Supporting Information

S1. Computational cost of each step in binding free energy calculation.

S2. Details of constructing 2-dimensional PMF.

S3. The detailed decomposition of the total binding free energy into the contribution from each stage for the 25 SH2 domain-phosphopeptide pairs.

S4. Structure database of peptide motifs that are known to bind SH2 domain.

S5. Comparison of PMF graphs before and after introducing the second distance for Stat1 binding free energy calculations.

S6. Comparison of sampling before and after introducing the second variable d2, where d1 is the center-of-mass distance between the SH2 domain and peptide and d2 the distance between the phophate group and phosphate binding residues.

S7. Superposition of structures containing the same SH2 domain and different peptides.

S8. Superposition of structures containing the same peptide motif and different SH2 domains.

I. DETAILS OF 2-DIMENSIONAL PMF

We first calculated the 1-dimensional PMF as a function of the SH2 domain-phosphopeptide center-of-mass (COM) distance using 31 umbrella sampling windows. The distance between two neighboring windows is 1 Åand the entire umbrella sampling span from 10 to 40 Å. Because of the strong electrostatic interactions between the phosphate group and the negatively charged residues in the phosphate binding site, there is usually no adequate sampling for the regions where the peptide is about to leave the binding site, as shown in Figure 1(a). In order to sample these regions, the distance between the peptide phosphate group and residues in SH2 domain that are in direct contact with phosphate group is introduced as a second variable to construct a 2-dimensional PMF. Here the Stat1 SH2 domain bound with the native phosphopeptide pYDKP is given as an example to illustrate how the 2-D PMF is constructed. Specifically, the second reaction coordinate r_2 is chosen to be the distance beteen the COM of the phosphate group PO4 and the COM of the sidechains of Lys7, Arg25, Ser27, Glu28 and Ser29, which have direct contact with the phosphate group of the phosphopeptide (Figure 1(b)). For each of the three windows with r_1 distance 24, 25 and 26 Å, 15 umbrella windows were sampled with r_2 distance spanning from 1 to 15 Å, as shown in Figure 3(b). Thus the total umbrella sampling has 73 windows and each lasts for 1 ns.

At the end of the simulation, a 2-dimensional PMF was calculated using the weighted histogram analysis methods (WHAM). By integrating the PMF along the r_2 distance, we obtained the 1-dimensinal PMF as a function of r_1 distance, shown in Figure 2. The free energy change associated with translating the ligand into SH2 domain binding site was calculated from this 1-dimensinal PMF. For comparison, the original 1-dimensinal PMF constructed from 31 windows is also given.

$$e^{-\beta \mathcal{W}(r_1)} = \int e^{-\beta \mathcal{W}(r_2, r_1)} d(r_2) \tag{1}$$

TABLE I: Computational cost of each step in binding free energy calculation.

Computation [a]	Number of windows	Sampling per window (ns)	Rate [b] (hrs/ns)	CPU time (hrs)
Equilibration	1	1	12	144
ΔG_c^{bulk}	32	1	1	384
ΔG_c^{site}	32	1	12	4,608
ΔG_a^{site}	11	1	12	1,584
ΔG_o^{site}	11	1	12	1,584
ΔG_a^{bulk}	52	1	12	9,984
Total per SH2-peptide				18,288

[a]All computations are repeated three times with different initial conditions except for ΔG_a^{bulk} term, which is repeated four times.

[b]Rate is based on running on four 2.2 GHz AMD Opteron 275 processors.

TABLE II: The detailed decomposition of the total binding free energy into the contribution from each stage for the 25 SH2 domain-phosphopeptide pairs (kcal/mol).

SH2	Peptide	ΔG_c^{bulk}	ΔG_o^{bulk}	$\Delta G_a^{\text{bulk},\circ}$	ΔG_a^{site}	ΔG_o^{site}	ΔG_c^{site}	ΔG_{total}
Lck	pYEEI	4.86±0.09	5.55	-16.6±0.8	0.35±0.03	0.87±0.06	1.70±0.13	-9.1±0.8
	pYVNV	4.62±0.42	5.66	-13.4±1.6	0.18±0.01	0.42±0.01	0.80±0.10	-4.5±1.7
	pYTPE	3.54±0.06	5.62	-9.4±1.8	0.22±0.01	0.69±0.07	0.53±0.02	-1.7±1.8
	pYMDM	5.09±0.01	5.64	-11.0±1.5	0.37±0.06	0.85±0.11	0.84±0.13	-2.3±1.5
	pYDKP	3.32±0.03	5.87	-7.8±1.3	0.84±0.04	3.84±0.37	3.25±0.14	-6.5±1.4
	pY	1.60±0.04	5.57	-7.5±1.5	0.37±0.04	1.84±0.09	1.40±0.05	-3.9±1.5
Grb2	pYEEI	2.76±0.05	5.64	-8.1±1.7	0.55±0.02	1.77±0.15	2.47±0.05	-4.5±1.7
	pYVNV	4.41±0.18	5.56	-19.4±1.7	0.15±0.01	0.42±0.01	0.38±0.05	-10.4±1.7
	pYTPE	2.40±0.04	5.66	-7.7±0.6	0.42±0.05	1.46±0.17	1.15±0.05	-2.7±0.6
	pYMDM	3.07±0.06	5.60	-10.5±2.0	0.22±0.01	1.14±0.24	2.39±0.31	-5.6±2.0
	pYDKP	6.67±0.22	5.57	-17.1±1.2	0.33±0.02	0.53±0.02	0.88±0.02	-6.6±1.2

Continued on next page

TABLE II – continued from previous page

SH2	Peptide	ΔG_c^{bulk}	ΔG_o^{bulk}	$\Delta G_a^{\text{bulk},\circ}$	ΔG_a^{site}	ΔG_o^{site}	ΔG_c^{site}	ΔG_{total}
	pY	1.32±0.06	6.05	-8.8±0.6	0.99±0.06	1.47±0.18	0.27±0.01	-4.2±0.6
Cbl	pYEEI	3.27±0.09	5.56	-10.4±0.9	0.31±0.03	0.99±0.03	1.24±0.04	-4.1±0.9
	pYVNV	3.76±0.06	5.56	-16.3±1.2	0.21±0.02	0.55±0.01	0.56±0.15	-8.3±1.2
	pYTPE	2.25±0.06	5.57	-15.1±1.0	0.31±0.07	0.77±0.05	0.81±0.13	-9.2±1.0
	pYMDM	3.01±0.12	5.56	-13.2±1.4	0.46±0.06	1.39±0.07	1.42±0.04	-7.9±1.4
	pYDKP	4.91±0.09	5.65	-10.0±1.3	0.65±0.21	3.51±0.21	3.59±0.52	-7.2±1.4
	pY	1.41±0.04	6.01	-9.2±1.6	0.53±0.01	1.62±0.12	0.38±0.17	-4.3±1.6
p85αN	pYEEI	3.31±0.02	5.57	-11.8±1.1	0.25±0.01	0.57±0.09	0.86±0.03	-4.6±1.1
	pYVNV	4.37±0.26	5.66	-18.9±1.9	0.34±0.08	0.33±0.02	0.13±0.01	-9.7±1.9
	pYTPE	5.95±0.37	5.68	-12.0±1.3	0.39±0.04	1.26±0.21	2.04±0.48	-4.1±1.4
	pYMDM	2.84±0.08	5.56	-11.3±0.7	0.33±0.04	0.91±0.19	0.93±0.12	-5.1±0.7
	pYDKP	2.94±0.18	5.93	-9.6±1.3	1.20±0.46	2.26±0.32	3.33±0.36	-7.5±1.5
	pY	1.33±0.11	6.07	-9.4±1.4	0.49±0.01	1.43±0.09	0.67±0.04	-4.6±1.4
Stat1	pYEEI	3.12±0.03	5.56	-8.4±1.3	0.77±0.11	1.39±0.16	1.54±0.06	-3.4±1.3
	pYVNV	3.51±0.12	5.57	-15.7±1.6	1.20±0.03	1.71±0.04	0.70±0.05	-10.2±1.6
	pYTPE	2.66±0.08	5.61	-7.8±1.8	0.50±0.02	2.46±0.34	1.48±0.05	-4.0±1.8
	pYMDM	3.06±0.05	5.55	-14.8±0.7	0.62±0.06	1.25±0.20	1.67±0.12	-9.7±0.7
	pYDKP	4.48±0.09	5.61	-16.8±1.2	0.67±0.04	0.43±0.01	0.52±0.04	-8.3±1.2
	pY	1.22±0.11	5.78	-9.3±0.3	0.49±0.04	1.57±0.07	0.35±0.06	-4.7±0.3

TABLE III: Structure database of peptide motifs that are known to bind SH2 domain.

	Ligand	PDB(resolution/Å)
1	pYADP	2FCI (NMR)
2	pYANF	1P13 (1.63)
3	pYAQP	1JU5 (NMR)
4	pYAQV	1D4T (1.10), 1D4W (1.80), 1M27 (2.50), 1I3Z (2.15)

Continued on next page

TABLE III – continued from previous page

	Ligand	PDB(resolution/Å)
5	pYDEV	2CIA (1.45), 2CI9 (1.50)
6	pYDKP	1YVL (3.00)
7	pYDVL	2OQ1 (1.90)
8	pYDYV	1X27 (2.70)
9	pYEEG	1LKL (1.80)
10	pYEEI	1LKK (1.00), 1KC2 (2.10), 1SHD (2.00), 2C0I (2.30), 1SPS (2.70), 1HCS (NMR), 1AOT (NMR)
11	pYEEV	1CWD (2.25)
12	pYELL	2HDX (2.35)
13	pYENL	1R1S (1.90)
14	pYEPI	1A81 (3.00)
15	pYEPL	1UUR (2.70), 1UUS (2.80)
16	pYEPP	2ETZ (NMR)
17	pYEpYI	1NZL (1.90)
18	pYESP	2FCI (NMR)
19	pYETD	1RQQ (2.60)
20	pYETL	2CSZ (NMR)
21	pYIIP	2PLD (NMR)
22	pYIKT	1BF5 (2.90)
23	pYINQ	1ZFP (1.80)
24	pYIpYV	1NZV (2.10)
25	pYLDI	1FU5 (NMR)
26	pYLKT	1BG1 (2.25)
27	pYLRV	1SHB (2.00)
28	pYMDM	2IUH (2.00), 1AYC (2.30)
29	pYMPM	1FU5 (NMR)
30	pYNEL	2OQ1 (1.90)
31	pYpYN	1JYQ (2.00)

Continued on next page

TABLE III – continued from previous page

Ligand	PDB(resolution/Å)
32 pYQGL	1TCE (NMR)
33 pYQPG	1FMK (1.50)
34 pYQPQ	1LCK (2.50)
35 pYSFY	1YVH (2.05)
36 pYSGL	1A81 (3.00)
37 pYSTV	2HMH (2.00), 2BBU (NMR)
38 pYTAV	1AYA (2.05)
39 pYTPE	2CBL (2.10), 1FBV (2.90)
40 pYVNI	1AYB (3.00)
41 pYVNQ	1MW4 (NMR)
42 pYVNV	1JYR (1.55), 1R1P (1.80), 1R1Q (1.80), 1BMB (1.80) 1TZE (2.10), 1FYR (2.40), 1QG1 (NMR)
43 pYVPM	1SHA (1.50), 1H9O (1.79), 2IUI (2.40), 1OO4 (NMR), 1PIC (NMR)

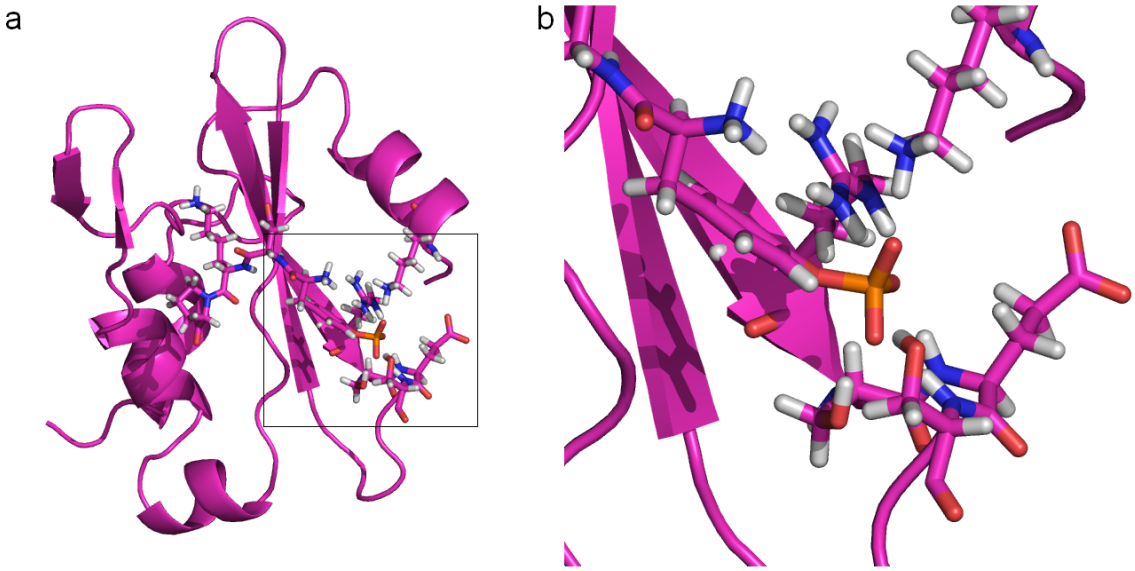

FIG. 1: a. Stat1 SH2 domain bound with a short phosphopeptide pYDKP. b. The phosphate group of the peptide has strong interaction with residue Lys7, Arg25, Ser27, Glu28 and Ser29.

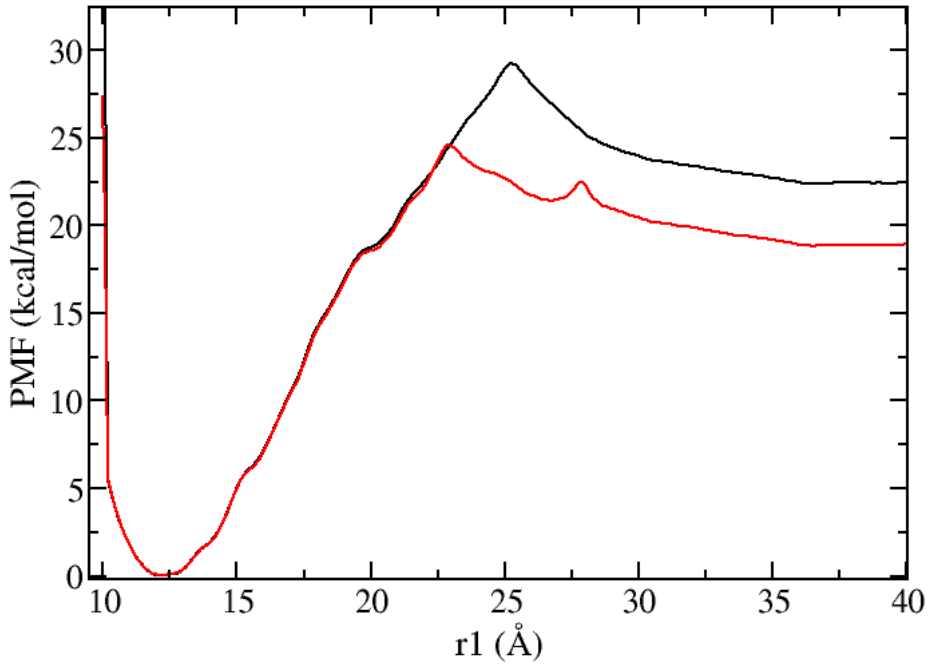

FIG. 2: Comparison of PMF graphs before and after introducing the second distance for Stat1 binding free energy calculations.

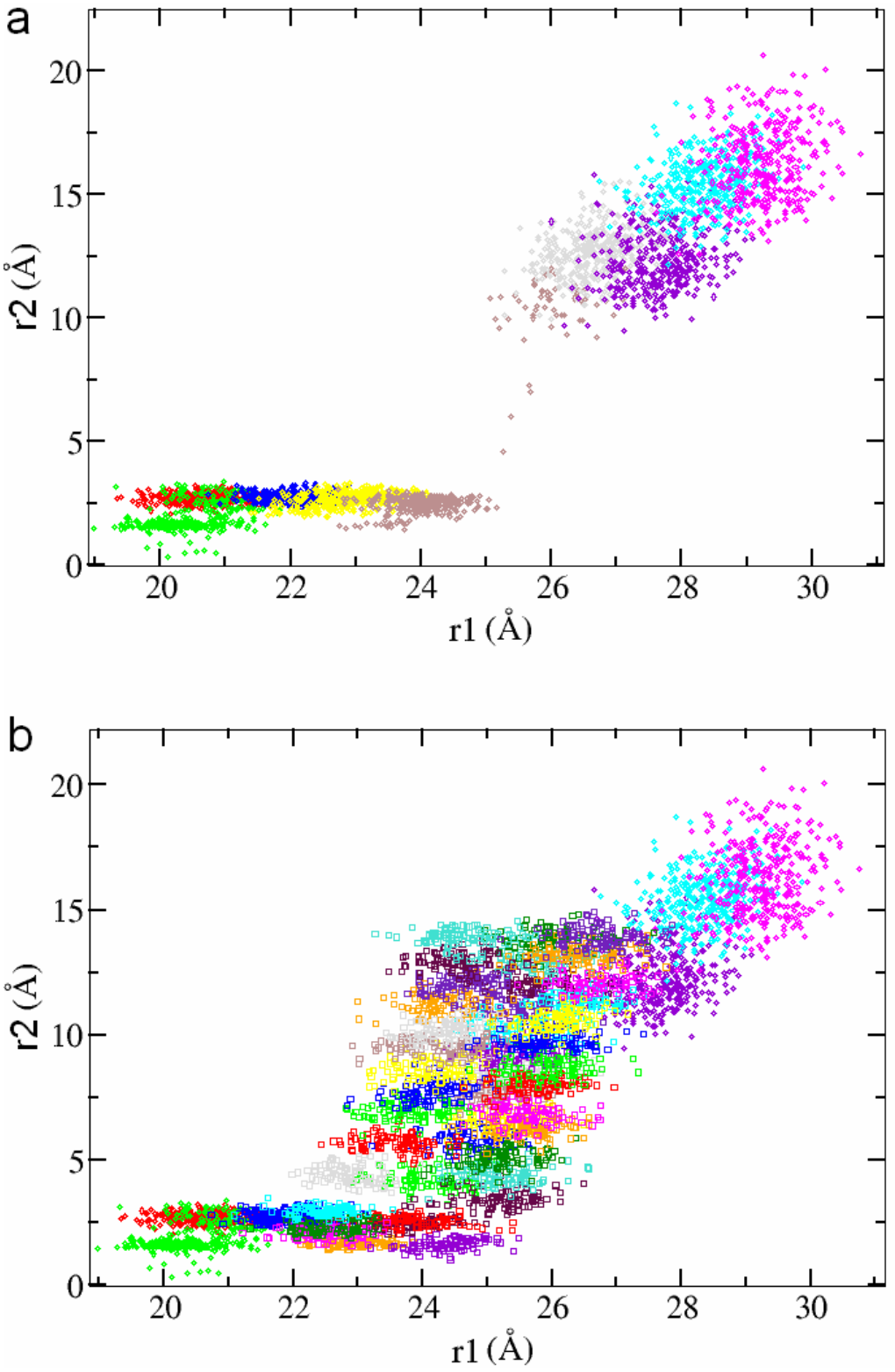

FIG. 3: Comparison of sampling before and after introducing the second variable r_2, where r_1 is the COM distance between the SH2 domain and peptide and r_2 the distance between the phophate group and phosphate binding residues.

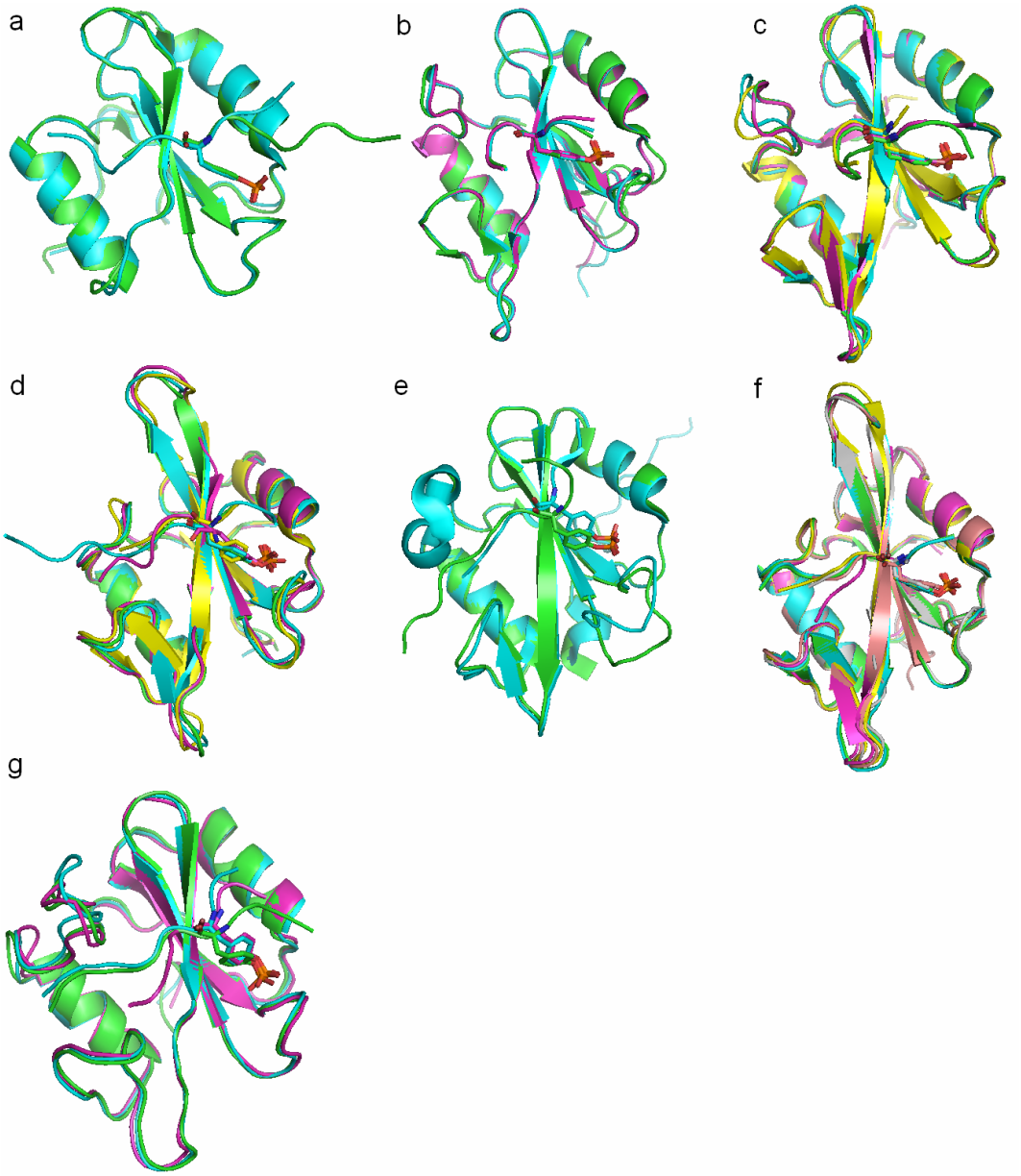

FIG. 4: Superposition of structures containing the same SH2 domain and different peptides. a. Cbl SH2 domain bound with SDGpYTPEPA and RAVENQpYSFY (PDB: 1FBV, 1YVH); b. Gads SH2 domain bound with DDpYVNV, AcREpYVNV and PDpYENL (PDB: 1R1P, 1R1Q, 1R1S); c. Grb2 SH2 domain bound with PSpYVNV, pYNpY, KPFpYVNV and EpYINQ (PDB: 1JYR, 1JYQ, 1TZE, 1ZFP); d. Lck SH2 domain bound with AcpYEEG, AcpYEEI, pYQPQ and DpYDYV (PDB: 1LKL, 1LCJ, 1LCK, 1X27); e. p85αN SH2 domain bound with SVDpYVPMLD and TNEpYMDMKP (PDB: 2IUI, 2IUH); f. Src SH2 domain bound with PQpYIpYVPA, PQpYEpYIPA, pYVPML, pYLRVA and AcpYEEI (PDB: 1NZV, 1NZL, 1SHA, 1SHB, 1SHD); g. Syp SH2 domain bound with SVLpYTAVQPNE, SPQEpYVNIEF and DGGpYMDMSKDE (PDB: 1AYA, 1AYB, 1AYC). The colors for different ligands are sequentially green, cyan, magenta, yellow, salmon.

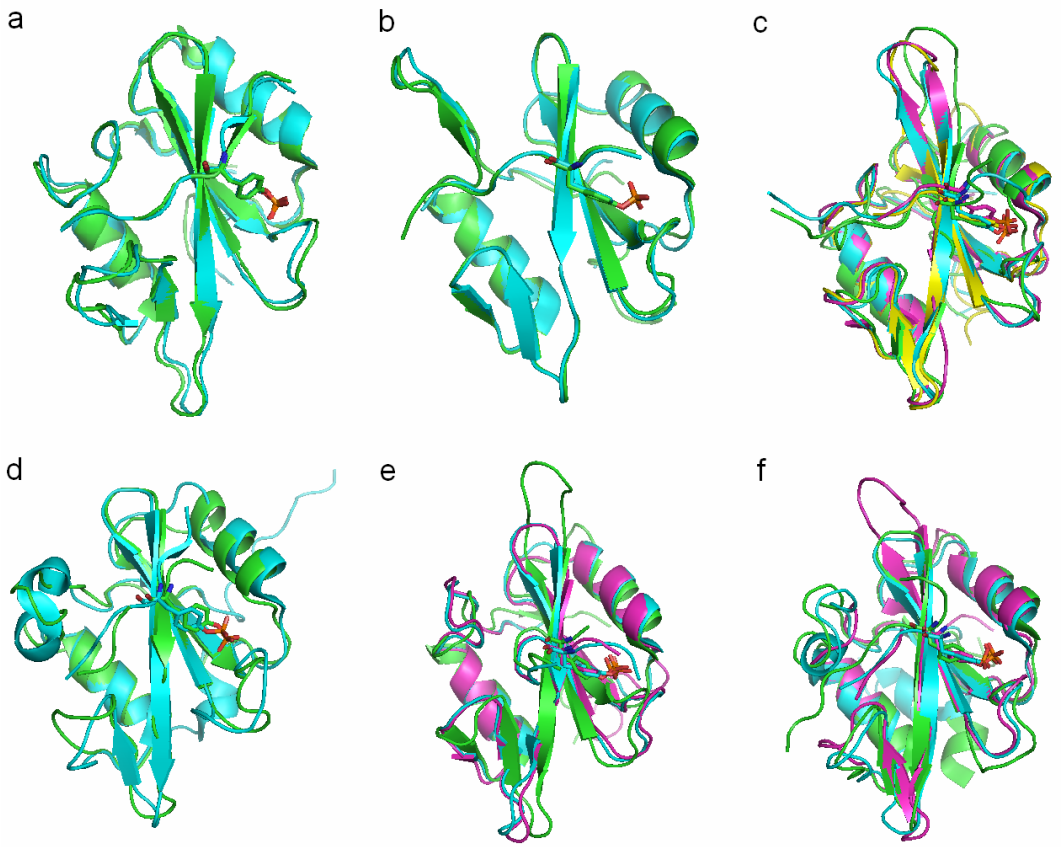

FIG. 5: Superposition of structures containing the same peptide motif and different SH2 domains. a. pYAQV bound with SAP SH2 and EAT2 SH2 (PDB: 1D4W, 1I3Z); b. pYDEV bound with Nck2 SH2 and Nck1 SH2 (PDB: 2CIA, 2CI9); c. pYEEI bound with Fyn SH2, Src SH2, Hck SH2 and Lck SH2 (PDB: 1AOT, 1SHD, 2C0I, 1LCJ); d. pYMDM bound with Syp SH2 and p85αN SH2 (PDB: 1AYC, 2IUH) e. pYVNV bound with Src SH2, Grb2 SH2 and Gads SH2 (PDB: 1F1W, 1JYR, 1R1P); f. pYVPM bound with p85αN SH2, p85αC SH2 and Src SH2 (PDB: 2IUI, 1H9O, 1SHA); The colors for different SH2 domains are sequentially green, cyan, magenta yellow.