Molecular Cell, Volume *42*

Supplemental Information

A De Novo Protein-Binding Pair

By Computational Design and Directed Evolution

John Karanicolas, Jacob E. Corn, Irwin Chen, Lukasz A. Joachimiak, Orly Dym, Sun H. Peck, Shira Albeck, Tamar Unger, Wenxin Hu, Gaohua Liu, Scott Delbecq, Gaetano Montelione, Clint Spiegel, David R. Liu, and David Baker

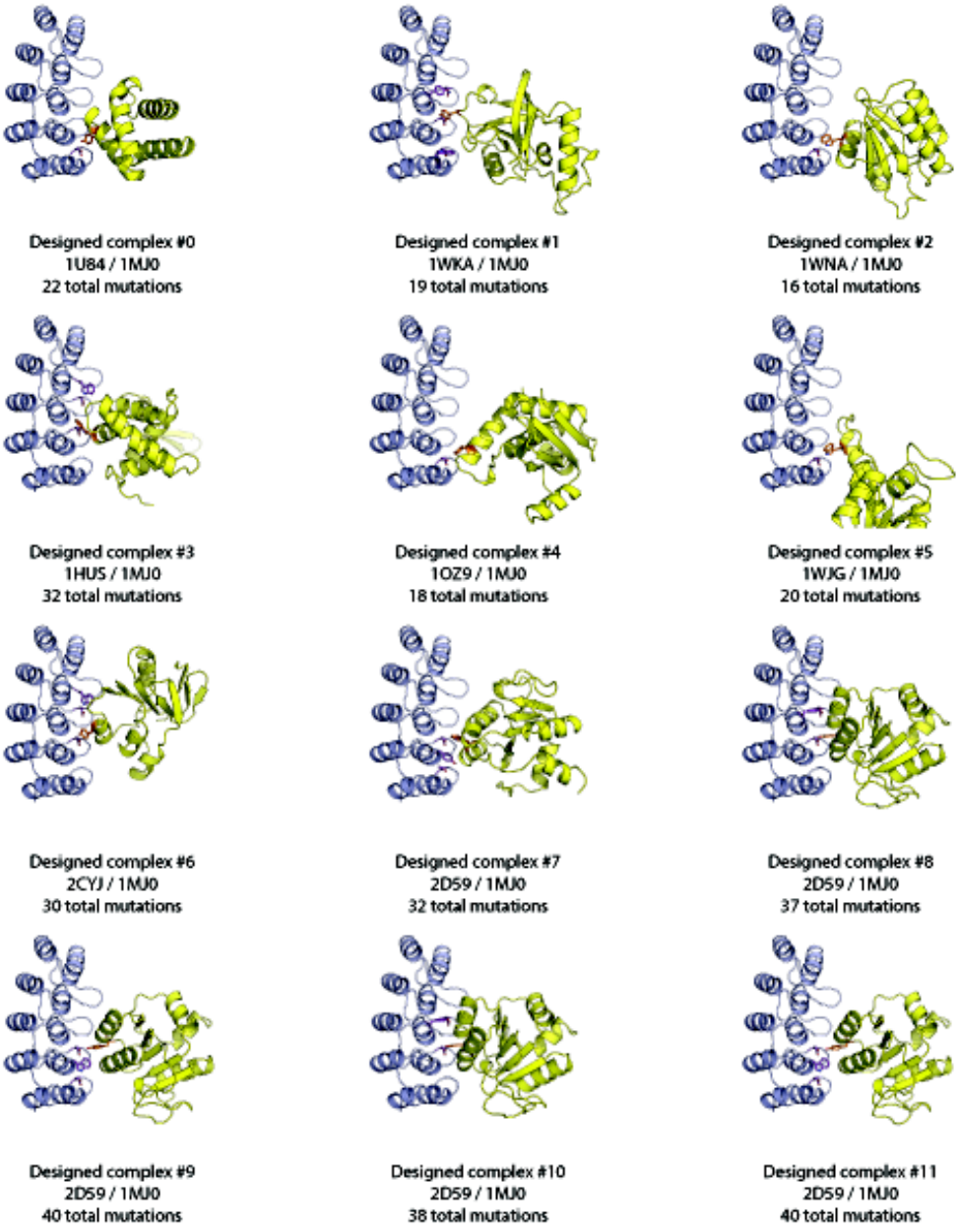

Figure S1. Design models of the 12 experimentally characterized pairs, related to Figure 1. Models are structurally oriented on the basis of the Ankyrin Repeat protein. Designed interfaces incorporated an

average of 29 total mutations from the wild-type starting protein pair. Complete sequence information is available in "Design models used for experimental characterization" section of Supplemental Materials.

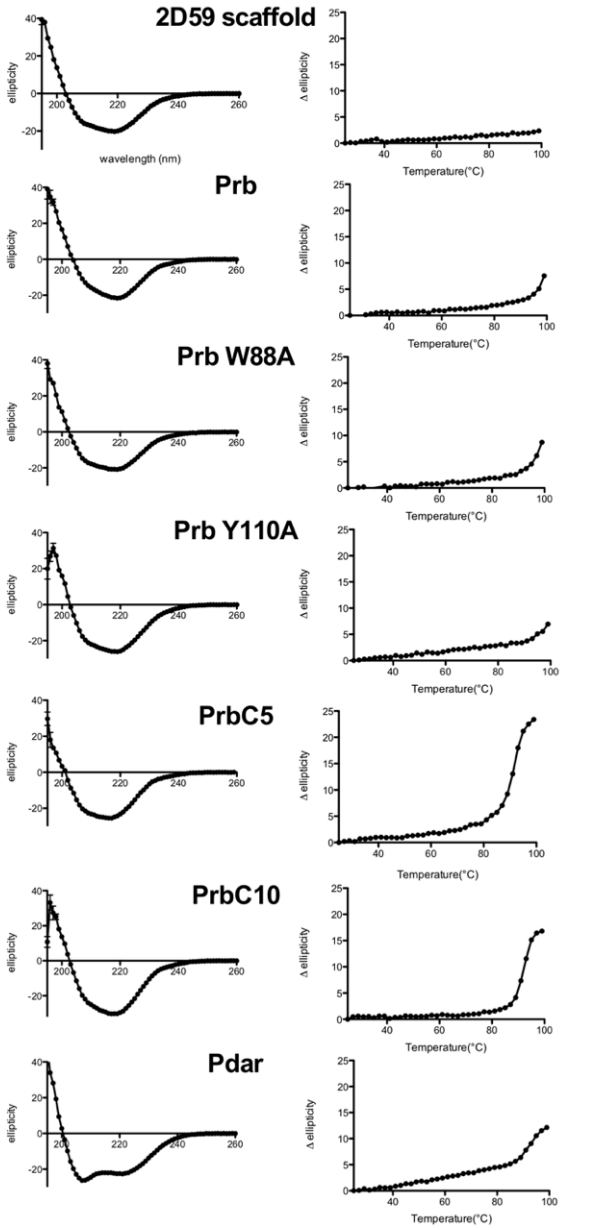

Figure S2. Structural integrity of Prb and Pdar, related to Figure 4. The Prb and Pdar designs are stable and well-folded at room temperature. The affinity-matured PrbC5 and PrcC10 clones display melting points lower than those observed for Prb or the PH1109 scaffold, consistent with a trade-off between binding affinity and thermal stability.

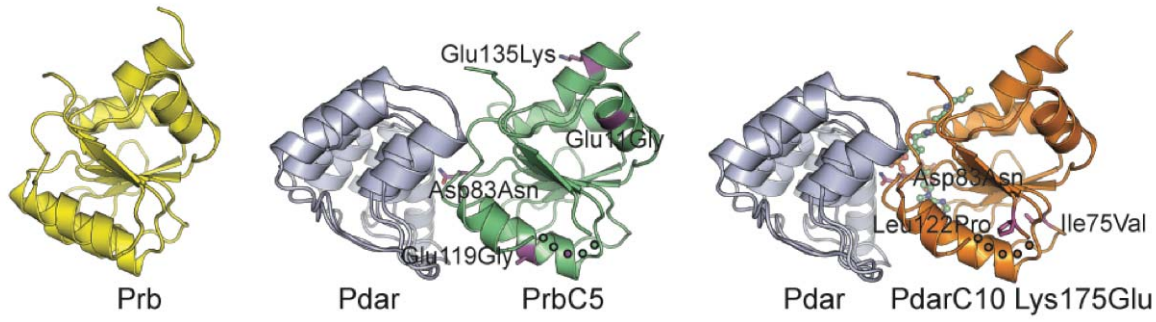

Figure S3. Prb mutations introduced during affinity maturation, related to Figure 6. Crystal structures of the Pdar-PrbC5 complex (gray and green) and the Pdar-PrbC10 Lys135Glu complex (gray and orange) are shown in equivalent orientations. Mutations introduced during affinity maturation are shown as magenta sticks. The crystal structure of Prb (yellow) is shown for reference.

Supplemental Tables

Table S1, related to Figure 2. Quantitative comparison interaction density in design models. The PrbC10-Pdar crystal structure is included for comparison, though CoA was not included in the calculation.

Complex	Interaction energy (kcal/mol)	Interface SASA ($Å^2$)	Interaction density (kcal/mol·1000$Å^2$)	Binding detected (see Figure 2)
Design #0	-10.6	795.06	-13.3	√
Design #1	-14.4	1237.33	-11.6	
Design #2	-14.6	880.51	-16.6	
Design #3	-18.8	1425.7	-13.2	√
Design #4	-13.0	862.45	-15.0	
Design #5	-11.9	1009.07	-11.8	√
Design #6	-16.1	1390.16	-11.6	√
Design #7	-14.9	1267.41	-11.8	
Design #8	-14.4	1323.08	-10.9	
Design #9	-15.2	1099.99	-13.8	
Design #10	-13.0	1180.69	-11.0	
Design #11 (Prb-Pdar)	-14.5	1021.68	-14.2	√
PrbC10-Pdar (crystal structure)	-15.1	1546.83	-9.8	N/A

Table S2, related to Figure 5. Dissociation constants between Prb and Pdar variants as measured by various methods.

Interaction Pair	$K_d^{F,S,YP,\ YH}$
Prb-Pdar	100-140 $nM^{F,S,\ YH}$
Prb-Pdar + 1 mM CoA	130 nM^{F}
PrbW88A-Pdar	> 1 μM^{F}
PrbY110A-Pdar	> 1 μM^{F}
PH1109-Pdar	> 1 μM^{F}
PrbC5-Pdar	1.3 nM^{YH}, 5.2 nM^{YP}
PrbC5(G11D)-Pdar	5 nM^{YP}
PrbC5(N83D)-Pdar	> 300 nM^{YP}
PrbC5(A92V)-Pdar	17 nM^{YP}
PrbC5(G119E)-Pdar	12 nM^{YP}
PrbC5(K135E)-Pdar	4 nM^{YP}
PrbC10-Pdar	4.3 nM^{YH}, 11.1 nM^{YP}
PrbC10(V75I)-Pdar	7 nM^{YP}
PrbC10(N83D)-Pdar	> 1 μM^{YP}
PrbC10(P122L)-Pdar	18 nM^{YP}
PrbC10(K135E)-Pdar	17 nM^{YP}
Prb(D83N)-Pdar	5.8 nM^{YH}
Prb-PdarC1	1.3 nM^{YH}, 18 nM^{YP}
Prb-PdarC1(D34N)	50 nM^{YH}
Prb-PdarC1(T76A)	2.2 nM^{YH}
Prb-PdarC1(A103T)	2.0 nM^{YH}

Prb-PdarC1(R113Q)	1.4 nM[YH]
Prb-Pdar(N34D)	4.7 nM[YH]
Prb(D83N)-Pdar(N34D)	180 pM[YH]
PrbC10-PdarC1	296 pM[YH], 4 nM[YP]

[F] measured by fluorescence polarization

[S] measured by surface plasmon resonance

[YP] measured by yeast display in 10 mM phosphate + 150 mM NaCl +0.1% BSA

[YH] measured by yeast display in 20 mM HEPES pH 7.5 +50 mM potassium glutamate +0.1% BSA

Table S3, related to Figures 4 and 6. Summary of data collection from crystals of Prb, Pdar-PrbC5 and Pdar-PrbC10.

Data Collection	Prb	Pdar_PrbC15	Pdar_PrbC10
Resolution range (Å)[a]	33.3-1.9 (1.97-1.9)	50.0-2.0 (2.03-2.0)	50.0-2.2 (2.24-2.2)
Space group	$P2_12_12_1$	*P*1	*P*1
Unit Cell Dimensions:			
a=(Å)	40.00	53.12	53.90
b=(Å)	59.39	56.56	57.62
c= (Å)	60.04	56.98	58.23
α=(º)	90.00	89.96	89.96
β=(º)	90.00	112.42	90.14
γ=(º)	90.00	90.03	113.44
Molecules in asymmetric unit	1	2	2
Total reflections measured	44,972	157,886	47,173
Unique reflections[a]	11,574	40,226 (1957)	27,228 (1023)
R_{sym}[a,b]	0.04 (0.124)	0.05 (0.37)	0.06 (0.26)
Completeness (%)[a]	98.2 (96.3)	98.0 (93.1)	83.4 (63.1)
Redundancy	3.89 (3.91)	3.9 (3.8)	1.7 (1.5)
<I> / <s (I) >	22.2 (8.1)	20.5 (3.5)	12.5 (2.1)
Refinement statistics			
Resolution limits (Å)	30.0 – 1.90	50.0-2.0	50.0-2.2
R_{work}[c](%)	21.32	20.6	25.0
R_{free}[c](%)	24.17	26.5	32.8

Wilson mean B factor ($Å^2$)	22.66	24.77	33.44
rms deviations (Å)	0.012	0.022	0.038
Bond angles (°)	1.614	1.99	1.99
Ramachandran Plot			
Most favored (%)	96.4	98.3	88.6
Allowed (%)	3.6	1.7	9.9
Disallowed regions (%)	0	0.0	1.6

[a] Values in parentheses correspond to the highest-resolution shell.

[b] $R_{sym}=\sum | \langle I_{hkl}\rangle - I_{hkl} | / I_{hkl} |$, where $\langle I_{hkl}\rangle$ is the average intensity of symmetry-related reflections and I_{hkl} is the observed intensity.

[c] $R= \sum | |F_o| - |F_c| | / \sum |F_o|$, where F_o denotes the observed structure factor amplitude and F_c the calculated one.

Supplemental Experimental Procedures

Protein scaffolds

Naturally occurring examples of AR modules have been used to construct a consensus AR sequence, which was found to maintain the AR fold and demonstrated increased thermodynamic stability (Kohl et al., 2003; Letunic et al., 2006; Mosavi et al., 2004). We used the crystal structure of this consensus AR as our starting point for designs (PDB accession code 1mj0).

The large number of AR sequences present in nature allows construction of a very deep multiple sequence alignment. The diverse set of AR binding partners makes it unlikely that any residues are conserved due to functional constraints, which in turn implies that strongly conserved residues are important for stability; the amino-acid identity at these positions were restricted during design calculations to those occurring frequently in the multiple sequence alignment. The identity of residues distant from the AR binding groove were also fixed. Using the residue position numbering of Kohl et al. (Kohl et al., 2003) for AR repeats, sequence restrictions were as follows:

Position 1 – restricted to Asp/Asn

Position 4 – restricted to Gly

Position 5 – restricted to Phe/Ile/Leu/Met/Val/Trp/Tyr

Position 6 – restricted to Thr

Position 7 – restricted to Pro

Position 8 – restricted to Leu

Position 9 – restricted to His

Position 10 – restricted to Ala/Ile/Leu/Val/Trp/Tyr

Position 11 – restricted to Ala

Position 15 – restricted to Gly

All amino acids were allowed at positions 2, 3, 12, 13, 14, 16, 30, 31, 32, and 33. Positions 17 to 29 were fixed in structure as well as sequence (ie. excluded from the design entirely).

We compiled a set of protein scaffolds to pair with the AR by identifying proteins from thermophilic organisms which are small, expressible in *E. coli*, stable, globular, and have a crystal structure with resolution not worse than 2.5 Å. The list of PDB accession codes for these targets was: 1BXE, 1HUS, 1IU9, 1IUK, 1MGT, 1O13, 1O6D, 1OZ9, 1RL6, 1RSS, 1SHE, 1TMY, 1TZV, 1U84, 1ULR, 1VJK, 1VJX, 1VKU, 1VMB, 1WJG, 1WKA, 1WNA, 1WRJ, 1WS6, 1X3O, 1XE1, 1YD0, 2CMX, 2CVK, 2CWR, 2CWY, 2CYJ, 2CZW, 2D59, 2ETD, 2HD9 and 2HIA. Amino-acid residues critical for stability of these proteins were identified by using RosettaDesign to individually estimate the energetic consequence of a substitution to alanine at each position. The wild-type amino-acid was enforced throughout the design protocol at any position for which an alanine substitution was found to destabilize the protein by at least 1.5 kcal/mol.

Docking Protocol

To build a set of bound orientations that exhibited overall shape complementarity, we carried out rigid-body docking using the PatchDock program(Schneidman-Duhovny et al., 2003). This program identified surface features on each component protein, then exhaustively matched each concave features on one protein with the convex features on its partner. A constraint was applied to restrict bound orientations to those involving the AR binding groove. The top 500 orientations were selected via PatchDock's shape complementarity metric, and carried forward. If PatchDock could not identify more than 500 bound orientations (because of the particular surface features of a given protein pair), all orientations were carried forward.

The PatchDock search was designed to exclude bound orientations which are structurally very similar, and focus instead on structurally diverse orientations with global shape complementarity. Having found that subtle changes in the relative orientation of the two proteins would lead to different designed sequences, we used the low-resolution "small perturbations" docking protocol of RosettaDock (Gray et al., 2003a; Gray et al., 2003b) to build 100 new bound orientations from each one generated by PatchDock.

In all stages of docking, alanine sidechains were used at all positions that would be allowed to vary in subsequent steps (these were specified as described in "Protein Scaffolds" above). After these docking steps were complete, the wild-type sidechains were restored from the crystal structures of the protein scaffolds.

Motif-searching Protocol

The goal for this step of our protocol was to rapidly identify those orientations from the docking step (above) capable of harboring at least two of the desired motifs (Figure 1). This is implemented in Rosetta version 2.3 (http://www.rosettacommons.org/), and can be accessed using the "-incorporate_hotspot" flag in the "-design" mode. For the designs described in this study, an additional flag "-AR_conservation_file" was added to allow the identity of the structurally conserved aspartate residues to be specified (these would serve as hydrogen bond acceptors).

Collectively, these command-line flags invoke the protocol used in this study. Each amino acid position on the non-AR that is within 7 Å of the AR was subjected to screening for an initial motif site. Screening entailed first building a large set of tyrosine and tryptophan (backbone-dependent) rotamers (Dunbrack and Cohen, 1997a) onto the current amino acid position of the non-AR backbone. The "-ex1" flag was used to increase the number of rotamers. Each of these rotamers were then evaluated to determine whether it made an intermolecular hydrogen bond worth at least -0.75 kcal/mol to one of the specified aspartate sidechains without a steric clash to any backbone or Cβ atoms.

For each motif thus identified, a secondary search was then carried out to determine whether an additional motif could also be incorporated. The only valid hydrogen bond acceptors in this secondary search were the aspartate residues one repeat adjacent to the one used by the primary motif (i.e. separated in sequence by exactly 33 residues). These could be satisfied via either an intramolecular or an intermolecular hydrogen bond. Intramolecular hydrogen bonds were identified by screening rotamers at two AR positions: 29 residues before the conserved aspartate (tryptophan only) and 9 residues after the conserved aspartate (tryptophan or tyrosine). Intermolecular hydrogen bonds were identified by screening

rotamers of all uncharged hydrogen bond donors at all amino-acid positions of the non-AR backbone. Orientations were discarded from further consideration if two non-overlapping motifs could not be identified. For each pair of motifs that were identified, the hydrogen bond geometries were then optimized by carrying out a energy minimization in which all intramolecular degrees of freedom were held fixed; this resulted in a subtle adjustment of the bound orientation.

Design Protocol

The goal for this step of our protocol was to identify optimal sidechains to stabilize the motifs identified above. This portion of the protocol is implemented in Rosetta version 2.3 (http://www.rosettacommons.org/), as an optional submode available using the "-incorporate_hotspot" flag in the "-design" mode.

For each pair of motifs, the hydrophobic layer surrounding the motifs (the "core" of the interface) was designed by using RosettaDesign (Kuhlman et al., 2003) to select optimal aliphatic amino acids at the backbone positions in the immediate vicinity of these two hydrogen-bonded aromatic sidechains. RosettaDesign implements a Monte Carlo search across discrete sidechain conformations that frequently occur in the Protein Data Bank (Dunbrack, 2002; Dunbrack and Cohen, 1997b). Sidechain energies are computed as the sum of a Lennard-Jones term, a hydrogen bonding term (Kortemme and Baker, 2002), an implicit solvation term (Lazaridis and Karplus, 1999), a statistical term representing the backbone-dependent internal free energies of amino acid rotamers (Kuhlman et al., 2003), and an amino acid dependent reference energy (Kuhlman et al., 2003). In order to encourage direct stabilizing interactions with the motif residues, these inter-residue energy were scaled by a factor of two relative to all other energetic contributions. To further optimize interactions around the motif residues, an energy minimization was carried out after the design step, in which all chi angles were optimized (backbone degrees of freedom were held fixed).

At this point, the interactions around the primary motif were re-evaluated to determine whether the surrounding residues provided a sufficiently good environment. Criteria required to move forward

were: a) the attractive part of the Lennard-Jones energy for this residue must be better than -7.0 kcal/mol, b) the solvent accessible surface area in the complex must be less than 40 $Å^2$, and c) the largest void volume immediately adjacent to the motif residue must have a radius smaller than 1.25 Å. We further required that the total solvent accessible surface area buried by the complex be less than 2000 $Å^2$, and the interface contain at most 4 buried polar groups not engaged in hydrogen bonds.

Provided all these criteria were met, our protocol then continued to design the periphery of the interface while keeping the "core" fixed. Any residue outside the "core" with a heavy atom within 7.0 Å of an atom on the binding partner was included in this stage of the design. A final energy minimization was then carried out over all backbone and sidechain dihedral angles.

Selection of "Native-like" Design Models

Finally, a series of filters were used to select the most "native-like" of the design models. These orthogonal filters were calibrated using a set of naturally occurring protein-protein interfaces (Gray et al., 2003a). Filters were applied using both physical properties as well as energy-based metrics. A complete list of filters is as follows:

a) the total solvent accessible surface area buried by the complex must be less than 2000 $Å^2$

b) all of the AR conserved aspartates must either participate in a hydrogen bond or be exposed to solvent

c) the total intermolecular Lennard-Jones energy must be better than -10.0 kcal/mol

d) the interaction density (defined as the total intermolecular Lennard-Jones energy divided by the solvent accessible surface area) must be better than -12.0 kcal/mol·1000$Å^2$

e) the largest single empty cavity in the interface, computed via RosettaHoles (Sheffler and Baker, 2009), must have a volume less than 75.0 $Å^3$

f) the score derived from buried polar chemical groups which do not participate in a hydrogen bond, calculated using the "-identify_interface_UNS" flag in Rosetta, must be less than 7.0

Using these filters we identified a set of design models which exhibited native-like features. These designed interfaces were compared to three previously identified interfaces involving AR proteins: GABPβ bound to GABPα (PDB code 1awc, K_d = 0.78 nM) (Batchelor et al., 1998; Desrosiers and Peng, 2005), a ribosome-display evolved AR "AR_3a" bound to aminoglycoside phosphotransferase (APH) (PDB code 2bkk, K_d = 1.7 nM) (Amstutz et al., 2005; Kohl et al., 2005), and a ribosome-display evolved AR "off7" bound to maltose binding protein (MBP) (PDB code 1svx, K_d = 4.4 nM) (Binz et al., 2004). The latter two interfaces may represent the most analogous comparison for this computational experiment, since these interfaces were the result of *in vitro* selections for binding alone, in the absence of any additional functional constraints acting on naturally-occurring protein sequences.

Overall, the close similarity of these features between this set of designed interfaces and the set of AR-containing interfaces indicates that this computational approach creates designed complexes that resemble naturally-occurring complexes which bind with high-affinity.

Design models used for experimental characterization

Sequences of each designed complexes are reported below, as a list of mutations from the starting scaffold proteins (sequences for these are available via the PDB). In all cases "a" refers to the Ankyrin Repeat protein and "b" refers to its partner. Note that numbering is relative to PDB entry of scaffold protein, hence numbering for Prb-Pdar does not match listed mutations below. For example, the Design 11a mutations are offset by +3 from Pdar, and the Design 11b mutations are offset by +12 from Prb.

Coordinates corresponding to all twelve design models have been made available on Proteopedia (www.proteopedia.org), which provides an interactive interface for examining them. They will be made publicly available pending publication.

Design #0a – 1MJ0 with 16 substitutions: N45D, Y48R, T79N, D110N, N111E, D112S, H114L, L119A, K122I, Y123T, K144S, F145Q, K147L, I152L, D155L, N156T

Design #0b – 1U84 with 6 substitutions: Q52L, S53L, E56Y, F57G, D60S, E61R

Design #1a – 1MJ0 with 17 substitutions: N45D, Y48N, D77N, T79N, I81M, L86Y, A89W, D112Q, H114M, L119A, K122S, Y123S, K144S, F145T, I152W, D155K, N156K

Design #1b – 1WKA with 2 substitutions: E231Y, K262P

Design #2a – 1MJ0 with 9 substitutions: N45D, D112L, H114T, K122F, K144S, F145M, K147R, I152L, D155L

Design #2b – 1WNA with 7 substitutions: E108G, E109R, L111W, K112W, R113Y, A116L, I117Y

Design #3a – 1MJ0 with 22 substitutions: R23Q, N45D, D46S, Y48W, S56R, L78N, T79N, A89G, T90S, N111S, D112G, H114F, L119I, K122L, Y123S, H125Q, D143T, K144S, F145L, K147L, D155N, N156Q

Design #3b – 1HUS with 10 substitutions: Y17F, K55L, D56P, E59W, E62M, Q63L, K66N, N67A, R95Q, N128Y

Design #4a – 1MJ0 with 11 substitutions: N45D, I81T, D112L, H114L, K122M, Y123R, F145M, K147L, I152L, D155M, N156M

Design #4b – 1OZ9 with 7 substitutions: E130P, E131Q, E132R, K133Y, K134F, R136E, E137F

Design #5a – 1MJ0 with 13 substitutions: N45D, Y48L, T79N, I81R, D110T, D112A, H114L, K122F, K144S, F145M, K147L, D155L, N156Q

Design #5b – 1WJG with 7 substitutions: E48M, F50N, F51M, E52W, L55W, R56A, L59R

Design #6a – 1MJ0 with 19 substitutions: K16N, E20K, R23W, N45S, D46E, Y48W, S56A, L78R, T79K, I81V, A89W, H92Q, N111Q, D112A, H114L, K122M, Y123S, H125Q, F145N

Design #6b – 2CYJ with 11 substitutions: K41T, H42K, V63I, T68M, S75E, E79K, K81M, K82Y, E85S, E88I, K108Y

Design #7a – 1MJ0 with 20 substitutions: R23Q, N45S, D46S, Y48R, L78R, T79M, I81L, L86Y, A89R, N111R, D112G, H114Y, K122A, Y123K, H125Q, D143T, F145E, K147L, D155N, N156Q

Design #7b – 2D59 with 12 substitutions: K88D, M91A, E95F, Q96F, K99E, N113F, R114S, E115V, S117A, K118T, K119W, D121K

Design #8a – 1MJ0 with 20 substitutions: N45D, D46N, Y48L, S56A, H59Q, L78S, T79Q, I81V, L86Y, T90N, H92W, D112A, H114L, K122L, Y123T, H125Q, F145L, D155N, N156Q, N158Q

Design #8b – 2D59 with 17 substitutions: K58N, K86A, K88G, T90A, M91L, E92V, Y93F, E95I, Q96F, N113Y, R114G, E115L, S117A, K118L, K119L, D121E, E122S

Design #9a – 1MJ0 with 23 substitutions: N45D, D46N, Y48L, S56Q, N57Q, H59Q, L78S, T79A, A89Y, T90Q, H92W, N111R, D112A, H114W, K122L, Y123S, H125Q, K144S, F145E, K147L, D155N, N156Q, N158Q

Design #9b – 2D59 with 17 substitutions: K58N, K88A, L89R, T90A, M91W, E92R, Y93F, E95V, Q96Y, N113Y, R114P, E115L, S117A, K118Q, K119Q, D121E, E122N

Design #10a – 1MJ0 with 21 substitutions: R23A, N45D, D46N, Y48L, S56A, H59Q, L78N, T79G, I81V, L86Y, T90N, H92W, D112A, H114L, K122A, Y123T, H125Q, F145L, D155N, N156Q, N158Q

Design #10b – 2D59 with 17 substitutions: K58N, K86A, K88G, T90A, M91L, E92N, Y93F, E95I, Q96F, N113Y, R114P, E115L, S117A, K118L, K119T, D121K, E122S

Design #11a – 1MJ0 with 23 substitutions: R23A, N45D, D46N, Y48L, S56A, H59Q, L78S, T79A, A89Y, T90D, H92W, N111R, D112A, H114W, K122L, Y123S, H125Q, K144A, F145L, K147L, D155N, N156Q, N158Q

Design #11b – 2D59 with 17 substitutions: K58N, K86D, K88A, L89K, T90A, M91W, E92R, Y93F, E95V, Q96Y, N113Y, R114P, E115L, S117A, K118R, K119Q, D121K

Construction and cloning of genes

Synthetic genes corresponding to the designed protein pairs (Codon Devices) were cloned into bi-cistronic constructs in a modified version of pET29b containing an N-terminal StrepII tag and C-terminal 6xhistidine tag (Novagen). The co-expressed bi-cistronic constructs were used for initial binding analysis and crystallization trials. For surface plasmon resonance (SPR) and fluorescence polarization binding assays, each protein was cloned separately into pET29b containing only a C-terminal 6x histidine tag.

ELISA screen for binding partners

Induced cells for each co-expressed design pair were resuspended in lysis buffer (100 mM potassium glutamate, 20 mM HEPES 7.5), lysed by sonication, and cell debris was removed by precipitation. Cleared lysates were incubated with Ni-NTA plates (Qiagen) for one hour to bind His_6-tagged proteins and washed with buffer (100 mM potassium glutamate, 20 mM HEPES pH 7.5, 0.05% Tween-20, 0.01 mg/ml BSA). Horseradish peroxidase conjugated mouse anti-StrepII antibody (Pierce) was incubated for an hour followed by 4 x 200 µl buffer washes. Colorimetry was developed using Ultra TMB-ELISA substrate (Pierce) and optical densities were measured on a plate reader (Molecular Devices).

Purification and separation of component proteins

All computationally selected variants were transformed into BL21DE3 *E. coli* (Novagen), expressed in LB media and purified by affinity chromatography using Ni-NTA resin (Qiagen) in 50mM Hepes pH 7.5, 150mM NaCl and a $NiSO_4$ gradient. Proteins were better than 95% pure as determined by SDS-PAGE.

Surface plasmon resonance

Prb and Pdar were cloned into a pET29b derivative containing tandem C-terminal AviTag (Avidity) and hexahistidine tags. Purified protein was 100% *in vitro* biotinylated with BirA (Avidity), as verified by electrospray ionization mass spectrometry (ESI-MS). Biotinylated Pdar was coupled to a separate flow cells of a streptavidin sensor chip (BiaCore) at approximately 200, 400, and 1000 response units, with one flow cell left uncoupled as a reference cell. Biotinylated Prb was coupled to a separate streptavidin chip at approximately 200, 300, and 800 response units. Prb or Pdar expressed with only a hexahistidine tag were purified as described above and diluted in 100 mM potassium glutamate, 20 mM HEPES (pH 7.5), 3 mM EDTA, 0.005% Tween-20 (Buffer HBG-EP). In several experiments, titrations of Prb or Pdar were measured against a chip containing the opposite binding partner at flow rates between 50 ul/min and 100 ul/min. Binding curves were analyzed using Scrubber or GraphPad Prism.

Fluorescence polarization

Purified wildtype Pdar was labeled with Alexa Fluor 488 TFP ester (Molecular Probes), yielding an approximately 1:2 molar ratio of dye:protein, as quantified by ESI mass spectrometry. Notably, none of the lysine residues potentially modified by the dye are located within the Pdar binding site. Fluorescence polarization was measured by holding fluorescently labeled Pdar constant at 10 nM and titrating in each binding partner. Binding reactions were performed in a buffer of 50 mM potassium glutamate, 20 mM HEPES (pH 7.5), 10% glycerol, 1 mM DTT, and 0.1 mg/ml BSA. Reactions were prepared in triplicate and allowed to equilibrate for 1 hour at room temperature, after which polarization was determined using a SpectraMax m5^3 plate reader (Molecular Devices). Dissociation constants were determined by non-linear curve fitting using a single site model of total binding (GraphPad Prism).

Gel Filtration Analysis

Size exclusion chromatography was performed using a Superdex S200 10/300 column (GE Life Sciences) in a buffer of 150 mM NaCl, 20 mM HEPES (pH 7.5), and 1 mM DTT. Individual proteins purified by nickel chromatography were further purified with an initial sizing chromatography step, pooling only fractions of the correct molecular weight, as determined by SDS-PAGE. These re-purified proteins were then diluted to 40 μM in column running buffer, equilibrated for 30 minutes at room temperature, and subjected to gel filtration as either individual monomers or a 1:1 molar ratio mixture.

Dynamic Light Scattering

Prb, Pdar, or a 1:1 stoichiometric mixture of Prb and Pdar were dialyzed overnight into buffer (150 mM NaCl, 10 mM HEPES pH 7.5) before data collection. Dynamic light scattering data were collected on a Wyatt DynaPro 99-CP at 25 °C. Data were analyzed with DYNAMICS v6.

Circular Dichroism

Each protein was dialyzed into PBS (137 mM NaCl, 10 mM phosphate buffer, pH 7.2) for 2 hours at room temperature, and diluted to 10-30 μM in PBS. Data were collected on an Aviv 62A DS spectrophotometer (Avid Biomedical, Inc) in a 1 mm pathlength cuvette (Hellma). Wavelength scans were collected from 195-260 nm at 25 °C. Thermal denaturations were collected at 222 nm from 25-98 °C in 2 °C increments with 1 minute equilibration at each temperature. Buffer-only scans and thermal denaturations were collected immediately after each protein sample and used for background subtraction.

NMR spectroscopy

The Prb domain was cloned, expressed and purified following largely automated NESG standard protocols to produce a uniformly ^{13}C, ^{15}N-labeled protein sample (Acton et al., 2005). Briefly, the

designed *Prb* gene was cloned into a pET29b+ (Novagen) vector, yielding the plasmid pOR13-29. The resulting construct contains 8 nonnative residues at the C-terminus (LEHHHHHH) to facilitate protein purification. *Escherichia coli* BL21 (DE3) pMGK cells were transformed with pOR13-29, and cultured in MJ9 minimal medium (Jansson et al., 1996) containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. *U*-^{13}C, ^{15}N Prb was purified using an ÄKTAxpress™ (GE Healthcare) based two step protocols consisting of IMAC (HisTrap HP) and gel filtration (HiLoad 26/60 Superdex 75) chromatography. The final yield of purified *U*-^{13}C, ^{15}N Prb (> 98% homogeneous by SDS-PAGE; 18.6 kDa by MALDI-TOF mass spectrometry) was ~19 mg/L. The *U*-^{13}C,^{15}N and 5%^{13}C, *U*-^{15}N Prb were dissolved, respectively, at concentrations of ~0.51 mM in 95% H_2O/5% ^2H_2O (20 mM MES, 200 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3) at pH 6.5. ^{15}N T_1 and T_2 relaxation data (data not shown) indicate that the protein is monomeric in solution under the conditions used for these NMR studies.

All NMR spectra were recorded at 30 ºC using cryogenic probes. Triple resonance NMR data were collected on Varian INOVA 600 MHz, a simultaneous 3D $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-edited NOESY(Shen et al., 2005) spectrum (mixing time 100 ms) in H_2O were acquired on a Bruker AVANCE 800 MHz spectrometer. ^{15}N T_1 and T_2 relaxation measurements were made using ^{15}N T_1 and T_2 (CPMG) relaxation experiments, respectively (Farrow et al., 1994). All NMR data were processed using the program NMRPipe (Delaglio et al., 1995) and analyzed using the program XEASY (Bartels et al., 1995). Spectra were referenced to external DSS. Resonance assignments were achieved as described previously (Liu et al., 2005). About 96 backbone amides were observed in NHSQC spectra (Figure 4), and ~30 residues were unobserved due to exchange broadening. Backbone resonance assignments (H^N/N, H^α/C^α, and H^β/C^β) assignments were obtained in largely automated fashion with the program AUTOASSIGN (Moseley et al., 2001). The simultaneously NOESY and CCH-TOCSY was analyzed manually to obtain side-chain assignments. Assignments were obtained for ~70% of backbone and side-chain chemical shifts assignable with the NMR experiments listed above (excluding N-terminal NH_3^+, Lys NH_3^+, Arg NH_2, OH of Ser, Thr and Tyr, $^{13}C^\gamma$ of Asp and Asn, $^{13}C^\delta$ of Glu and Gln, and aromatic $^{13}C^\gamma$ shifts). Based on chemical shifts, the locations of regular secondary structure elements were identified (Wishart and Sykes, 1994).

Affinity Maturation

The Prb-Pdar complex was first affinity matured by creating a library of random Prb mutants and selecting for binding against Pdar using a combination of phage and yeast display. In a second directed evolution experiment, a library of random Pdar mutants was selected for binding against Prb using yeast display. The yeast display plasmid pCTCon2 and *S. cerevisiae* strain EBY100 were kind gifts from Dane Wittrup (MIT).

Prb phagemid library construction

The Prb gene was randomly mutated using error-prone PCR with mutagenic dNTP analogs to generate 1-5 mutations per gene (Zaccolo et al., 1996). The amplified Prb gene library was digested with *BamHI* and *Pst*, and ligated into similarly-digested pDST23 phagemid vector to yield a library of 2.9×10^8 transformants (Steiner et al., 2006). Following electroporation into *E. coli* XL1-Blue MRF (Stratagene), the entire SOC outgrowth was plated on 2xYT agar supplemented with 1% glucose, 34 μg/mL chloramphenicol, and 12.5 μg/mL tetracycline. After overnight incubation at 37 °C, the lawns of *E. coli* colonies were scraped into fresh 2xYT media for further propagation. Phage were rescued from XL1-Blue cells harboring the phagemid libraries according to the procedure described by Pluckthun and co-workers (Steiner et al., 2006).

Biotinylation of Pdar and Prb for selections

The AviTag sequence (GLNDIFEAQKIEWHE) was appended onto the N-terminus of the Pdar gene cloned into the pET29 expression plasmid. AviTagged Pdar was expressed from *E. coli* BL21 (DE3), purified using Ni-NTA affinity chromatography, and biotinylated *in vitro* by incubation overnight with 0.1 mol% BirA, 2 equivalents of biotin, and 5 mM MgATP. Excess biotin was removed by dialysis against PBS. AviTagged Prb was biotinylated at its C-terminus in the same manner.

Prb phage library panning (Rounds 1-3)

Three rounds of phage panning against Pdar were performed. For each round, the input phage particles (ranging from 4.5×10^{11} to 1.3×10^{12} colony-forming units) were first incubated in 1-2% BSA/PBS at 30 °C for at least 30 minutes. Biotinylated Pdar was then added to a final concentration of 500 nM and allowed to incubate with the phage for one hour at room temperature. Phage that bound to the biotinylated Pdar ligand were captured in a neutravidin-coated Maxisorp well (Nunc) that was pre-blocked by incubation in 3% BSA/PBS. After 15 minutes, the wells were washed five times with PBS +0.1% Tween and five times with 3% BSA/PBS. The captured phage were eluted from the wells by incubation in 100 mM glycine pH 2.2 for 10 min, and the eluate was immediately neutralized with 2M Tris base (5.6 µL Tris per 100 µL of glycine). The neutralized eluate was used to infect a culture of mid-log-phase *E. coli* XL1-Blue for 40 minutes at 37 °C. Following infection, the cells were spread on 2xYT agar supplemented with 1% glucose, 34 µg/mL chloramphenicol, and 12.5 µg/mL tetracycline. After overnight incubation at 37 °C, colonies were scraped into fresh 2xYT media for further propagation. Phage were rescued from a culture of the recovered cells as described by Pluckthun and co-workers (Steiner et al., 2006).

Prb yeast display selections (Rounds 4-7)

After the third round of phage panning, the remaining mutant Prb genes were recovered by PCR. The genes were transformed into *S. cerevisiae* EBY100 together with *NheI*- and *BamHI*-digested pCTCon2 to obtain a starting yeast library of 1.2×10^{7} transformants by gap repair homologous recombination. Yeast display selections were performed as described by Wittrup and co-workers (Chao et al., 2006). Briefly, yeast cells displaying the Prb library were incubated with biotinylated Pdar for one hour at room temperature. The cells were chilled on ice, pelleted by centrifugation, and washed with ice-cold PBS +0.1% BSA. The washed cells were stained at 4 °C with chicken anti-c-myc antibody/Alexa488-conjugated goat anti-chicken antibodies (Invitrogen) to monitor display levels and streptavidin-(R)-phycoerythrin (Jackson Research) to detect Pdar binding. The antibody-stained yeast

libraries were sorted on a MoFlo Legacy Cell sorter (Beckman Coulter**)** in the Harvard University FAS Center for Systems Biology.

Increased selection pressure for improved Pdar binding was applied by successively decreasing the concentration of Pdar employed in the selection and increasing the stringency of the sort gate (see below).

Round	Pdar concentration	Cells sorted	Cells recovered
4	1000 nM	$2.0x10^7$	$3.0x10^5$
5	100 nM	$3.0x10^7$	$1.8x10^7$
6	100 nM	$9.0x10^6$	$3.0\ x10^4$
7	10 nM	$1.0x10^7$	$1.4\ x10^4$

Pdar yeast display selections (Rounds 1-5)

The Pdar gene was randomly mutated using error-prone PCR with mutagenic dNTP analogs to generate 2-11 mutations per gene (Zaccolo et al., 1996). The amplified Pdar gene library was electroporated into EBY100 together with *NheI*- and *BamHI*-digested pCTCon2 yeast display vector to yield a starting yeast library of 2.6×10^8 transformants by gap repair homologous recombination. Sorting of the yeast-displayed Pdar library for improved Prb binding was performed essentially as described above for the Prb library except for the inclusion of negative selections against binding toward the following panel of negative control proteins: the off7 ankyrin repeat protein (Binz et al., 2004) which binds Pdar with weak affinity (I. Chen, D. R. Liu, unpublished results), the Prb Y110A and W88A knockout mutants, and the PH1109 scaffold protein (see below). Biotinylated proteins for negative selections were prepared as described above for Prb and Pdar.

In the first round of sorting, the starting Pdar library was selected for binding to 18 nM of biotinylated Prb that was preloaded onto a streptavidin-(R)-phycoerythrin conjugate to increase the binding signal. In the next two rounds, negative selections against binding toward a panel of control

proteins were performed. Positive selections for improved Prb binding were performed on cells collected and re-grown after the negative selections. In the final two rounds (4 and 5A), simultaneous negative selections against PH1109 scaffold binding and positive selections for Prb binding were performed. Due to the high level of Prb-binding activity observed by Round 4, a parallel round of sorting was performed consisting only of a negative selection against PH1109 scaffold binding (Round 5C).

Round	Prb concentration	Negative selection proteins	Cells sorted	Cells recovered
1	18 nM*		1.0×10^{9}	1.3×10^{7}
2-neg		5 μM each of A,B, C	1.3×10^{8}	6.1×10^{6}
2-pos	1 μM		3.0×10^{7}	3.6×10^{5}
3-neg		220 nM* each of A, B, C, D	1.8×10^{7}	1.3×10^{6}
4	100 nM	5 μM D	1.0×10^{7}	1.5×10^{5}
5A	10 nM	5 μM D	1.0×10^{7}	1.8×10^{4}
5C		7.5 μM D	1.0×10^{7}	3.2×10^{4}

* protein pre-loaded with streptavidin-(R)-phycoerythrin to increase binding signal

Proteins for negative selection:

A- off7 ankyrin repeat

B- Prb Y110A

C- Prb W88A

D- PH1109 scaffold

Measuring K_d by yeast surface display

Dissociation constants (K_d) were determined by the yeast surface titration method reported by Wittrup and co-workers (Chao et al., 2006). Yeast cells displaying Prb or Pdar mutants were treated with varying concentrations of biotinylated protein partner in either PBS +0.1% BSA or 20 mM HEPES pH 7.5 +50 mM potassium glutamate +0.1% BSA. The cells were stained with streptavidin-(R)-phycoerythrin, and the mean phycoerythrin fluorescence intensities of the cells were measured on a BD LSRII cell analyzer. Dissociation constants were determined by non-linear curve fitting of the

fluorescence intensities using a standard binding equation that included a linear term to account for non-specific sticking to the yeast cell surface (Origin). The K_d's determined by yeast surface display have been shown to match well with those measured by other techniques such as surface plasmon resonance (Gai and Wittrup, 2007).

X-ray crystallography

Prb

The gene encoding Prb was subcloned into a pET29b(+) vector (Novagen) using *Nde*I (5') and *Xho*I (3') restriction sites (New England Biolabs) to generate a C-terminal 6X histidine tag expression vector. Prb was expressed in BL21(DE3) cells by addition of 0.5 mM IPTG at 15°C for 14-16 hours with shaking. Cells were harvested by centrifugation in a GS3 rotor at 7000 rpm for 10 minutes at 4°C, resuspended in lysis buffer (50 mM Tris-HCl (pH 7.0), 100 mM NaCl, 10 mM imidazole), and lysed by sonication. The soluble fraction was separated by centrifugation in an SS34 rotor at 18,000 rpm for 35 minutes at 4°C and filtered through at 0.45 micron syringe filter. Prb was purified at room temperature from the filtered soluble fraction by nickel affinity chromatography as described by the manufacturer (Ni-NTA, Qiagen), dialysed into storage buffer (20 mM TrisHCl (pH 7.0), 50 mM NaCl), and fractionated by size exclusion chromatography using a Superdex 200 column.

Crystals of Prb were grown at 10 mg/mL by hanging drop vapor diffusion in 100 mM sodium acetate (pH 4.5), 3.0 M NaCl at 24°C. Crystals were transferred stepwise to the crystallization mother liquor containing glycerol to 30% (v/v) as a cryoprotectant and subsequently flash frozen in liquid nitrogen. Crystallographic data were collected using a rotating anode Micromax HF-007 X-ray generator (Rigaku, Inc.) equipped with a Saturn 994+ CCD detector and scaled using d*TREK (Pflugrath, 1999). Phases for Prb were determined by molecular replacement with PHASER (McCoy et al., 2007) using pdb ID 2D59. Model building was performed using COOT (Emsley and Cowtan, 2004) and refinement was completed using CNS (Brunger et al., 1998) and Refmac5 (Murshudov et al., 1997) while monitoring the R_{free}. The overall geometric quality of the refined model was assessed using PHENIX (Adams et al.).

The coordinates and structure factors of the Prb monomer have been deposited in the RCSB PDB under accession No. 3QA9.

Prb-PdarC5 and PrbPdarC10

Prb and Pdar mutants were produced and purified individually as described for Prb. However, the His tags were removed by TEV protease prior to their stoichiometric coupling. The complex was then isolated following fractionation on a Superdex 200 column. Crystals of the Pdar_PrbC10 complex obtained at 19C° by the sitting drop vapor diffusion technique using a Mosquito robot (TTP LABTECH LIMITED, Melbourn, Royston, U.K.) diffracted at best to 2.0Å resolution. The crystals obtained in the presence of 1mM CoA were grown from a solution of 100mM LiCl, 50mM MES (pH 6.0), 20% polyethylene glycol 6000, and 10mM Sarcosine. The crystals formed in the P1 space group with cell constants a=53.12 Å, b=56.56 Å, c=56.98 Å, α=89.78º, β=112.42º, and γ=90.03º, and contained two monomers in the asymmetric unit cell with a Vm of 2.4 $Å^3$/Da. A complete data set up to 2.0 Å was collected at the European Synchrotron Radiation Facility (ESRF) beam line, ID14-1. Crystals of the Pdar_PrbC5 complex obtained at 19C° by the sitting drop vapor diffusion diffracted at best to 2.2Å resolution. The crystals obtained in the presence of 1mM CoA were grown from a solution of 100mM NaF, 20% polyethylene glycol 3350, and 0.5% ethyl acetate. The crystals formed in the P1 space group with cell constants a=53.90 Å, b=57.62 Å, c=58.23 Å, α=89.96º, β=90.14º, and γ=113.44º, and contained two monomers in the asymmetric unit cell with a Vm of 2.4 $Å^3$/Da. A complete data set up to 2.2 Å was collected at the European Synchrotron Radiation Facility (ESRF) beam line, ID23-2. The diffraction of both data sets were indexed and integrated using the program, HKL2000 (Otwinowski, 1997). Integrated intensities were scaled using the program SCALEPACK (Otwinowski, 1997). The structure factor amplitudes were calculated using TRUNCATE from the CCP4 program suite. The Pdar_PrbC10 and the Pdar_PrbC5 complex structures were each solved by molecular replacement using the known structure of the hypothetical protein from Pyrococcus horikoshii OT3 (PDB code 2D59) which has high sequence homology to Prb and the structure of artificial Ankyrin repeat protein (PDB code 1MJ0) which has high sequence homology to Pdar as starting models. The refinement was carried out using the program,

CCP4/Refmac5 (Murshudov et al., 1997) and phenix.refine (Adams et al.). The model was rebuilt on the basis of the electron density maps ($2F_{obs}-F_{calc}$ and $F_{obs}-F_{calc}$) using the program COOT (Emsley and Cowtan, 2004). Water molecules were built into peaks greater than 3σ in the $F_{obs}-F_{calc}$ maps. The details of the refinement statistics of the Pdar_PrbC10 and the Pdar_PrbC5 complex structures are presented in Supplemental Table S3. The coordinates and structure factors for the Pdar_PrbC10 and the Pdar_PrbC5 structures have been deposited in the RCSB PDB under accession No. 3Q9N and 3Q9U respectively.

Supplemental References

Acton, T.B., Gunsalus, K.C., Xiao, R., Ma, L.C., Aramini, J., Baran, M.C., Chiang, Y.W., Climent, T., Cooper, B., Denissova, N.G., *et al.* (2005). Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. Method Enzymol *394*, 210-243.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D *66*, 213-221.

Amstutz, P., Binz, H.K., Parizek, P., Stumpp, M.T., Kohl, A., Grutter, M.G., Forrer, P., and Pluckthun, C. (2005). Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. J Biol Chem *280*, 24715-24722.

Bartels, C., Xia, T.H., Billeter, M., Guntert, P., and Wuthrich, K. (1995). The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. J Biomol Nmr *6*, 1-10.

Batchelor, A.H., Piper, D.E., de la Brousse, F.C., McKnight, S.L., and Wolberger, C. (1998). The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA. Science *279*, 1037-1041.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. Nat Biotechnol *22*, 575-582.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M., and Wittrup, K.D. (2006). Isolating and engineering human antibodies using yeast surface display. Nat Protoc *1*, 755-768.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. J Biomol Nmr *6*, 277-293.

Desrosiers, D.C., and Peng, Z.Y. (2005). A binding free energy hot spot in the ankyrin repeat protein GABPbeta mediated protein-protein interaction. J Mol Biol *354*, 375-384.

Dunbrack, R.L., Jr. (2002). Rotamer libraries in the 21st century. Curr Opin Struct Biol *12*, 431-440.

Dunbrack, R.L., Jr., and Cohen, F.E. (1997a). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci *6*, 1661-1681.

Dunbrack, R.L., Jr., and Cohen, F.E. (1997b). Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci *6*, 1661-1681.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Cryst D *60*, 2126-2132.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry *33*, 5984-6003.

Gai, S.A., and Wittrup, K.D. (2007). Yeast surface display for protein engineering and characterization. Curr Opin Struct Biol *17*, 467-473.

Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C.A., and Baker, D. (2003a). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol *331*, 281-299.

Gray, J.J., Moughon, S.E., Kortemme, T., Schueler-Furman, O., Misura, K.M., Morozov, A.V., and Baker, D. (2003b). Protein-protein docking predictions for the CAPRI experiment. Proteins *52*, 118-122.

Jansson, M., Li, Y.C., Jendeberg, L., Anderson, S., Montelione, G.T., and Nilsson, B. (1996). High-level production of uniformly N-15- and C-13-enriched fusion proteins in Escherichia coli. J Biomol Nmr *7*, 131-141.

Kohl, A., Amstutz, P., Parizek, P., Binz, H.K., Briand, C., Capitani, G., Forrer, P., Pluckthun, A., and Grutter, M.G. (2005). Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. Structure *13*, 1131-1141.

Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Pluckthun, A., and Grutter, M.G. (2003). Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci U S A *100*, 1700-1705.

Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A *99*, 14116-14121.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364-1368.

Lazaridis, T., and Karplus, M. (1999). Effective energy function for proteins in solution. Proteins *35*, 133-152.

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006). SMART 5: domains in the context of genomes and networks. Nucleic Acids Res *34*, D257-D260.

Liu, G.H., Shen, Y., Atreya, H.S., Parish, D., Shao, Y., Sukumaran, D.K., Xiao, R., Yee, A., Lemak, A., Bhattacharya, A., *et al.* (2005). NMR data collection and analysis protocol for high-throughput protein structure determination. Proc Natl Acad Sci U S A *102*, 10487-10492.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. Protein Sci *13*, 1435-1448.

Moseley, H.N.B., Monleon, D., and Montelione, G.T. (2001). Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B *339*, 91-108.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst D *53*, 240-255.

Otwinowski, Z.a.M., W (1997). Processing of X-ray data collected in oscillation mode. Method Enzymol *276*, 307-326.

Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. Acta Cryst D *55*, 1718-1725.

Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R., *et al.* (2003). Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. Proteins *52*, 107-112.

Sheffler, W., and Baker, D. (2009). RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. Protein Sci *18*, 229-239.

Shen, Y., Atreya, H.S., Liu, G.H., and Szyperski, T. (2005). G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. J Am Chem Soc *127*, 9085-9099.

Steiner, D., Forrer, P., Stumpp, M.T., and Pluckthun, A. (2006). Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. Nat Biotechnol *24*, 823-831.

Wishart, D.S., and Sykes, B.D. (1994). The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. J Biomol Nmr *4*, 171-180.

Zaccolo, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J Mol Biol *255*, 589-603.