Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.

Supplementary Information for

Parallelized Identification of On- and Off-Target Protein Interactions

Jiayi Dou, Inna Goreshnik, Cassie Bryan, David Baker, Eva-Maria Strauch

Supplementary Methods

Plasmid Extraction and DNA Preparation

Plasmids from cells of the starting and selected pools were extracted as previously described[22]. Briefly, about $3x10^7$ cells were treated with Zymolase (50 U) in 400 µL Solution buffer 1 (Zymo Research, yeast plasmid miniprep II), incubated at 37°C for 4 h and vortexed every hour. Cells were freeze-thawed once; for quick thawing, tubes were submerged in a 42°C water bath and treated as instructed by the Zymo kit manual, with the exception that lysate was applied to higher-yield columns (QIAgen, plasmid miniprep kit). After washing with the QIAgen PB and PE buffers twice, plasmid DNA was eluted with 30 µL elution buffer (QIAgen). Possible contaminating genomic DNA was removed by digesting the sample with 2 µL ExoI (NEB) and 1 µL Lambda exonuclease (NEB)[22] using the supplied Lambda exonuclease buffer. Following a QIAgen PCR clean-up step producing a 30 µL DNA solution, 15 µL were subjected to PCR for the addition of selection-specific barcodes and flow cell adapters. For that, two PCR steps were performed. The first PCR uses a set of “inner primers” to add the Illumina-specific primer annealing site that enables to use with commercial sequencing kit without the addition extra primers. Additionally, we included a short 12 bp sequence as a second set of barcode to label the selected gene pools. The 12mer sequences were designed to have maximal nucleotide diversity for 4 different sets of primers, which can be helpful when sequencing low sequence diversity libraries, as it increases the apparent complexity to the machine; the latter seemed to be important in context of cluster assignments. Primers were designed to have a lower annealing temperature for the first reaction (51°C) (Table S2). To add the Illumina flow-cell adapters and selection-specific barcodes, a second PCR step with a higher annealing temperature (64°C) was performed using primers outer-F and a set of reverse primers containing various barcodes (Table S2). Due to the significant difference between the two melting

temperatures, a purification step for amplicon of the first PCR was not necessary and 2 µL of the first reaction served directly as template for the second reaction. All primers were PAGE purified. The first PCR step was performed for 14 cycles, whereas the second PCR was performed for 15 cycles. However, the cycles necessary for the first reaction depend on the efficiency of the DNA preparation from the yeast cells and may need more cycles, which can be monitored using qPCR. Resulting DNA fragments were gel purified and amounts were quantified by qPCR as instructed (Illumina qPCR manual).

As genes selected from the pool 1 libraries (PPI pools) were all barcoded, only the gene-specific barcodes needed to be sequenced. We choose a reverse primer site arbitrarily 450 bp downstream of the forward primer site(Fig. S1.B). Given the primer annealing site and the 12 bp long forward barcodes at least 46 positions have to be sequenced in order to cover the 18 bp long gene-specific barcodes (Table S2). For sequencing a 10 pM solution of an equimolar mix of the amplicons of each pool was treated as instructed by the Illumina sequencing manual and sequenced using a 50 cycle kit on a Miseq Desktop Sequencer (Illumina)[22]. Forty-eight forward reactions and 6 indexing reactions were performed.

As the pool 2 library (SMPI libraries) did not contain gene-specific primers, we amplified the whole gene. Plasmid-specific primers at the 5' (upstream of the *NheI* site) and 3' site (including the *XhoI* site) were used as inner primers (Fig. S1, Table S2). For sequencing whole genes (SMPI libraries, from pool 2 selection), 300 forward and 300 reverse reactions were performed and 5 times the amount of DNA was used for the reference pool.

Protein Expression and Purification

As control proteins, we used variants of the translocated intimin receptor (TIR) interacting with lectin domain of intimin. TIR variants were displayed on yeast, whereas the last C-terminal amino acids residues (positions 752-939) of intimin from enteropathogenic *Escherichia coli* were cloned into the pET29b expression vector, together with an optimized biotinylation site on its C-terminal end (sequences can be sent upon request). The domain was expressed from the T7 promoter in BL21 (DE3, pLys) cells using Studier autoinduction[23]. After conventional Ni^{2+}-NTA affinity

chromatography purification (QIAgen), intimin was subjected to size exclusion chromatography using a Sephacryl S100 column (GE). Similarly, all other target proteins were expressed with an C-terminal biotinylation site.

Biotinylation

Intimin containing a C-terminal recognition tag for BirA was biotinylated using a kit from Avidity (Aurora, CO). Aliquots were thawed on ice and added to a final concentration of 40 μM to a mixture containing (per 100 μL) 10 μL biomix A, 10 μL biomix B, 10 μL d-biotin (500 mM stock), 1 μL E. coli biotin ligase (3 mg/mL stock), and the balance buffer of 25 mM Hepes at pH 7.4 with 250 mM potassium glutamate. After incubation at 22°C for 5 h, intimin was separated from the biotin ligase by Ni^{2+}-NTA affinity chromatography and desalted into buffer HBS (20 mM Hepes, 150 mM sodium chloride pH 7.4) using a P10 desalting column.

Supplementary Figures and Tables

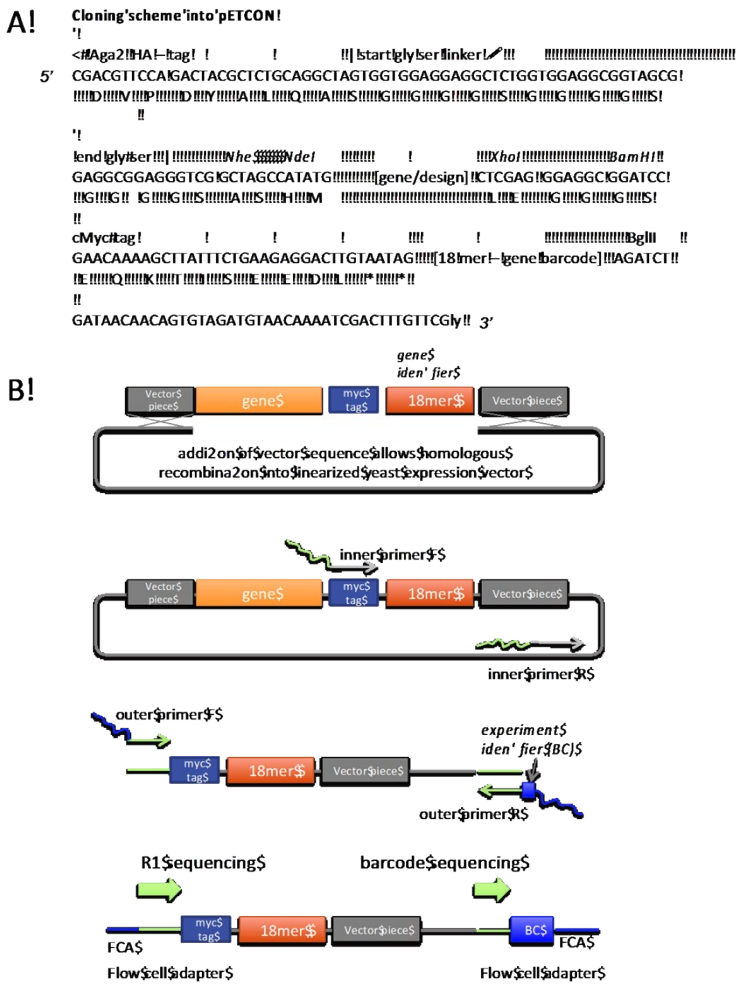

Fig S1.Cloning and Primer Design for Sequencing. (A) Genes for the PPI screen were synthesized with flanking homologous region for *in vivo* cloning into the yeast expression vector. The 5'-end contained sequences encoding parts of the HA-tag and the glycine-serine linker along with the restriction sites, *NheI* and *NdeI*. After the gene of interest, a *XhoI* restriction site is followed by a sequence encoding a short glycine-serine linker, a DNA sequence encoding the cMyc-tag and 2 stop codons so the unique 18mer barcode that is added downstream of each individual coding sequence of pool I is not translated. Lastly, a short sequence homologous to the surface expression vector (pETCON) is included to allow recombination. (B) Overview of cloning and PCR. For next generation sequencing, two nested PCR reactions were performed: the inner PCR is specific to the construct and aligns to the Cmyc-tag and a short sequence ~450 bp 3' of the coding region; this set-up allows the amplification of same-length fragments containing the unique gene specific barcode; the primers for the inner PCR add a short segment for sequencing (supplied in Illumina sequencing kits). The outer reaction adds flow cell adapters and experiment-specific barcodes reporting on the selection target. For sequencing, only a forward and a barcode-specific sequencing reaction are necessary.

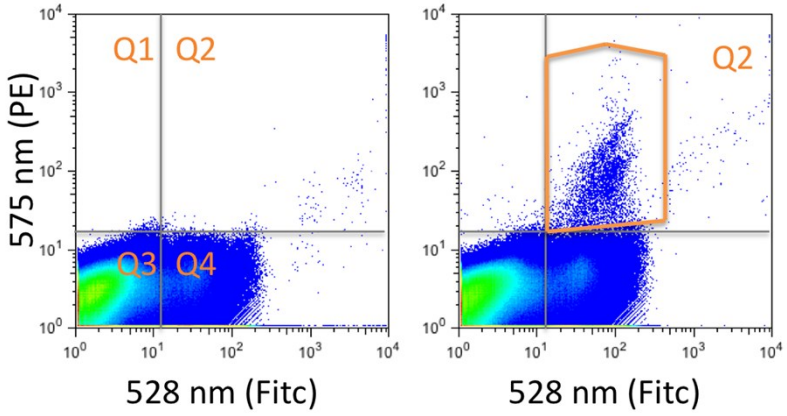

Figure S2. Gates for Identification of Weak Binders. Gating for sorts to identify weak binding proteins to any of the target proteins. Cells only labeled with anti-myc FITC-conjugated antibody were used as a reference to split the 2-dimensional histogram describing expression and binding levels into quarters. Cells from quarter 2 (Q2) were selected as potential binders.

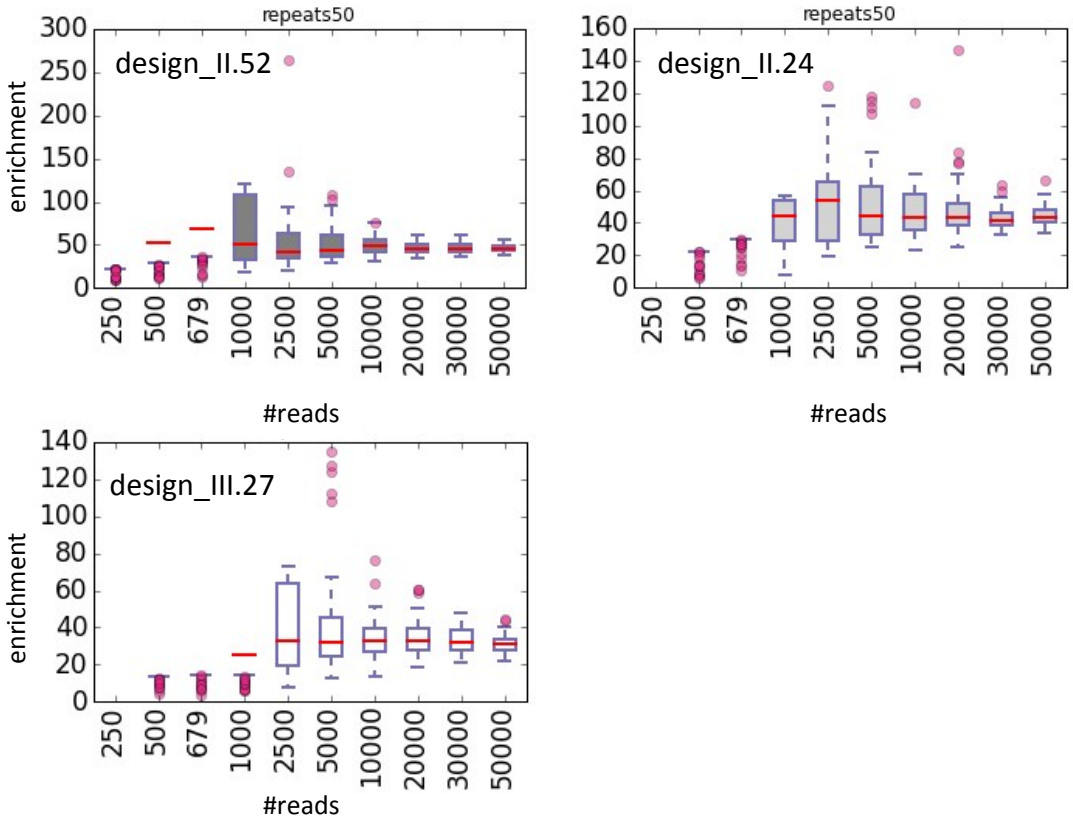

Fig. S3. Simulating smaller sized sequencing reads to estimate minimum read counts for 3 different genes. Using a custom python program, indicated numbers of sequencing reads were pulled out of the raw sequencing data (without replacement).

Genes were counted and enrichment values determined. Calculations were repeated for 50 times for each number of reads.

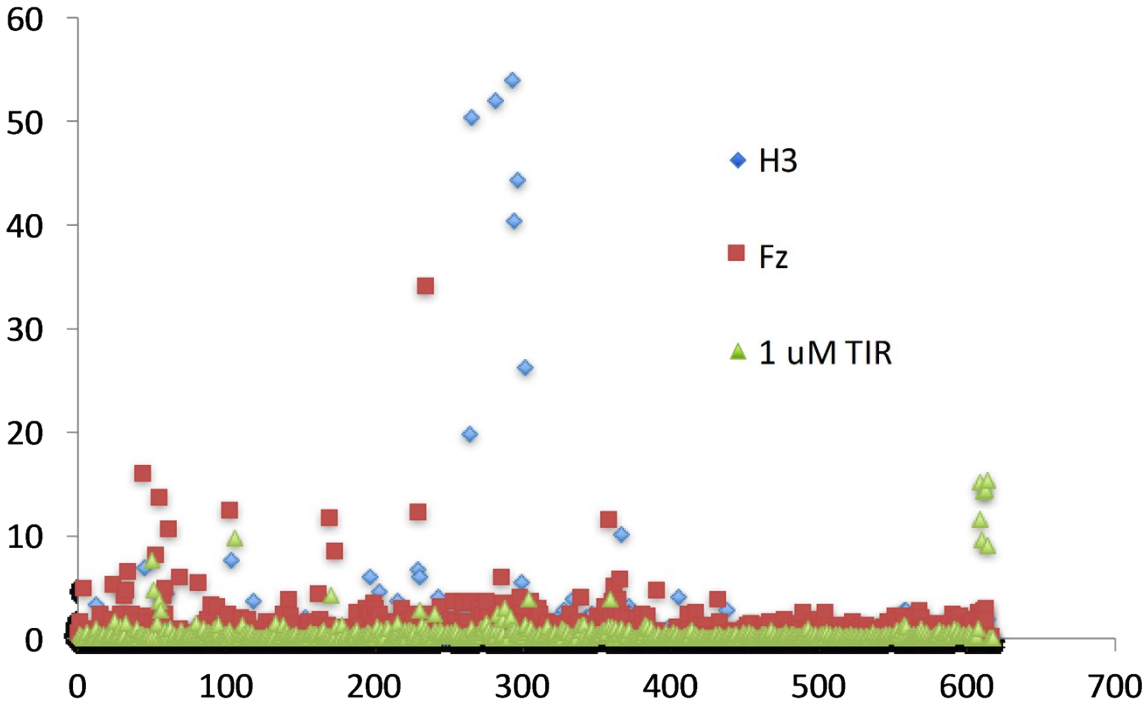

Fig. S4. Enrichment versus sequencing count. (X-axis = sequencing count, Y-axis enrichment)

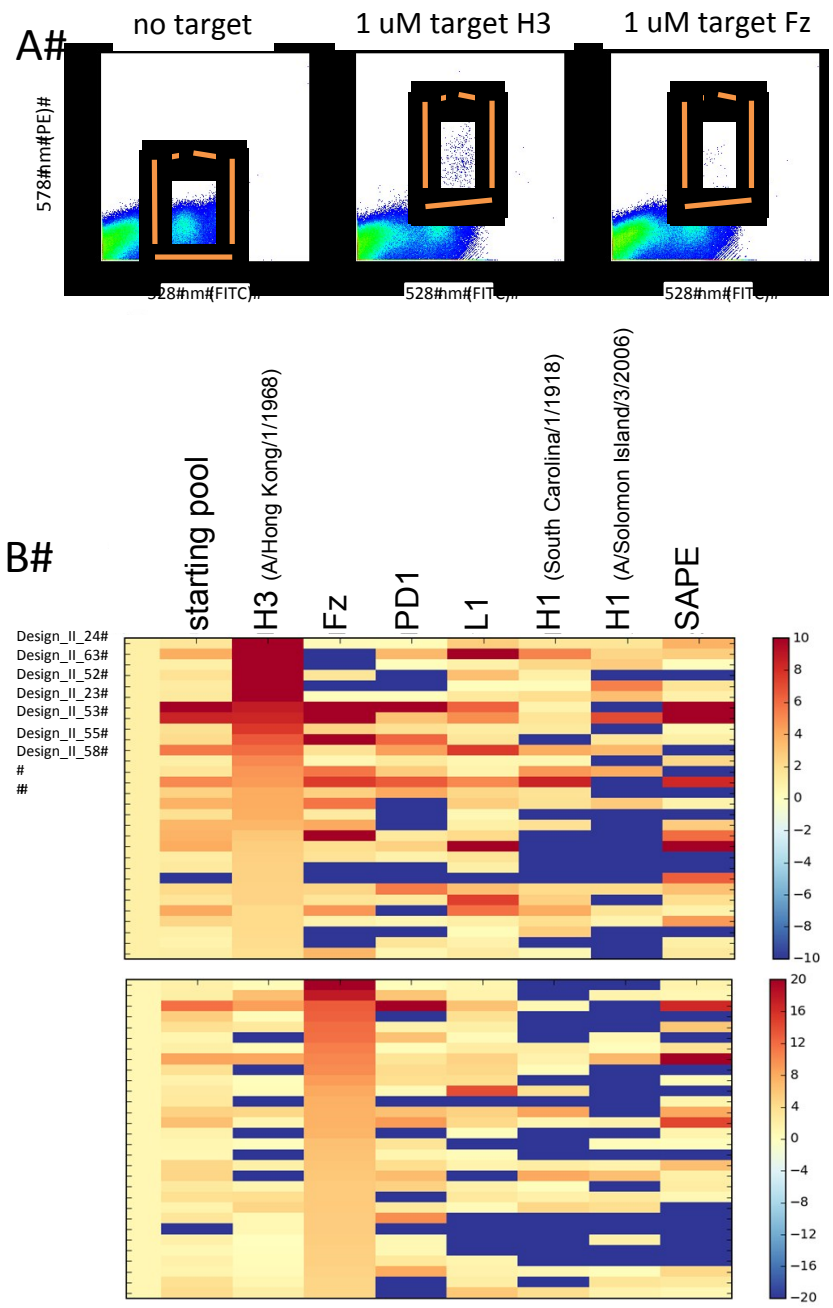

Figure S5. Protein-Protein Interaction Screen of Pool IB. A pool of proteins designs with overlapping designs as in pool I was screened for binding using targets proteins H3 (A/Hong Kong/1/1968), frizzle, PD1, smallpox L1 protein, H1 (A/South Carolina/1/1918) and H1 (A/Solomon Islands/3/2006) and SAPE.

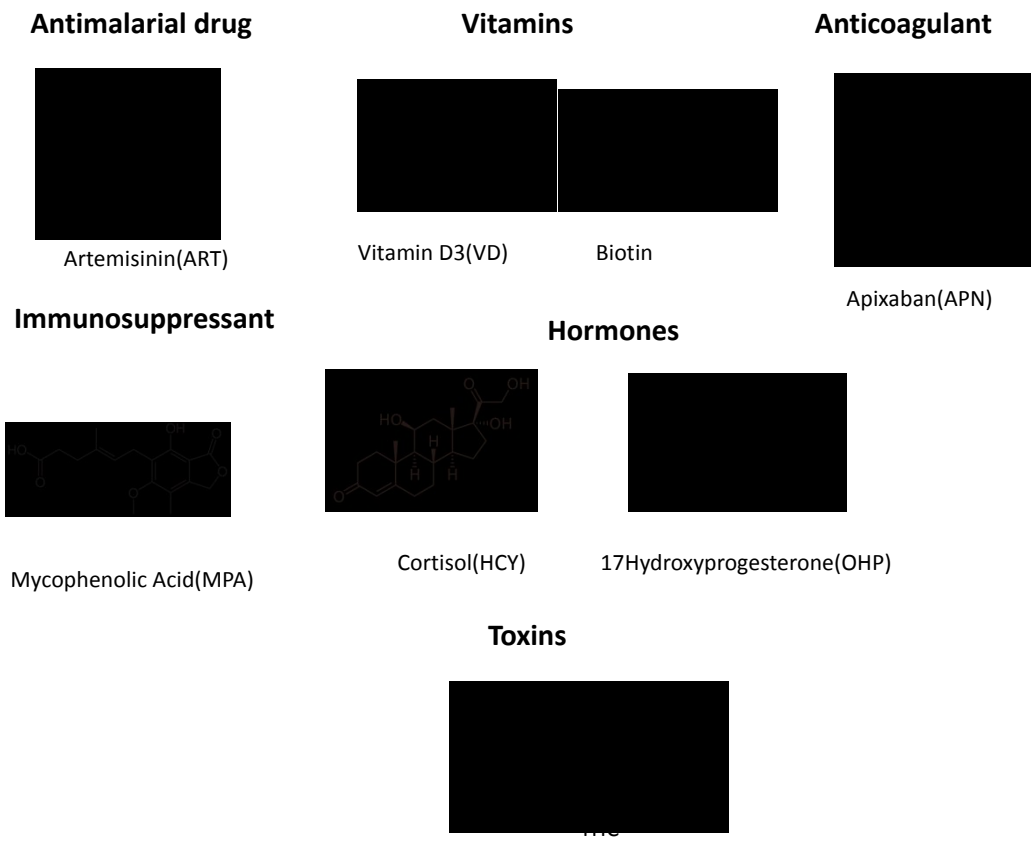

Figure S6. Structures of ligands using in screen.

Table S1. List of crystal structures from the PDB used for the design of new protein binders. If given, the fifth letter specifies the chain used for redesign.

1bse,1cc3,1dz0,1ifc,1ilu,1lpj,1sjv,1urr,1xb3,1xb6,1xb8,1xtn,2a9i,2aza,2bp2,2g3r,2g4s,2gi0,2hx8,2i5l,2o2w,2vq4,2z9t,3a07,3c0d,3fq1,3gxw,3jt2,3k1x,3lgl,3lwe,3o49,3o4a,3oan,3p5i,1ifva,2a0b,2i5ua,1p9ya,1rj1a,1v2za,1wpaa,2cj7a,2huja,2hxxa,2qffa,3g7ca,1yo7a,4f0ab,2ck2,3r8q,1bxv,2c9r,1bxv,3i57,1jwfa,1nzna,1rj1a,1srva,1tqga,1v84a,1yo7a,2d48a,2fupa,2oeba,2qr3a,2qsba,2qvpa,2w9ya,2zrra,3beea,3ezla,3iqta,3lysa,1b88a,1gena,1grja,1ifga,1lu4a,1neua,1tjxa,2coqa,2h71a,2i1ua,2peta,2pkda,2pnda,2qtda,2r2ya,2wfba,3f0pa,3gg6a,3iosa,3k74b,3mwza,1dzoa,1f46a,1hnfa,2c3va,2c60a,2IGBA,2ptva,2w9qa,3BN0A,3khqa,1YE8A,1r6ja,1WMMA,2rb8a,2ve8a,1c0eb,1L2TB,1nu0a,1qaua,1tj6b,2cxha,2EULB,2feoa,2qgvb,2wmyf,3ec3a,3ii3a,3jtba,3klqb,3ld7a,1opd,1paq,3fz9,2cbpa,3ijja,3c7la,3fgxa,1shma,1u9pa,2fi9a,3gaxa,1cewa,1sifa,1xaka,2j5aa,2cs7c,1pgaa,1kh0b,1k52b,2zcba,1bxua,1bxva,3htyk,3htym,3htyn,3htyk,3htym,3htyn,2vwra,1SQRA,2vwra,3ld7c,2vwra,2bk8a,1yn3a,3ft9a,2wbxa,1yn5a,2ywqc,2cj3b,2YWQB,2ywqc,3kgrc,3pz8d,3kgrc,3pz8d,3a2eb,1zoxa,1xeda,1xedb,1xeda,1xedb,2nmsa,1em7a,1kh0b,2zcba,2pkta,1SQRA,2vwra,3ld7c,2bk8a,1yn3a,3ft9a,1yn5a,1yn5b,1yn5a,1yn5b,2YWQB,2ywqc,2cj3b,2YWQB,3kgrc,1bxv,3r8q,2ck2,2c9r,3i57,1em7a,1k52b,1SQRA,1bxva,1em7a,1k52b,1bxva,1SQRA,1bse,1cc3,1dz0,1ifc,1ilu,1lpj,1sjv,1urr,1xb3,1xb6,1xb8,1xtn,2a9i,2aza,2bp2,2g3r,2g4s,2gi0,2hx8,2i5l,2o2w,2vq4,2z9t,3a07,3c0d,3fq1,3gxw,3jt2,3k1x,3lgl,3lwe,3o49,3o4a,3oan,3p5i,1b88a,1ep8a,1jo0a,1joia,1nrva,1nwmx,1O13A,1r6ja,1ULRA,1V76A,1xaka,1ysra,2croa,2d58a,2e7va,2h71a,2i6va,3mxza,1h4xa,1svya,2bl7a,2igpa,3ZZPA,3C0De,3fv5b,3egyx,2wi1a,1l7ma,3hrva,3fv5a,1r75a,2bxwa,3LFGa,2BQAa,1brga_2,1brga_1,3D9Ac,3CS

Pa,2BQKa,1TIGa,1GBWa,1bsbc,3hrva,1r0fa,1qkxa,1kp5a,1g13a,2cw4a,2bxwa,2akpa,1uyla,1rhoc,3ii2a,2x8xx,2vw5a,2qg2a,3ckfa,2zmwd,3nhmb,3ii2a,3hksb,3pu2h,3EAKa,3A0Va,2DRZa,2zmwd,1J5PA,3fv5b,3lnfa,1Q9UB,3lnfb,1vcaa,1jfmb,1ft3a,2DRZa,3lnfb,3hmba,2wi3a,2iwsa,1BRIb,3egyx,3fv5b,3ekca,2qfob,1jfmc,3rga,3ke7,2z76,3ke7,3juq,3juo,3jun,3jum,6std,4std,3std,3ke7,2z7a,2z76,1ohp,1oh0,1ogx,2z77,1ohp,2z7a,6std,2z76,1ohs,1w6y,2z77,2z76,1ogx,1oh0,2z7a,2z77,3fka,2x34,1xj4,1iwm,1y0g,3fka,1wub,2vv8,1y0g,2vv7,1ew0,3ef8,2z77,2x32,1wub,2z77,1gvf,3g16,3fka,1wub,1y0g,1wub,1vzz,1wub,3h3h,3hpe,2x32,1y0g,2v1a,1sjw,1nww,1iwm,1xj2,3b2j,3t4h,3nr3,2erv,3stm,1hmt,2rct,3vg2,3hx8,3hpe,1iwm,1iiu,3gn8,2b46,3gn8,1w01,1oh0,1ikt,1iiu,1hbq,1axk,1iwm,1iiu,1ikt,1oh0,1sjw,3akn,2b46,1ikt,3gn8,1iwm,1axk,1iiu,1oh0,3gn8,1iwm,3hx8,2z77,2z76,2o62,1wub,1erb,3r56,1lnm,3mf6,3ke7,2r56,3emm,2oz7,3fel,4a8u,1f5j,2r56,1lnm,3ke7

Table S2

Summary of origins and pdbs for the PPI pools

Origin	number of proteins
Phage 434	1
Pseudomonas fluorescens bv. a	1
Streptococcus pyogenes	1
Streptococcus sp. gx7805	1
Pseudomonas syringae pv. tomato	1
Orectolobus maculatus	1
Lupinus luteus	1
Methylibium petroleiphilum pm1	1
Anabaena sp.	1
Pseudomonas putida	5
Geobacillus stearothermophilus	2
Lactococcus lactis	1
Silicibacter pomeroyi dss-3	1
Enterococcus faecium	1
Streptococcus pneumoniae	1
Thermosynechococcus elongatus	2
Aeropyrum pernix	1
Sinorhizobium meliloti	1
Myxococcus xanthus	1
Streptomyces lasaliensis	1
Saccharomyces cerevisiae	4
Magnaporthe grisea	3
Enterococcus mundtii	1
Pseudomonas testosteroni	2
Lumbricus terrestris	1
Pyrococcus horikoshii	2
Acidianus filamentous virus 1	2

Chlamydomonas reinhardtii	1
Pseudomonas aeruginosa	13
Novosphingobium aromaticivorans dsm 12444	1
Aquifex aeolicus	3
Desulfovibrio gigas	1
Haemophilus influenzae	1
Vibrio cholerae	2
Arabidopsis thaliana	3
Nostoc punctiforme	1
Xenopus tropicalis	1
Parabacteroides distasonis atcc 8503	1
Camelus dromedarius	1
Bacteroides thetaiotaomicron vpi-5482	1
Bacillus subtilis	4
Thermus thermophilus	8
Avena sativa	1
Cucumis sativus	1
Synechococcus elongatus	2
Ashbya gossypii	1
Borrelia burgdorferi	1
Drosophila melanogaster	1
Mesorhizobium loti	1
Helicobacter pylori	1
Bacteroides fragilis	1
Saccharophagus degradans	2
Pyrococcus furiosus	1
Shigella flexneri 2a	1
Vibrio parahaemolyticus rimd 2210633	3
Streptomyces nogalater	1
Bradyrhizobium japonicum	4
Actinomycete	1
SARS coronavirus	1
Betula pendula	1
Gallus gallus	4
Rhodococcus erythropolis	1
Dictyostelium discoideum	1
Thermotoga maritima	4
Aequorea victoria	1
Burkholderia pseudomallei 1710b	1
Lactobacillus reuteri	1

Thermoplasma acidophilum dsm 1728	1
Phleum pratense	1
Shewanella amazonensis	1
Achromobacter xylosoxidans	1
Ginkgo biloba	1
Bacillus caldolyticus	1
Thermopolyspora flexuosa	1
Candida glabrata	1
Methanocaldococcus jannaschii dsm 2661	1
Leishmania major	1
Rattus norvegicus	7
Xenopus laevis	1
Bacillus halodurans	2
Burkholderia thailandensis e264	1
Escherichia coli k-12	1
Staphylococcus aureus	2
Clostridium pasteurianum	1
Fungia concinna	1
Mycobacterium tuberculosis	7
Bos taurus	5
Bacillus sphaericus	1
Alcaligenes xylosoxidans	1
Archaeoglobus fulgidus	1
Listeria innocua	2
Dictyoglomus thermophilum	1
Nicotiana tabacum	2
Staphylococcus aureus subsp. aureus	1
Pseudomonas aeruginosa pak	1
Pieris brassicae	1
Thermoanaerobacter tengcongensis	1
Bacillus stearothermophilus	1
Escherichia coli	19
Finegoldia magna	2
Homo sapiens	67
Caenorhabditis elegans	1
Mus musculus	17
Ethanocaldococcus jannaschii	2
Ixodes scapularis	1
Lama glama	3
Bacillus amyloliquefaciens	5

Burkholderia sp.	4
Bartonella henselae	1
Streptococcus sp.	1
Solanum tuberosum	1
Synthetic construct	2
Rhodobacter sphaeroides	1
Enterococcus faecalis	1
Rhizopus oryzae	1
Populus tremula x populus tremuloides	1

Summary of origins for the SMP pools

Origin	number of proteins
pyrococcus horikoshii	1
pyrobaculum aerophilum	1
dictyoglomus thermophilum	1
streptomyces lasaliensis	1
burkholderia sp.	4
gallus gallus	1
rhodococcus erythropolis	1
pieris brassicae	1
arabidopsis thaliana	1
planctomyces limnophilus	1
trypanosoma cruzi	1
methylibium petroleiphilum pm1	1
escherichia coli	8
homo sapiens	23
pseudomonas putida	13
streptomyces roseochromogenes subsp. oscitans	1
argas monolakensis	1
clostridium difficile	1
bacillus subtilis	5
burkholderia glumae	1
mus musculus	1
silicibacter pomeroyi dss-3	1
ectothiorhodospira halophila	1
rhodococcus sp. rha1	1
catenulispora acidiphila	1
rhizobium loti	1
helicobacter pylori	1
streptococcus pneumoniae	1

thermus thermophilus	1
avena sativa	1
aedes aegypti	1
streptomyces galilaeus	1
klebsiella pneumoniae	1
halorhodospira halophila	1
nostoc punctiforme	1
comamonas testosteroni	2
thermopolyspora flexuosa	1
enterobacter cloacae	1
sinorhizobium meliloti	1
betula pendula	1
burkholderia thailandensis e264	1
rattus norvegicus	4
halomicrobium mukohataei	1
anopheles gambiae	1
mesorhizobium loti	3
kribbella flavida	1
magnaporthe grisea	4
pectobacterium atrosepticum	1
marinobacter aquaeolei	1
pseudomonas testosteroni	2
bacillus licheniformis	2
shigella flexneri	1
mycobacterium tuberculosis	5
saccharophagus degradans	2
novosphingobium aromaticivorans	1
candida albicans	1
novosphingobium aromaticivorans dsm 12444	1
bos taurus	4
pseudomonas putida kt2440	1
parabacteroides distasonis atcc 8503	1
rhodobacter capsulatus	1
drosophila melanogaster	1
rhodnius prolixus	2
pseudomonas aeruginosa	2
streptomyces nogalater	1
bradyrhizobium japonicum	5
sphingomonas paucimobilis	1

Table S3. Primers for Sequencing

InnerF1-A
TCTTTCCCTACACGACGCTCTTCCGATCT ACCAGGCGCTGG
GCTTATTTCTGAAGAGGACTTGTAATAG
InnerF1-B
TCTTTCCCTACACGACGCTCTTCCGATCT GAGGCCTTGGCC
GCTTATTTCTGAAGAGGACTTGTAATAG
InnerF1-C
TCTTTCCCTACACGACGCTCTTCCGATCT CTTTAAAATATA
GCTTATTTCTGAAGAGGACTTGTAATAG
InnerF1-D
TCTTTCCCTACACGACGCTCTTCCGATCT TGACTTGCACAT
GCTTATTTCTGAAGAGGACTTGTAATAG

Inner-R1
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GGCCAGTGAATTGTAATACGAC

OuterF
AATGATACGGCGACCACCGAGATCTACAC
TCTTTCCCTACACGACGCTCTTCCGATCT

OuterR-idx_1
CAAGCAGAAGACGGCATACGAGAT CGTGAT
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx2
CAAGCAGAAGACGGCATACGAGAT ACATCG
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx3
CAAGCAGAAGACGGCATACGAGAT GCCTAA
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx4
CAAGCAGAAGACGGCATACGAGAT TGGTCA
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx5
CAAGCAGAAGACGGCATACGAGAT CACTGT
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx6
CAAGCAGAAGACGGCATACGAGAT ATTGGC
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx7

CAAGCAGAAGACGGCATACGAGAT GATCTGG
TGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx8
CAAGCAGAAGACGGCATACGAGAT TCAAGT
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx9
CAAGCAGAAGACGGCATACGAGAT AAGCTA
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx10
CAAGCAGAAGACGGCATACGAGAT GTAGCC
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx11
CAAGCAGAAGACGGCATACGAGAT TACAAG
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx12
CAAGCAGAAGACGGCATACGAGAT TTGACT
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx13
CAAGCAGAAGACGGCATACGAGAT GGAACT
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx14
CAAGCAGAAGACGGCATACGAGAT TGACAT
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx15
CAAGCAGAAGACGGCATACGAGAT GGACGG
GTGACTGGAGTTCAGACGTGTGCTCTTC

OuterR-idx16
CAAGCAGAAGACGGCATACGAGAT CTCTAC
GTGACTGGAGTTCAGACGTGTGCTCTTC

InnerF2-A
TCTTTCCCTACACGACGCTCTTCCGATCT ACCAGGCGCTGc GGT CGG CTA GCC ATA TG
InnerF2-B
TCTTTCCCTACACGACGCTCTTCCGATCT GAGGCCTTGGCg GGT CGG CTA GCC ATA TG
InnerF2-C
TCTTTCCCTACACGACGCTCTTCCGATCT CTTTAAAATATA GGT CGG CTA GCC ATA TG
InnerF2-D
TCTTTCCCTACACGACGCTCTTCCGATCT TGACTTGCACAT GGT CGG CTA GCC ATA TG

InnerR2-A
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT ACCAGGCGCTGG
TCCGCCTCCCTCGAG
InnerR2-B
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT GAGGCCTTGGCC
TCCGCCTCCCTCGAG
InnerR2-C
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT CTTTAAAATATA
TCCGCCTCCCTCGAG
InnerR2-D
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT TGACTTGCACAT
TCCGCCTCCCTCGAG