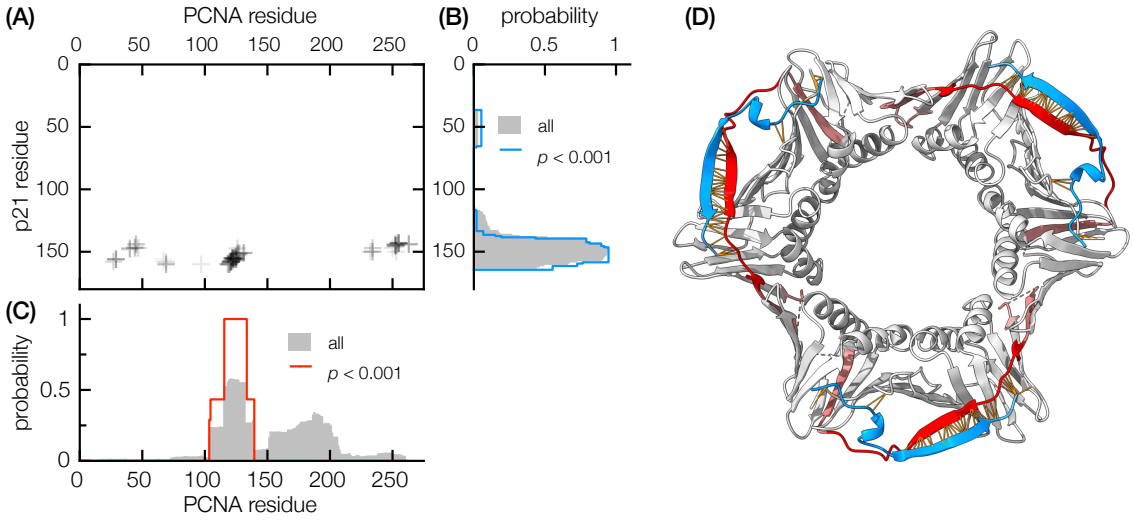

Figure 3. Mapping the p21-PCNA interface. **(A)** Map of atomic contacts between PCNA and p21 residues from the co-crystal structure of PCNA with $p21^{139\text{-}160}$; each (+) represents an atomic contact between indicated residues. **(B)** Heatmap of PCNA binding sites on p21; solid gray is heatmap compiled from all prey peptides, red outline is heatmap compiled using only prey peptides whose abundances significantly exceed the random peptide background. **(C)** Heatmap of p21 binding sites on PCNA. **(D)** Co-crystal structure (PDB ID: 1AXC) of PCNA homo-trimer in complex with $p21^{139\text{-}160}$ peptide (blue). PCNA is colored by its affinity heatmap for p21 (red/white = most/least probable). Orange lines are pseudo-bonds denoting atomic contacts between PCNA and $p21^{139\text{-}160}$.

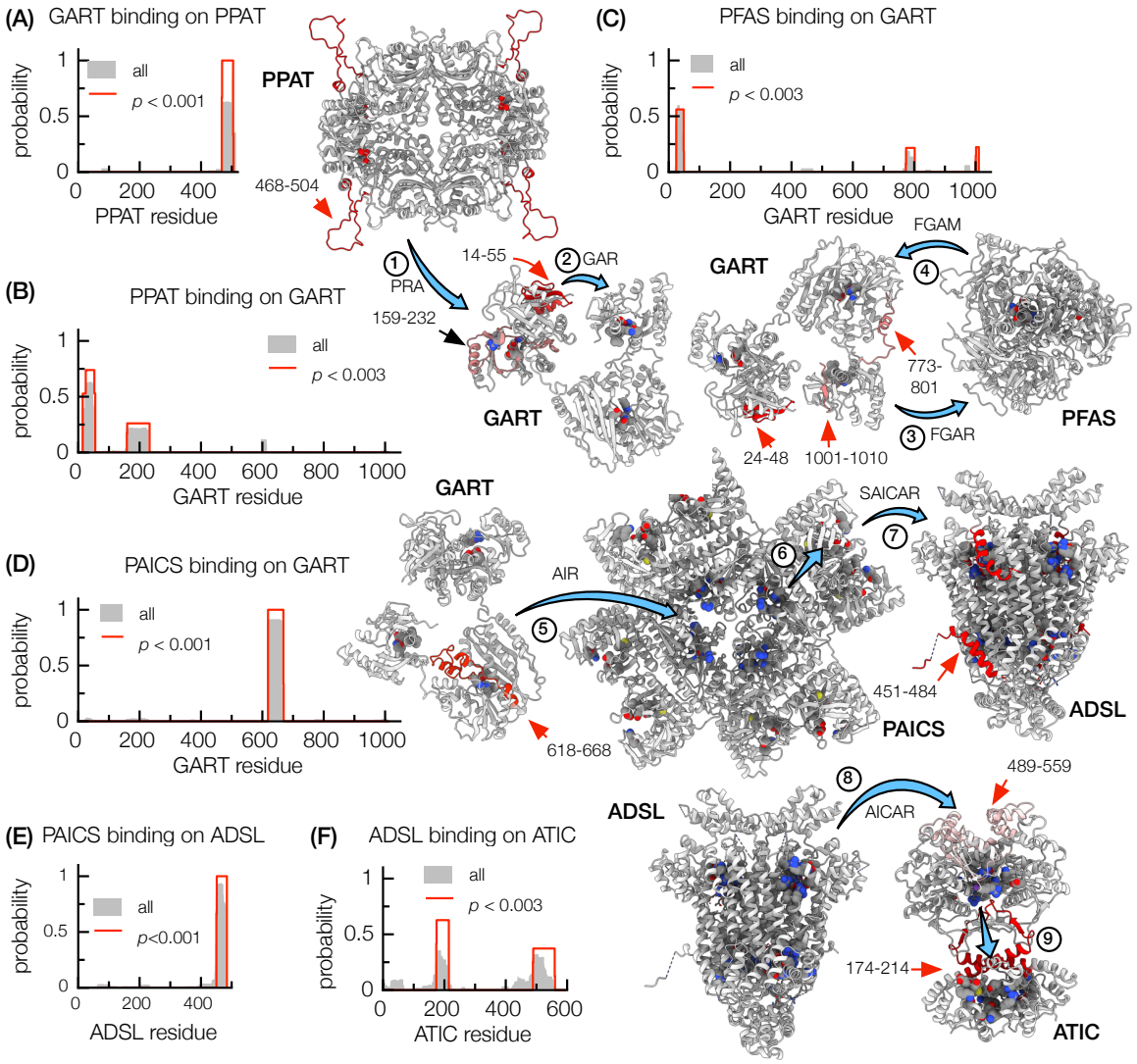

Figure 4. High confidence (p<0.003) PPI interfaces between sequential enzymes in the purinosome. Protein structures are colored by affinity heat map (red/white = most/least probable). Spherical atoms highlight enzyme active sites and ligand binding sites. Blue arrows indicate the direction of substrate transfer from upstream to downstream enzyme (domain). Heatmap plots show probable **(A)** GART (monomer) binding sites on PPAT (homo-tetramer); **(B)** PPAT binding sites on GART; **(C)** PFAS (monomer) binding sites on GART; **(D)** PAICS (homo-octamer) binding sites on GART; **(E)** PAICS binding sites on ADSL (homo-tetramer); **(F)** ADSL binding sites on ATIC (homo-dimer).

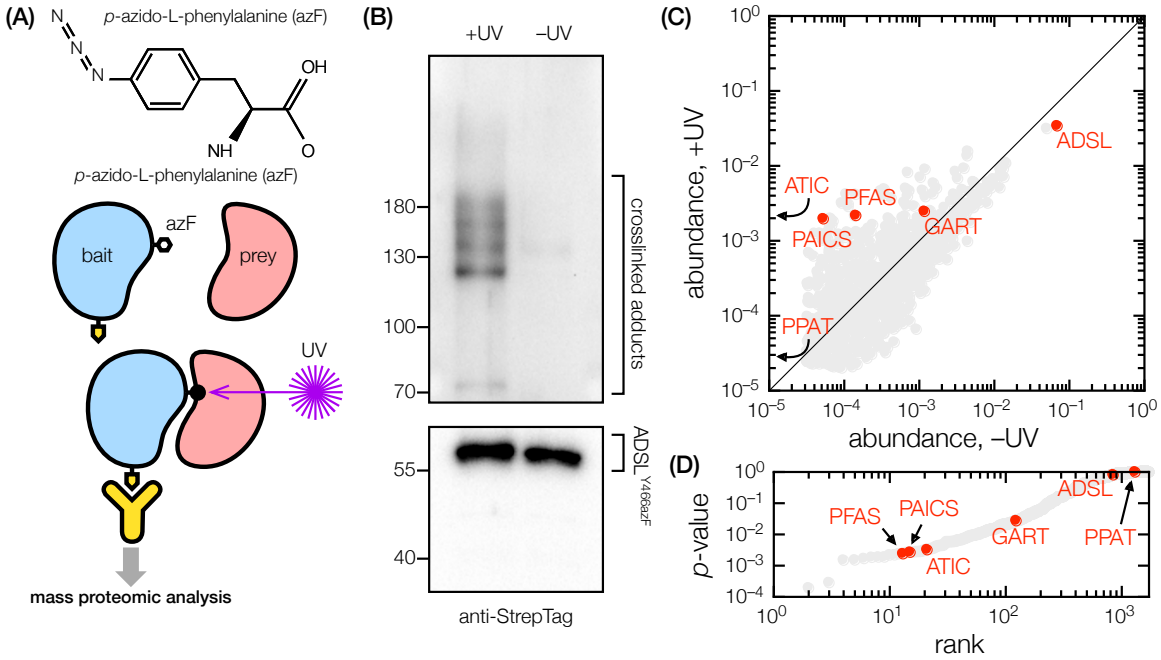

Figure 5. Site-specific photocrosslinking with ADSL^{Y466azF} in live cells. **(A)** Photoactive p-azido-L-phenylalanine (azF) is inserted onto a PPI interface of a bait protein. Upon UV-irradiation, the azF can form a covalent crosslink between the bait and any prey that interact with the bait via the specific interface. The resulting adduct can then be pulled down and analyze. **(B)** Western blot showing anti-StrepTag pull-down from HEK293T transfected with 2×Strep-ADSL^{Y466azF}. High molecular weight crosslinked adducts are apparent following UV-irradiation. **(C)** Mass proteomic analysis of protein abundance in anti-StrepTag pull-downs with and without UV irradiation. The six purinosome enzymes are highlighted in red. **(D)** *p*-values in ascending rank, indicating PFAS, PAICS, and ATIC are significantly enriched in UV-irradiated vs non-irradiated samples.

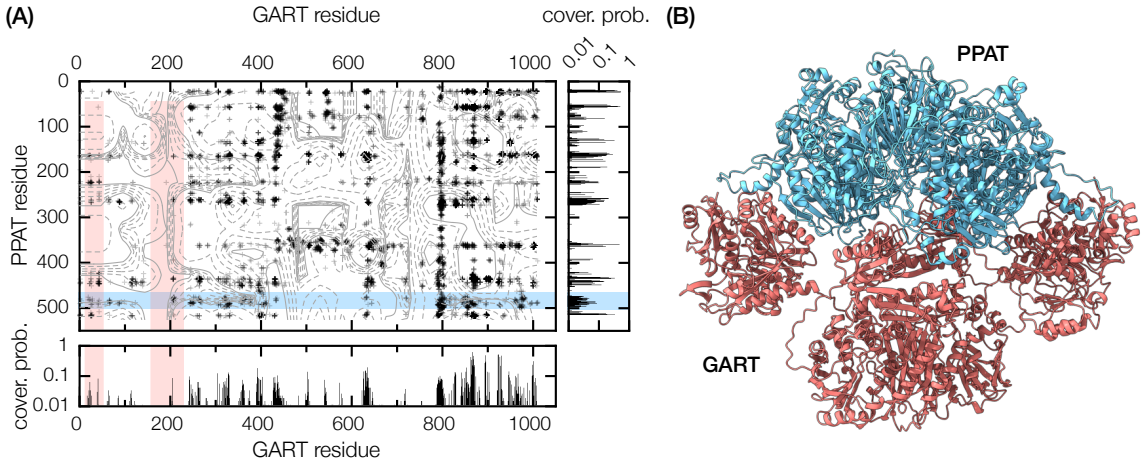

Figure 6. Modeling of the PPAT-GART complex with AlphaFold3. **(A)** Map of atomic contacts between PPAT and GART monomers across 289 AF3 models. Each data point is a PPAT-GART atomic contact found in at least one model; density contours are quadratic interpolations. Also shown are heatmaps indicating the fraction of models that has a specific residue on PPAT (right sub-panel) or GART (bottom sub-panel) to be a part of the PPAT-GART interface. Shaded regions (light blue on PPAT, pink on GART) are regions identified by SpARC-map as probable PPAT-GART interfaces. **(B)** Example of an AF3 model of PPAT-GART complex that is compatible with SpARC-map results; PPAT (homo-tetramer) is show in light blue, GART (homodimer) is in pink.

Supplemental materials for "Multiplex mapping of protein-protein interaction interfaces"

Jingxuan He, Ling-Nan Zou, Vidhi Pareek, Stephen J. Benkovic

Mathematical model of affinity-selection in SpARC-map

Consider a bacterial host in an environment with kanamycin concentration $[\text{kan}]_{\text{out}}$. The kanamycin concentration $[\text{kan}]_{\text{in}}$ inside the host cell depends on the balance between kanamycin transport of kanamycin into the cell, and the degradation of kanamycin by reconstituted neoR inside the cell,

$$\frac{d[\text{kan}]_{\text{in}}}{dt} = \mu\left([\text{kan}]_{\text{out}} - [\text{kan}]_{\text{in}}\right) - k_{\text{cat}}[\text{neo}^{\text{R}} \cdot \text{kan}] \ . \tag{2.1}$$

The first term in (2.1) describes kanamycin influx into the cell, characterized by a rate μ; the second term describes the degradation of kanamycin by (reconstituted) neoR with catalytic rate k_{cat}. We assume the inactivation of kanamycin by neo^{R} follows Michaelis-Menten kinetics, such that

$$[\text{neo}^{\text{R}} \cdot \text{kan}] = \frac{[\text{kan}]_{\text{in}}[\text{neo}^{\text{R}}]}{K_{\text{M}} + [\text{kan}]_{\text{in}}}, \tag{2.2}$$

and that the reconstitution of intact neo^{R} is entirely due to the 1:1 binding (with dissociation constant K_{D}) of the bait protein and prey peptide to which the neo^{R}N and neo^{R}C fragments are fused,

$$[\text{neo}^{\text{R}}] = [\text{bait} \cdot \text{prey}] = \frac{[\text{bait}][\text{prey}]}{K_{\text{D}}}. \tag{2.3}$$

In steady state, the kanamycin concentration inside the cell is given by

$$[\text{kan}]_{\text{in}} = \frac{[\text{kan}]_{\text{out}}}{1 + \frac{k_{\text{cat}}}{\mu K_{\text{M}}}\frac{[\text{bait}][\text{prey}]}{K_{\text{D}}}} \ , \tag{2.4}$$

where we have assume $[\text{kan}]_{\text{in}} < K_{\text{M}}$ first order kinetics apply. The host cell can survive only if $[\text{kan}]_{\text{in}}$ is below some critical value, assumed to be to the minimum inhibitory concentration of kanamycin for *E. coli*, $[\text{kan}]_{\text{MIC}}$ = 3.4 μM. In terms of the K_{D} between the bait protein and the prey peptide, the condition for host survival is

$$K_{\text{D}} \leq K_{\text{D}}^{*}, \quad K_{\text{D}}^{*} = \frac{k_{\text{cat}}}{\mu K_{\text{M}}}\left(\frac{[\text{bait}][\text{prey}]}{[\text{kan}]_{\text{ext}}/[\text{kan}]_{\text{MIC}} - 1}\right). \tag{2.5}$$

The selection threshold K_{D}* depends on several kinetic parameters: for the split neoR system, $k_{\text{cat}}/K_{\text{M}} \approx 0.1$ μM^{-1} s^{-1}; for kanamycin transport into E coli, $\mu \approx 0.002$ s^{-1} (1). Taking $[\text{kan}]_{\text{ext}}/[\text{kan}]_{\text{MIC}}$ = 5 – 100, and assuming [bait] ~ [prey] = 0.1 – 10 μM (equivalent to ~ 0.01 – 1 mg of a 30 kDa recombinant protein expressed per liter of culture), yields a tunable range for K_{D}* between 5 nM and 1 mM.

In practice, even for cells that host the identical bait/prey pair, cell-to-cell variation in bait and prey expression (and therefore neo^{R} reconstitution), kanamycin uptake, and kanamycin sensitivity means K_{D}* is not a single hard threshold, but will vary from cell to cell. The probability of survival $p_{\text{surv}}(K_{\text{D}}|K_{\text{D, avg}}^{*})$, where $K_{\text{D, avg}}$* is the average selection threshold across all cells, will be a sigmoidal function with limits $p_{\text{surv}} \to 1$ when $K_{\text{D}} \ll K_{\text{D, avg}}^{*}$,

$p_{surv} \to 0$ when $K_D \gg K_{D,\,avg}^*$ (Fig. S1). For cells transformed with a SpARC library containing many bait-binding prey peptides of varying K_D, characterized by some distribution $P_{lib}(K_D)$, the NGS species abundance extracted from the post-selection cell population will be directly proportional to $p_{surv}(K_D \,|\, K_{D,\,avg}^*)$. If selection is weak (large $K_{D,\,avg}^*$), it may be difficult to distinguish strong versus moderate binding peptides because their p_{surv} will be nearly equivalent. Conversely, strong selection (small $K_{D,\,avg}^*$) can drive the library to extinction. Optimal selection is when the decline of $p_{surv}(K_D \,|\, K_{D,\,avg}^*)$ coincides with the left wing of $P_{lib}(K_D)$. Note that the fraction of transformed cells that survives a given selection condition (characterized by $K_{D,\,avg}^*$) is simply

$$f_{\mathrm{surv}}(K^*_{\mathrm{D,\ avg}}) = \int_0^\infty P_{\mathrm{lib}}(K_{\mathrm{D}})\; p_{\mathrm{surv}}(K_{\mathrm{D}}|K^*_{\mathrm{D,\ avg}})\; dK_{\mathrm{D}}\ . \tag{2.6}$$

Generically, $f_{surv}(K_{D,\,avg}^*)$ will be a sigmoid curve, with limits $f_{surv} \to 1$ when $K_{D,\,avg}^* \to 0$ (no selection) and $f_{surv} \to 0$ when $K_{D,\,avg}^* \to \infty$ (maximum selection).

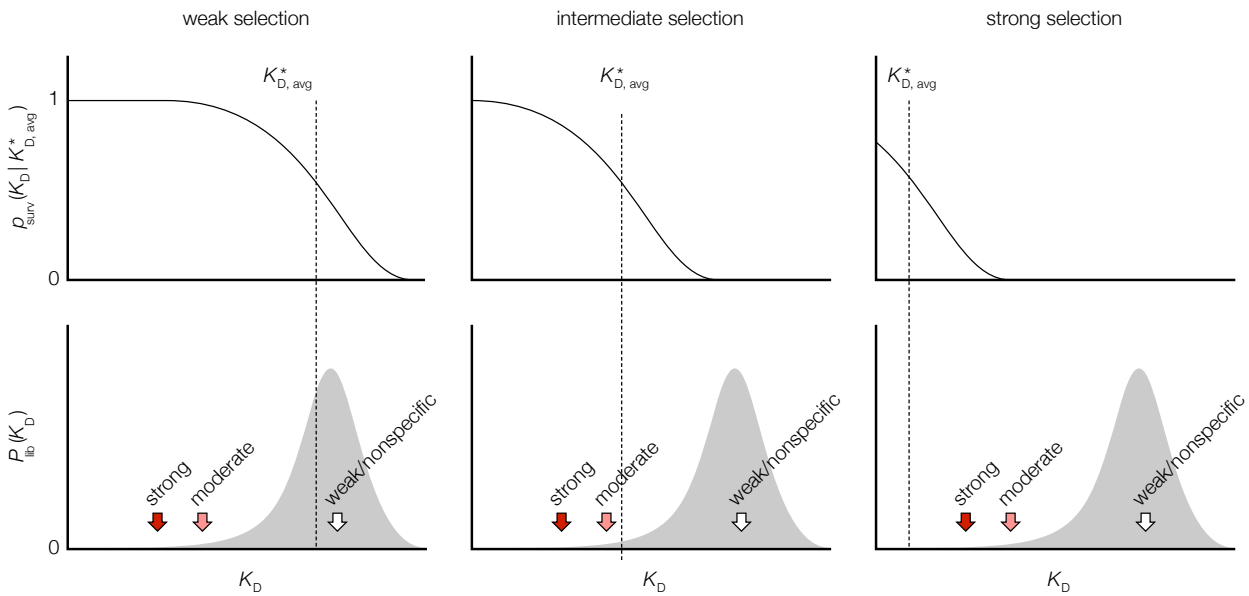

Figure S1. Sketch of affinity selection in SpARC-map. Changing the selection conditions (kanamycin and inducer concentrations) shifts the survival probability $p_{surv}(K_D \,|\, K_{D,\,avg}^*)$ (top row) to sample different portions of the library affinity distribution $P_{lib}(K_D)$ (bottom row, showing locations of strong, moderate, and weak/nonspecific binders).

Defective rescue phenotype in C-terminally tagged ADSL

ADSL deficient HeLa (crADSL) requires adenine supplementation in culture media for survival and proliferation (2). We assessed the ability of CMV-driven expression of exogenous ADSL in crADSL to rescue the deficient phenotype by transfecting crADSL with plasmids expressing 2×Strep-ADSL or ADSL-2×Strep fusion. Transfected crADSL was transferred to purine-depleted media and monitored for survival and proliferation. We found crADSL transfected with 2×StrepTag-ADSL had approximately twice as many surviving cells after 2 weeks as crADSL transfected with ADSL-2×StrepTag (Fig. S2A). This suggests ADSL with C-

terminal 2×StrepTag fusion is defective. After extended culture, both rescue experiments yielded colonies of cells (with many fewer colonies in the ADSL-2×StrepTag) that had re-integrated the exogenous ADSL into the genome; these stably rescued cells can now survive and proliferate indefinitely in purine-depleted media. This indicates both 2×StrepTag-ADSL and ADSL-2×StrepTag retained ADSL's essential enzymatic activity. However, when examined for protein expression, the crADSL::ADSL-2×StrepTag stable rescue expressed (re-integrated) ADSL at ~35× that of wildtype HeLa, while ADSL expression in crADSL:: 2×StrepTag-ADSL is comparable to that of wildtype HeLa (Fig. S2B). Since the C-terminus of ADSL does not participate in known active or ligand binding sites, we interpret the overexpression of ADSL in crADSL::ADSL-2×StrepTag as a compensation for disturbed PPIs in ADSL-2×StrepTag.

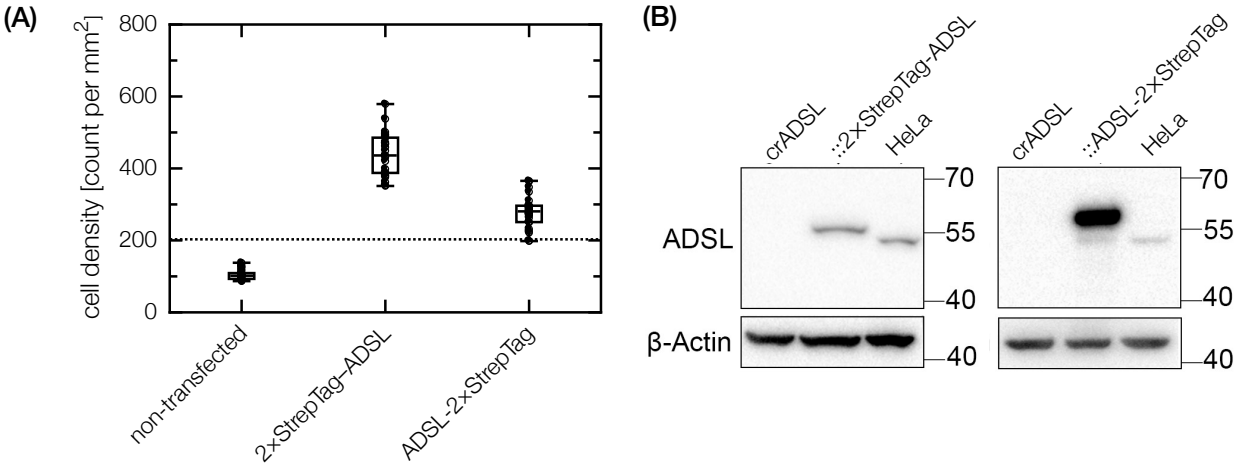

Figure S2. C-terminal tagging of ADSL compromises its ability to rescue ADSL-deficient crADSL cells. **(A)** Proliferation of crADSL rescued by transfection with 2×StrepTag-ADSL or ADSL-2×StrepTag. Dotted line indicates initial crADSL cell seeding density at the start of the rescue assay. Rescue with ADSL-2×StrepTag leads to only half as many surviving cells as rescue with ADSL-2×StrepTag. **(B)** Western blot analysis of reintegrated ADSL expression in crADSL stably rescued with either N-terminally tagged 2×StrepTag-ADSL or C-terminally tagged ADSL-2×StrepTag, showing ADSL hyper-expression is crADSL::ADSL-2×StrepTag.

Supplementary References

1. R. Nakae, T. Nakae, Diffusion of aminoglycoside antibiotics across the outer membrane of Escherichia coli. *Antimicrob Agents Chemother* **22**, 554–559 (1982).
2. V. Baresova, *et al.*, CRISPR-Cas9 induced mutations along de novo purine synthesis in HeLa cells result in accumulation of individual enzyme substrates and affect purinosome formation. *Molecular Genetics and Metabolism* **119**, 270–277 (2016).

Supplementary Tables

Table S1: Summary of SpARC-map libraries screened in this study.

Library (bait-prey)	Size (as ligated)	Size (pre-selected)	Selection Condition (final)
p21 - PCNA	157k	11k	40, 56, 80 µg/ml kan no ITPG, 100 ng/ml aTc
PCNA - p21	275k	26k	40, 56, 80 µg/ml kan no ITPG, 100 ng/ml aTc
PPAT - DNPB	344k	51k	28, 40, 56 µg/ml kan 1 mM ITPG, 100 ng/ml aTc
GART - DNPB	61k	2k	28, 40, 56 µg/ml kan 1 mM ITPG, 100 ng/ml aTc
PFAS - DNPB	188k	24k	20, 28, 40 µg/ml kan 1 mM ITPG, 100 ng/ml aTc
PAICS - DNPB	120k	10k	40, 80, 160 µg/ml kan 1 mM ITPG, 100 ng/ml aTc
ADSL - DNPB	428k	35k	40, 80, 160 µg/ml kan 1 mM ITPG, 100 ng/ml aTc
ATIC - DNPB	92k	8k	40, 80, 160 µg/ml kan 1 mM ITPG, 100 ng/ml aTc

Table S2. Probable PPI interfaces in the purinosome identified by SpARC-map (p<0.01, **bold** p<0.003)

		Bait					
		PPAT	**GART**	**PFAS**	**PAICS**	**ADSL**	**ATIC**
Prey	**PPAT**	**485-503**	**468-504** **503-517**	359-375	**359-382** 485-504	77-109 **358-385** 456-502	**340-382** **459-500**
	GART	**14-55** **159-232** 598-617	**122-162** **219-256**	**24-48** **773-801** **1001-1010**	24-47 174-232 **618-668**	24-55 163-187 841-875 966-986	25-55 966-985
	PFAS	ND	ND	151-161 **463-511** 1328-1338	463-512 **539-556**	ND	463-499
	PAICS	413-425	ND	1-16	ND	251-270 281-348 413-425	ND
	ADSL	**456-484**	**391-402**	**52-91** 405-429 **448-484**	**451-484**	**105-126** **451-484**	73-130 447-484
	ATIC	ND	59-81 180-214	**163-220** 422-435	**180-215** **431-532**	**174-214** **489-559**	**28-84** 493-538