Cell, Volume *166*

Supplemental Information

Origins of Allostery and Evolvability in Proteins:

A Case Study

Arjun S. Raman, K. Ian White, and Rama Ranganathan

Origins of allostery and evolvability in proteins: Supplementary Information

Arjun S. Raman[1], Kristopher I. White[1] & Rama Ranganathan[1,2*]

[1]The Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390

[2]Departments of Pharmacology and Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390

Contents

I. Supplementary Experimental Procedures 1
A. Global analysis of PDZ ligand specificity. 1
B. Construction of the ligand library 1
C. Expression and purification of PSD95^{pdz3} proteins. 2
D. Crystallization and structure determination of PSD95^{pdz3} variants. 2
E. Computational simulations and codes 3

II. Supplementary Tables 7
A. TableS1: Sequencing statistics of the ligand library selection experiments 7
B. TableS2: Crystallographic data collection and refinement statistics - Part 1 8
C. TableS3: Crystallographic data collection and refinement statistics - Part 2 9
D. TableS4: Crystallization conditions for PSD95^{pdz3} variants 10

References 11

I. SUPPLEMENTARY EXPERIMENTAL PROCEDURES

A. Global analysis of PDZ ligand specificity.

Comprehensive study of PDZ binding specificity is made possible by a modified version of a bacterial two-hybrid system (McLaughlin et al., 2012) in which transcription of the chloramphenicol acetyl transferase (CAT) reporter gene (pZE1RM plasmid, pRM+ promoter, ampicillin resistant) is made quantitatively dependent on the binding between a PDZ domain (fused to the pRM+ promoter-binding bacteriophage λ-c1 DNA binding domain, pZS22 plasmid, IPTG inducible, trimethoprim (trm) resistant) and its ligand (fused to the N-terminal domain of *E.coli* RNA polymerase α subunit, pZA31 plasmid, anhydrotetracycline (aTC) inducible, kanamycin resistant) (Fig. S1). Electrocompentent MC4100-Z1 cells containing pZE1RM-CAT and pZS22-PDZ3 variant plasmids were transformed with 1 µl of 20 ng/µl pZA31-RNAα-ligand library (see below), recovered for one hour in LB media, grown in 20 µg/ml trm, 50 µg/ml kan, 100 µg/ml amp to OD_{550} of 0.04, and induced using 50 ng/ml doxycyline plus antibiotics for 3 hours to an OD_{550} of 0.1. 10ml of the induced culture was used to innoculate 100mL LB + antibiotics as above for selection; the remainder was reserved as the pre-selection population for deep sequencing. Selection was carried out with 150 µg/ml chloramphenicol for 6 hours (taking care that $OD_{550} \leq 0.1$), washed in LB medium, and grown overnight at 37°C. Both pre- and post-selection cultures were subject to plasmid DNA isolation, PCR amplification of the ligand region of pZA31, and standard preparation for Illumina Hi-Seq 2500 sequencing (UT Southwestern genomics core). Sequencing data were analyzed using home-written codes and MATLAB (Mathworks Inc., provided upon request) and used to compute $\Delta E_x = \log(f_x^s/f_x^u) - \log(f_o^s/f_o^u)$, the enrichment of each ligand x in the selected (s) and unselected (u) libraries relative to a reference sequence o with similar affinity for each PDZ variant. The reference sequence was CRIPT for wild-type and G330T variants, and T-2F for H372A and the double mutant variants.

B. Construction of the ligand library

The library of PDZ ligands (randomized in the C-terminal four amino acid positions, total theoretical library complexity $20^4 = 160,000$) was generated as C-terminal fusions with the N-terminal domain of *E.coli* RNA polymerase α subunit. The library was made using NNS oligonucleotide-directed mutagenesis with a pZA31-RNAα template containing a non-binding PDZ ligand (N-TKNYKQGGG-COOH) to eliminate background binding. Two oligonucleotides (one sense, one antisense) were synthesized (IDT) with each sequence complementary to 15 base-pairs (bp) on either side but with one oligo containing four consecutive NNS codons at the target positions; N is a mixture of A, T, C, and G, and S is a mixture of G and C. This results in 32 codons at each position encoding all 20 amino acids. The

oligos incude a type IIs restriction site (BsaI), designed to optimize cloning efficiency by enabling a unimolecular ligation protocol. We carried out a single round of PCR, amplifying the entire plasmid while encoding the full library of ligand sequences. This product was subsequently restricted with BsaI, subject to a unimolecular ligation reaction (1 ml, incubated overnight at 16°C), and purified into a final volume of 10 µl (Zymo purification kit). Ten individual transformations into MaxDH10B *E. coli* (Invitrogen) were made, grown overnight after recovery, and plasmid DNA prepped so as to minimize any possible bottlenecking effect. Transformation of the final library into MC4100-Z1 cells for selection yielded greater than 10^8 transformants, and a near complete representation of the theoretical complexity (Table S1).

C. Expression and purification of PSD95^{pdz3} proteins.

pGEX-4T-1 plasmids containing Glutathione-S-transferase (GST)-fusions of wild-type or mutant PSD95^{pdz3} (amino acid range 297-415) were transformed into *Escherichia coli* BL21(DE3) cells and grown overnight on LB plus 100 μg/mL ampicillin (amp) plates. Streaks of colonies were used to start overnight cultures (LB + amp), used to innoculate 1L cultures (Terrific Broth or ZYM-5052 auto-inducing medium (Studier, 2005) + 100 µg/mL amp). Cultures were grown to an optical density (600 nm) of 0.6-0.8 at 37 ^{o}C, induced overnight at 18 ^{o}C (supplemented with 1 mM isopropyl--D-thiogalactopyranoside if manual induction), and then harvested by centrifugation. Pellets were resuspended in lysis buffer composed of phosphate-buffered saline (PBS; 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) supplemented with 1% glycerol, 1 mg/ml hen egg white lysozyme, 1 mM dithiothreitol (DTT), EDTA-free protease inhibitor cocktail (Roche). The cell suspension was subjected to sonication and centrifugation, and clarified lysate was then incubated with glutathione sepharose 4B resin (GE Healthcare). Bound protein was washed with PBS supplemented with 1% glycerol and 1 mM DTT, and the GST tag was cleaved through bovine thrombin (Calbiochem) proteolysis overnight at room temperature in PBS supplemented with 10% glycerol and 1 mM DTT. Thrombin was removed by benzamidine sepharose (GE Healthcare) and the PDZ domain was purified to near-homogeneity using a Source 15Q anion exchange (GE Healthcare) column employing a linear gradient from low salt (20 mM Tris HCl pH 7.5, 1% glycerol, 1.0 mM DTT) to high salt (20 mM Tris HCl pH 7.5, 1 M NaCl, 1% glycerol, 1 mM DTT). The protein was dialyzed into 10 mM HEPES pH 7.2, 10 mM NaCl, concentrated and subject to size-exclusion chromatography (Superdex 75, GE Healthcare). Peak fractions were pooled, concentrated to 35 mg/mL, and subsequently either flash frozen in liquid N_2 for storage at -80°C or used immediately for crystallization. Substrate peptides for co-crystallization (CRIPT (Acetyl-TKNYKQTSV-COOH), T-2F (Acetyl-TKNYKQFSV-COOH)) were synthesized using standard FMOC chemistry (UTSW Proteomics Core Facility), HPCL purified, and lyophilized.

D. Crystallization and structure determination of PSD95^{pdz3} variants.

Crystallization of PSD95^{pdz3} variants was performed by the vapor diffusion hanging drop method. In all cases, purified protein was diluted to a final concentration of $\sim$ 9 mg/ml in protein buffer (10 mM HEPES pH 7.2, 10 mM NaCl). Where applicable, peptide was included in protein buffer to a final molar ratio of 2:1 relative to protein. Reservoir solutions typically contained 1 M sodium citrate, pH 7.0; specific crystallization conditions for each mutant are shown in Table S4. Equal amounts (1.5 µl) of protein and reservoir solution were mixed and equilibrated against 500 µl of crystallization buffer at 16°C. Diamond-shaped crystals appeared either spontaneously or with microseeding after 1–5 days and grew to 100–200 µm in length over several weeks. To prepare microseeding solutions, wild-type crystals of the appropriate state were crushed and resuspended in crystallization buffer. Single crystals were cryoprotected by serial equilibration into crystallization buffer with increasing amounts of glycerol (up to 25%) and flash frozen in liquid N_2.

Diffraction data were collected at 100 K at either at the UT Southwestern structural biology laboratory or at the Advanced Photon Source (Argonne National Laboratory, 19-ID) and indexed and scaled in HKL-2000 (Otwinowski and Minor, 1997) (HKL Research). Resolution cutoffs were chosen based on I/σ and CC 1/2 (Tables S2-S3). Phasing and automated refinement was carried out using PHENIX (Adams et al., 2010) with manual modeling in COOT (Emsley et al., 2010); the data collection and refinement statistics are summarized in Tables S2-S3. An initial model was obtained from rigid body and temperature factor refinement using published structures of PSD95^{pdz3} (PDB 1BFE and 1BE9, with ligand removed), and subject to 0.5 Å coordinate randomization followed by Cartesian simulated annealing to help reduce phase bias. Further computational refinements steps involved iterative rounds of positional and temperature factor minimization, manual model building, solvent placement, and TLS refinement, guided by decrease in crystallographic R-factors. Figures were prepared with PyMol (DeLano, 2002). The atomic coordinates and structure factors have been deposited in the Protein Data Bank with the following acces-

sion numbers: 5HEB (PSD95^{pdz3}(WT)-CRIPT), 5HED (PSD95^{pdz3}(WT)-T-2F), 5HET (PSD95^{pdz3}(G330T)-apo), 5HEY (PSD95^{pdz3}(G330T)-CRIPT), 5HF1 (PSD95^{pdz3}(G330T)-T-2F), 5HFB (PSD95^{pdz3}(H372A)-CRIPT), 5HFC (PSD95^{pdz3}(H372A)-T-2F), 5HFF (PSD95^{pdz3}(G330T, H372A)-T-2F).

E. Computational simulations and codes

The model shown in Fig. 4 and Fig. S3 simulates the dynamics of a constant sized population (here, $N = 1000$, large relative to the number of genotypes) comprising the four PDZ variants (wild-type, G330T, H372A, and the double mutant) with mutation and selection under a fluctuating condition of fitness. In each generation, single mutations occur with probability μ, double mutations with probability μ^2, and selection re-draws the frequency of each genotype according to its ability to bind ligand relative to all other available genotypes. The fractional binding of each genotype is determined from the experimentally measured equilibrium dissociation constants (Fig. 1B), and the ligand (CRIPT or T-2F) switches every τ generations. For each trial of switching from CRIPT to T-2F in which the double mutant ultimately goes to fixation in the population, we compute the fraction of G330T and H372A in the interval from the switch to fixation of the double mutant; limits for integration were automatically determined by empirical fitting of the probability density of the double mutant in each trial (see codes). The data in Figs. 4C-F were obtained from ∼500 trials of CRIPT to T-2F ligand switching each. The codes were executed using a custom shell script on a high-performance LINUX cluster (BioHPC, UT Southwestern Medical Center).

```
function [out] = evSim_RWR(pinit, mu, tau, nswitches,Kds,switch_mode)
% [sim]=evSim_RWR([1000;0;0;0],mu,tau,nswitches, switch_mode,f);

% The population dynamics model in Raman et al, "Origins of allostery and
% evolvability in proteins: a case study".  This function models the
% dynamics of population shifts in WT, G330T, H372A, and the double mutant
% given an initial population structure, mutation rate, ligand switching
% time, and other parameters, described below.

%  Inputs:

%   (1) "pinit" is the initial population vector, in order wild-type,
%   G330T, H372A, and the double mutant. For example, pinit=[1000;0;0;0] to
%   start with 1000 WT individuals. sum(pinit) gives the total population
%   size, N. In paper pinit = [1000;0;0;0].

%   (2) "mu" is the mutation rate - the probability of a single mutation at
%   each generation.  So, for example, mu=0.001 gives N*mu=1. The paper
%   describes simulations at mu = 0.0001, 0.001, and 0.01.

%   (3) "tau" is the wait time in generations for ligand switching. So,
%   tau=100 means the ligand switches every 100 generations.

%   (4) "nswitches" is total number of ligand switches, and so
%   nswtiches*tau is the total number of generations simulated.
%
%   (5) "Kds" is a 2 X 4 matrix of equilbrium dissociation constants,
%   assumed in units of micromolar. The columns correspond to WT, G330T,
%   H372A, G330T,H372A in order, and rows to CRIPT or T-2F ligand in
%   order. Values are given in Fig. 1B. Kds(:,1)=[0.8 2.2 26.9 22.1] and
%   Kds(:,2)=[36 1.8 1.9 0.5].

%   (5) "switch_mode" is a flag that determines whether ligand switching is
%   regular (every tau generations, switch mode=0), or is Poisson
%   distributed with a mean wait time of tau generartions
%   (switch_mode~=%0).  Default switch_mode=0.

% Outputs:

% "out" is a structure with two fields...out.input has all the inputs and
% out.output has "P", the population vector at each generation,
% "gen_switch", the generation at which each ligand switch happens
% (always starts with CRIPT).

% ****************************************

% ****************************
% Step 1: Set Inputs
% ****************************
if nargin>6
    switch_mode=0;
end
% Determine fractions bound from the input dissociation constants (from
% experiments in Fig. 1, in units of micromolar).  Ligand concentration is
% fixed at 10 micromolar, as explained in the main text.
```

```
56
57 if size(Kds,1)==2; Kds=Kds';end % in case Kds matrix needs transposition
58 f=10./(10+Kds);
59
60 % ******************************
61 % Step 2: Initialization
62 % ******************************
63
64 len=(nswitches+1)*tau; % set number of generations
65
66 % Defines svect, which stores the generations at which ligands switch.
67 % Regular switching every tau generations if switch_mode=0 and Poissonian
68 % switching otherwise.
69 svect=zeros(1,len);
70 if switch_mode==0
71     svect([1:tau:len])=1;svect(1)=0;
72 else
73     draws=round(exprnd(tau,1,nswitches));
74     index=0;
75     for i=1:numel(draws)
76         svect(draws(i)+index)=1;
77         index=index+draws(i);
78     end
79     svect(1)=0;
80 end
81 % set initial ligand environment, 1 is CRIPT, -2 is T-2F
82 env=1;sel=1;
83
84 % initialize population vector (numbers of each species, a 4 X 1 vector)
85 % and individual vector (the actual members of the population, labelled
86 % according to type, a 1000 X 1 vector).
87 P=zeros(4,len);
88 P(:,1)=pinit;
89 psize=sum(pinit); % total population size
90 labels = [0 1 2 3]; % the genotypes (labels) in decimal. 0-WT, 1-G330T, 2-H372A, 3-the double mutant
91 f_vect=pinit./sum(pinit); % frequncies
92 i_vect=(labels*mnrnd(1,f_vect,psize)')'; % make initial individual vector from frequencies (in decimal)
93 m_tmp=dec2bin(i_vect(:,1),2); % binary version of the population vector
94 m_vect=[str2num(m_tmp(:,1)) str2num(m_tmp(:,2))]; %change from char to num. This is the initial popoulation vect
       in binary
95
96 % ********************************
97 % Step 3: The simulation
98 % ********************************
99
100 % The idea is to pick an environment (that defines the selections),
101 % randomly draw single and double mutations using binary representation,
102 % convert to decimal population, apply selection to adjust frequencies,
103 % redraw population, convert to binary representation, and iterate.  Very
104 % likely more efficient ways to do this...
105
106 for i=2:len
107     if svect(i)==1
108         env=-env; % switch ligands if at the generation specified in svect
109         sel=sel-env;
110     end
111     fit_vect=(f(:,sel))./(sum((P(:,i-1)./psize).*f(:,sel))); % convert fraction bound to relative fitness
112     m_vect=abs((poissrnd(mu,2,psize)~=0)'- m_vect); % apply random mutation given mu
113     p_vect=bin2dec(num2str(m_vect)); % switch to decimal
114     P_tmp=[numel(find(p_vect==0));numel(find(p_vect==1));numel(find(p_vect==2));numel(find(p_vect==3))]; % get
          population counts after mutation
115     f_vect=(fit_vect.*P_tmp)./sum(fit_vect.*P_tmp); %get frequencies with applying selection on species in
          population
116     i_vect=(labels*mnrnd(1,f_vect',psize)')'; % generate new population, now after mutation and selection
117     m_tmp=dec2bin(i_vect,2); % back to binary matrix
118     m_vect=[str2num(m_tmp(:,1)) str2num(m_tmp(:,2))]; % switch from char to num
119     P(:,i)=[numel(find(i_vect==0));numel(find(i_vect==1));numel(find(i_vect==2));numel(find(i_vect==3))]; %write
          new population vector after mutation and selection
120 end
121
122 % ********************************
123 % Step 4: Make output structure
124 % ********************************
125
126 % first, we determine the generations at which CRIPT changes to T-2F (every
127 % other one)
128 gen_switch_either=find(svect);
129 gen_switch=gen_switch_either(1:2:numel(gen_switch_either));
130 n_trials=numel(gen_switch);
131
132 % the output structure
133 out.input.pinit = pinit;
134 out.input.mu = mu;
```

```
out.input.tau = tau;
out.input.nswitches = nswitches;
out.input.f = f;
out.input.switch_mode = switch_mode;
out.output.P = P;
out.output.gen_switch = gen_switch;

% ************************************
% Step 5: Analysis
% ************************************
% This is optional, and is best avoided for high-throughput study of many
% conditions.  Comment out below if not desired.

% a plot of the population dynamics over the simulation length
h_sim=figure;clf;hold on;grid on;
plot(P(1,:),'k','LineWidth',1.5);
plot(P(2,:),'g','LineWidth',1.5);
plot(P(3,:),'r','LineWidth',1.5);
plot(P(4,:),'b','LineWidth',1.5);
plot(svect*psize,'--k');
hold off;

% analysis of the simulation. Here, we extract the population dynamics
% following each trial of ligand switching (CRIPT to T-2F), isolate the
% events in which the double mutant goes to near fixation, determine the
% inteval over which the intermediate genotypes should be intergrated, and
% compute the fractional flux through the G330T state.

% pre-allocate variables
events_post=zeros(4,tau,numel(gen_switch)); % population dymamics in each trial
lim=zeros(1,numel(gen_switch)); % limit for integration in each trial
Npost=zeros(2,numel(gen_switch)); % integrated counts of G330T and H372A per trial
frac_G330T=zeros(1,numel(gen_switch)); % fraction of G330T per trial
count_double=0;

for k=1:numel(gen_switch)
    events_post(:,:,k)=P(:,gen_switch(k):gen_switch(k)+(tau-1));
    if max(events_post(4,:,k))>800 % minimal double mutant level to consider trial (arbitrary)
        count_double=count_double+1;
        ydat=smooth(diff(events_post(4,:,k)),20);
        xdat=[1:numel(ydat)];
        try
            pd=fit(xdat',diff(events_post(4,:,k))','gauss1'); % Guassian fit of double mut population
            lim(k)=pd.b1+(2*pd.c1); % limit for integrating G330T, H372A populations
            if lim(k)<tau
                Npost(1,k)=sum(events_post(2,1:floor(lim(k)),k),2); %number G330T
                Npost(2,k)=sum(events_post(3,1:floor(lim(k)),k),2); %number H372A
                frac_G330T(k)=Npost(1,k)./(Npost(1,k)+Npost(2,k)); % the fraction G330T
            else
                frac_G330T(k)=-Inf; % if limit is not less than tau
            end
        catch
            frac_G330T(k)=-Inf; % if fitting throws an exception
        end
    end

end

% clean up events for exceptions
ind=find(~isinf(frac_G330T) & ~isnan(frac_G330T) & frac_G330T~=0 & frac_G330T~=-Inf); % the indices of valid
    trials
frac_G330T_clean=frac_G330T(ind); % clean frac_G330T for valid switching trials
lim_clean=lim(ind); % clean lim for valid switching trials
mean_fracG330T=mean(frac_G330T_clean); % mean fraction G330T over simulation
sem_fracG330T=std(frac_G330T_clean)/sqrt(numel(ind)); % standard error of the mean, since we want confidence in
    the mean value, and not the scatter over trials.

% plotting
h_analysis=figure;clf;
plot(frac_G330T_clean,'ok-','LineWidth',1.5);hold on;grid on
plot([0 numel(gen_switch)],[mean_fracG330T mean_fracG330T],'--or','LineWidth',1);
if ~isempty(frac_G330T_clean)
    text(1,(0.05*max(abs(frac_G330T_clean))),['fraction G330T = ' num2str(mean_fracG330T)],'FontWeight','bold','
        FontSize',12);
    axis([0 numel(gen_switch) 0 1]);
end
P_double=count_double./numel(gen_switch);
[mean_fracG330T sem_fracG330T P_double]

% ************************************
% Step 6: Add analysis to output structure
% ************************************

```

```
215  out.analysis.frac_G330T=frac_G330T_clean;
216  out.analysis.index_valid=ind;
217  out.analysis.integration_limits=lim_clean;
218
219  end
```

II. SUPPLEMENTARY TABLES

Unselected Library Statistics

	Number of reads	Number of Ligands (> 50 counts)	% Library Coverage
Total Input Library	1.07 x 10^8	154,521	96.7

Selected Library Statistics

	Number of reads	Number of Ligands	Number of Ligands (> 50 counts)	Number of Ligands Bound by Protein (> 15 μM)
WT	46,598,840	56,640	55,278	185
G330T	51,195,397	90,735	83,056	846
H372A	29,419,042	59,935	43,488	1051
H372A/ G330T	48,989,769	123,295	86,255	1894

TABLE S1 **Sequencing statistics of the ligand library selection experiments**. Related to Figures 2-3. Sequencing statistics from Illumina HiSeq2500 runs for the unselected and selected populations of peptide libraries for wild-type, G330T, H372A, and the double mutant experiments. The unselected populations were combined over all experiments. The total number of reads, (1.07×10^8) represented approximately 97% coverage of all peptides in the library.

	WT-CRIPT	WT-T$_2$F	G330T-Apo	G330T-CRIPT
Data Collection				
PDB ID	5HEB	5HED	5HET	5HEY
Source	UTSW SBL	UTSW SBL	APS 19-ID	APS 19-ID
Wavelength (Å)	1.54178	1.54178	0.97937	0.97937
Resolution range (Å)	40.15–1.65 (1.709–1.65)	31.66–1.7 (1.761–1.7)	36.52–2.001 (2.073–2.001)	27.21–1.5 (1.554–1.5)
Space group	$P4_132$	$P4_132$	$P4_132$	$P4_132$
Unit cell (Å, º)	89.771 89.771 89.771 90 90 90	89.551 89.551 89.551 90 90 90	89.445 89.445 89.445 90 90 90	90.23 90.23 90.23 90 90 90
Total reflections	209981	128220	87993	454547
Unique reflections	15083 (1156)	13972 (1299)	8697 (805)	20265 (1645)
Multiplicity	13.9 (2.5)	9.2 (3.0)	10.0 (8.4)	21.9 (14.7)
Completeness	0.98	0.99	0.99	0.98
Mean I/σ	61.571 (2.478)	60.017 (2.627)	35.095 (2.077)	30.057 (1.448)
Wilson B-factor	14.71	15.35	25.36	14.99
R-merge	0.034 (0.289)	0.031 (0.302)	0.064 (N/A)	0.110 (N/A)
R-meas	0.034 (0.353)	0.033 (0.361)	0.068 (N/A)	0.112 (N/A)
R-pim	0.008 (0.199)	0.010 (0.193)	0.022 (0.397)	0.035 (0.665)
CC1/2	0.87	0.885	0.713	0.555
Refinement				
Reflections used in refinement	15082 (1156)	13971 (1299)	8696 (805)	20265 (1645)
Reflections used for R-free	1509 (116)	1398 (130)	871 (80)	2024 (165)
R-work	0.1666 (0.2273)	0.1701 (0.2291)	0.1837 (0.2309)	0.1724 (0.2705)
R-free	0.1945 (0.2745)	0.1973 (0.2694)	0.2130 (0.2586)	0.2026 (0.2765)
Number of non-hydrogen atoms	1220	1262	1054	1268
Macromolecules	1086	1110	956	1121
Ligands	12	N/A	N/A	N/A
Protein residues	126	126	118	127
RMS (bonds)	0.007	0.014	0.006	0.011
RMS (angles)	1.22	1.46	0.83	1.84
Ramachandran favored (%)	97	99	98	95
Ramachandran allowed (%)	2.1	0.72	1.6	4.2
Ramachandran outliers (%)	0.71	0	0	0.69
Rotamer outliers (%)	2.6	3.4	2.9	3.3
Clashscore	3.19	7.29	1.06	4.02
Average B-factor	19.78	20.48	32.04	22.33
Macromolecules	18.51	19.21	31.36	21.16
Ligands	62.39	N/A	N/A	N/A
Solvent	26.84	29.79	38.7	31.27
Number of TLS groups	3	3	1	16

TABLE S2 **Crystallographic data collection and refinement statistics - Part 1**. Related to Figure 5. All proteins were crystallized in the same space group ($P4_132$) and showed unit cell constants within 0.5% of each other. Ligand-bound datasets were collected to a resolution higher than 2.0 Å.

	G330T-T$_2$F	H372A-CRIPT	H372A-T$_2$F	G330T-H372A-T$_2$F
Data Collection				
PDB ID	5HF1	5HFB	5HFC	5HFF
Source	APS 19-ID	UTSW SBL	UTSW SBL	APS 19-ID
Wavelength (Å)	0.97918	1.54178	1.54178	0.97918
Resolution range (Å)	40.13–1.747 (1.81–1.747)	40.07-1.617 (1.675-1.617)	36.52–1.851 (1.918–1.851)	36.57–1.749 (1.812–1.749)
Space group	$P4_132$	$P4_132$	$P4_132$	$P4_132$
Unit cell (Å, º)	89.724 89.724 89.724 90 90 90	89.592 89.592 89.592 90 90 90	89.445 89.445 89.445 90 90 90	89.575 89.575 89.575 90 90 90
Total reflections	358023	203441	68030	264373
Unique reflections	13007 (1217)	15108 (556)	10759 (914)	12879 (1179)
Multiplicity	27.3 (26.3)	13.3 (1.6)	6.3 (2.5)	20.3 (20.1)
Completeness	0.99	0.93	0.98	0.99
Mean I/σ	67.804 (1.844)	65.515 (0.936)	35.543 (1.000)	50.222 (1.826)
Wilson B-factor	19.04	15.74	21.65	18.64
R-merge	0.055 (N/A)	0.035 (0.338)	0.040 (0.448)	0.063 (N/A)
R-meas	0.056 (N/A)	0.036 (0.470)	0.044 (0.547)	0.065 (N/A)
R-pim	0.013 (0.581)	0.009 (0.325)	0.017 (0.304)	0.016 (0.478)
CC1/2	0.721	0.695	0.689	0.706
Refinement				
Reflections used in refinement	13006 (1217)	15109 (556)	10759 (914)	12880 (1179)
Reflections used for R-free	1301 (122)	1513 (56)	1078 (92)	1285 (114)
R-work	0.1993 (0.2911)	0.1751 (0.3004)	0.1835 (0.2894)	0.1740 (0.2314)
R-free	0.2179 (0.3097)	0.2033 (0.3281)	0.2051 (0.3146)	0.2233 (0.2761)
Number of non-hydrogen atoms	1182	1204	1161	1191
Macromolecules	1080	1055	1037	1080
Ligands	N/A	N/A	N/A	12
Protein residues	124	123	124	128
RMS (bonds)	0.006	0.006	0.012	0.011
RMS (angles)	0.94	0.95	1.46	1.27
Ramachandran favored (%)	99	99	95	97
Ramachandran allowed (%)	0.73	0.73	4.5	0.71
Ramachandran outliers (%)	0	0	0.76	2.1
Rotamer outliers (%)	4.4	13.3 (1.6)	2.7	6.2
Clashscore	4.63	1.91	10.18	4.16
Average B-factor	28.71	22.31	30.76	25.02
Macromolecules	27.95	20.99	30.12	24.07
Ligands	N/A	N/A	N/A	57.25
Solvent	36.72	31.72	36.13	31.53
Number of TLS groups	5	5	8	7

TABLE S3 **Crystallographic data collection and refinement statistics - Part 2**. Related to Figure 5. All proteins were crystallized in the same space group ($P4_132$) and showed unit cell constants within 0.5% of each other. Ligand-bound datasets were collected to a resolution higher than 2.0 Å.

Protein	[sodium citrate] (M)	crystallization buffer pH	[protein] (mg/ml)
WTapo	1.0	6.9	9
WT,CRIPT	1.0	7.0	9
WT,T-2F	1.125	7.1	9
G330Tapo	1.05	7.0	7
G330T,CRIPT	1.2	7.0	8
G330T,T-2F	1.2	6.8	9
H372Aapo	0.95	7.0	9
H372A,CRIPT	1.05	7.0	7
H372A,T-2F	1.05	7.0	7
G330T,H372Aapo	1.0	7.0	13
G330T,H372A,CRIPT	1.25	7.0	9
G330T,H372A,T-2F	1.2	6.75	7

TABLE S4 **Crystallization conditions for PSD95^{pdz3} variants**. Related to Figure 5. Details of protein expression, purification, and general aspects of crystallization are given in the methods section.

References

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213–21.

DeLano, W. L. (2002). The PyMol molecular graphics system.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486–501.

McLaughlin, Jr, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S. and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–42.

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Methods in Enzymology, (Carter, Jr, C. W. and Sweet, R. M., eds), vol. 276, pp. 307–326. Academic Press New York macromolecular crystallography, part A edition.

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*, 207–34.