Supplementary Materials:

Part S1. The influence of different k_+ values on predictions

The dataset we used to train the SVM classifiers was quite unbalanced: the ratio of binders versus non-binders was 1:20. When using penalty parameters, the objective function of SVM becomes:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C_+ \sum_{y_i=1} \xi_i + C_- \sum_{y_i=-1} \xi_i \\ \text{subject to} \quad & y_i (w_T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, ..., l. \end{aligned} \tag{1}$$

where $C_+ = k_+ C$ and $C_- = k_- C$ are the penalty parameters of error terms for the positive and the negative class, respectively; k_+ and k_- are the weight parameters for the positive and the negative class, respectively; $C > 0$ is a constant that represents the upper boundary of the penalty parameters.

In this study, we set a higher weight to the binder class (k_+) while keeping k_- to 1. Because different values of k_+ and k_- would affect the classification sensitivity and specificity of the SVM, we plotted a receiver operating characteristic (ROC) curve by varying k_+ (Figure S3). The area under the curve in the ROC plot represents of the performance of the model and our model achieved a value of 0.948. To determine the value of k_+ that would achieve balanced sensitivity (*SE*), prediction accuracy for binders (Q_+) and Matthews correlation coefficient (*C*), we analyzed how these classification criteria changed with the value of k_+ (Figure S4 and S5 in the supplementary materials). The value of k_+ is decided by the user: higher value of k_+ gives higher *SE* but lower Q_+ and *C*; lower of k_+ gives lower *SE* but higher Q_+ and *C*. Figure S4 and S5 suggest that $k_+ = 4$ may be a balanced value.

Table S1. The number of binders and non-binders included in the data set for training the MIEC-SVM classifier

	Binder	Non-binder
Abl_human	31	620
Boi1_yeast	25	500
c-Src_human	61	1220
Fyn_human	27	540
Grb2_mouse	19	380
Itk_human	5	100
Lsb3_yeast	25	500
Lyn_human	28	560
Myo3_yeast	6	120
Myo5_yeast	52	1040
Nbp2_human	25	500
P85a_human	29	580
Rvs167_yeast	19	380
Sla1_yeast	30	600
Spta2_human	20	400
Yes_human	29	580
Yha2_yeast	40	800
Ysc84_yeast	20	400
Total	491	9820

Table S2. The definition of the interaction pairs for the Abl SH3 domain for calculating MIECs

Peptide residue	SH3 residue
Ala_{-6}	Gly13, Asp14, Asn15, Thr16, Asn31, His32, Asn33, Trp36, Glu38, Trp47
Pro_{-5}	Gly13, Asp14, Asn15, Thr16, Asn31, His32, Asn33, Gly34, Trp36, Glu38, Trp47
Ser_{-4}	Ser12, Gly13, Asp14, Asn15, Thr16, His32, Asn33, 00, Gly34, Trp36, Trp47
Tyr_{-3}	Phe9, Val10, Ala11, Ser12, Gly13, Asp14, Asn15, Thr16, Glu35, Trp36, Trp47
Ser_{-2}	Ser12, Asn33, 00, 00, Gly34, Glu35, Trp36, Trp47
Pro_{-1}	Gly34, Glu35, Trp36, Pro49, Ser50, Asn51, Tyr52
Pro_{0}	Phe9, Trp36, Pro49, Asn51, Tyr52
Pro_{1}	Tyr7, Asp8, Asn51, Tyr52
Pro_{2}	Leu6, Tyr7, Asn51, Tyr52
Pro_{3}	Tyr7, Asp8, Asn51, Tyr52

Table S3.The performance of the SVM classifiers using various kernel functions and various combinations of domain-peptide MIECs

Model	MIECs*	Kernel function	SE_{train} (%)	SP_{train} (%)	SE_{test} (%)	SP_{test} (%)	Q_+ (%)	Q_- (%)	C
1	ΔE_{vdw},ΔE_{ele}	Linear	87.5	88.2	76.9	87.4	23.4	98.7	0.377
		Polynomial	65.0	85.4	62.4	85.2	17.4	97.8	0.269
		RBF	80.4	88.8	74.9	88.3	24.3	98.6	0.381
		Sigmoid	71.3	86.9	67.3	86.7	20.3	98.1	0.316
2	ΔE_{vdw}, ΔG_{polar}	Linear	89.9	88.3	79.5	87.6	24.3	98.8	0.394
		Polynomial	69.8	87.4	66.2	87.3	20.7	98.1	0.317
		RBF	71.1	88.1	67.5	87.8	21.8	98.2	0.333
		Sigmoid	66.3	85.4	63.8	85.4	17.9	97.9	0.279
3	ΔE_{vdw},ΔE_{ele},ΔG_{GB}	Linear	89.2	89.1	77.5	88.2	24.8	98.7	0.393
		Polynomial	48.6	87.5	46.5	87.3	15.5	97.0	0.206
		RBF	79.1	88.6	74.1	88.2	24.0	98.5	0.374
		Sigmoid	72.4	86.8	69.0	86.5	20.4	98.2	0.322

*ΔE_{vdw}, ΔE_{ele}, and ΔG_{GB} are van der Waals, electrostatic and polar contribution to desolvation, respectively. $\Delta G_{polar}=\Delta E_{ele}+\Delta G_{GB}$

Table S4. Distinct distributions of the binding free energies calculated by MM/GBSA and MM/PBSA for binders and non-binders.

No.	Domains	*p-value*[*]	
		MM/GBSA	MM/PBSA
1	Abl_human	$2.77e^{-8}$	$8.15e^{-12}$
2	Boi1_yeast	$1.75e^{-6}$	0.0050
3	c-Src_human	$1.27e^{-9}$	$7.71e^{-8}$
4	Fyn_human	$8.16e^{-5}$	$1.95e^{-4}$
5	Grb2_mouse	$1.09e^{-9}$	$8.16e^{-4}$
6	Lsb3_yeast	$1.52e^{-8}$	$3.12e^{-5}$
7	Lyn_human	$4.67e^{-22}$	0.0023
8	Myo5_yeast	0.55	$5.37e^{-4}$
9	Nbp2_human	$5.76e^{-11}$	$3.96e^{-10}$
10	P85a_human	$6.32e^{-23}$	$1.24e^{-9}$
11	Rvs167_yeast	$8.36e^{-13}$	$1.03e^{-12}$
12	Sla1_yeast	$7.29e^{-8}$	$3.92e^{-5}$
13	Spta2_human	$5.82e^{-4}$	0.010
14	Yes_human	$5.56e^{-12}$	$9.04e^{-7}$
15	Yha2_yeast	$4.24e^{-22}$	$7.63e^{-19}$
16	Ysc84_yeast	$6.73e^{-10}$	$2.40e^{-5}$

[*] P-value was calculated by Student's *t*-test. Two domains, Itk and Myo3, were not included in the calculation because they only had a small number of known binders (5 and 6, respectively).

Table S5. The classification accuracies of SVMs trained on the total binding free energies calculated by MM/GBSA.

No.	Domains	*SE* (%)	*n*+	*N*+	*SP* (%)	*n*-	*N*-	Q_+ (%)	Q_- (%)
1	Abl_human	90.3	28	31	75.3	467	620	15.5	99.4
2	Boi1_yeast	56.0	14	25	83.4	417	500	14.4	97.4
3	c-Src_human	44.3	27	61	87.0	1062	1220	14.6	96.9
4	Fyn_human	92.6	25	27	79.9	383	540	13..7	99.5
5	Grb2_mouse	63.2	12	19	79.7	303	380	13.5	97.7
6	Lsb3_yeast	76.0	19	25	78.8	473	600	13.0	98.7
7	Lyn_human	92.9	26	28	86.6	485	560	25.7	99.6
8	Myo5_yeast	82.7	43	52	21.1	219	1040	4.98	96.1
9	Nbp2_human	96.0	24	25	74.6	373	500	15.9	99.7
10	P85a_human	93.1	27	29	88.6	514	580	29.0	99.6
11	Rvs167_yeast	15.8	3	19	93.2	354	380	10.3	95.7
12	Sla1_yeast	96.7	29	30	79.0	474	600	18.7	99.8
13	Spta2_human	46.7	10	20	87.7	367	400	23.3	97.3
14	Yes_human	82.8	24	29	81.0	470	580	17.9	98.9
15	Yha2_yeast	75.0	30	40	84.9	679	800	19.9	98.5
16	Ysc84_yeast	90.0	18	20	80.3	321	400	18.6	99.4
	Total	74.8	359	480	75.9	7361	9700	13.3	98.4

Table S6. Comparison of MIEC-SVM and SH3-hunter on predicting the interactions between twenty peptides and four SH3 domains. These 20 peptides were not included in the training set of MIEC-SVM. It is not clear to us whether they were included in the training set of SH3-hunter.

No.	Peptides	Src			Yes			Abl			Grb2		
		Exp	MIEC-SVM	SH3-hunter	Exp	MIEC-SVM	SH3-hunter	Exp	MIEC-SVM	SH3-hunter	Exp	MIEC-SVM	SH3-hunter
1	LASRPLPLLP	√	√	√	√	√	√	×	×	√	×	×	×
2	ISQRALPPLP	√	√	√	√	×	√	×	×	√	×	×	×
3	ITMRPLPALP	√	√	√	√	√	√	×	×	√	√	√	√
4	RSGRPLPPIP	√	√	√	√	√	√	×	×	√	×	√	×
5	PPWWAPPPIP	×	×	√	×	√	√	√	√	√	×	√	×
6	APTYPPPPPP	×	×	×	×	×	√	√	√	√	×	√	×
7	LTPQSKPPLP	×	×	×	×	×	×	×	×	×	×	×	×
8	LGEFSKPPIP	×	×	×	×	×	×	×	×	×	×	×	×
9	SSAPQRPPLP	×	×	×	×	×	×	×	×	×	×	×	×
10	VVPLGRPEIP	×	×	×	×	×	×	×	×	×	×	×	×
11	MPPPVPPRPP	√	×	×	×	×	√	×	×	√	×	×	√
12	VPPLVAPRPP	×	×	×	×	×	×	×	×	×	×	×	×
13	GQPAGDPDPP	×	×	×	×	×	×	×	×	×	×	×	×
14	ATSEGLPILP	×	×	×	×	×	×	×	×	×	×	×	×
15	KWDSLLPALP	√	×	×	√	×	×	×	×	×	√	×	×
16	YWDMPLPRLP	√	√	×	×	×	×	×	×	×	√	√	×
17	YYQRPLPPLP	√	√	×	√	√	×	×	×	×	√	√	×
18	YFSRALPGLP	√	×	√	√	×	√	×	×	√	√	×	×
19	SLWDPLPPIP	√	√	×	√	√	×	×	×	×	√	√	×
20	DPYDALPETP	√	×	×	√	×	×	×	×	×	√	×	×
	Accuracy		16	13		15	13		20	14		14	11

Table S7. The 210 peptides tested in the peptide array experiment[a]

No.	Peptide	Prediction	Experiments
1	PPKFSPPPPP	√	√
2	PPHWAPPAPP	√	√
3	PPTWTPPKPP	×	√
4	KPTYPPPPPP	√	√
5	GPRWSPPPVP	√	√
6	GPRFPVPPVP	√	√
7	APKKPAPPVP	√	×
8	APTMPPPLPP	√	√
9	PPPYPPPPVP	√	√
10	PPPYPPPDIP	√	√
11	AAAMQKPSLP	×	×
12	AAMFQAPKTP	×	×
13	AFSFPHPINP	×	×
14	AGLMLTPTGP	×	×
15	AIWYTLPILP	×	×
16	AKRYVVPGHP	×	×
17	ALSMFAPLLP	×	×
18	ALWWFIPESP	×	×
19	ASQMLRPFAP	×	×
20	CEAFLKPWAP	×	×
21	DDFWPNPKFP	×	×
22	DDIMMSPHSP	×	×
23	DEAFKNPTKP	×	×
24	DGMMLGPKYP	×	×
25	DNGWIHPLDP	×	×
26	DNYYGTPKPP	×	×
27	DQGYPDPPNP	√	×
28	DVFMGPPGSP	×	×
29	ECLFPSPTRP	×	×
30	EDDFLEPATP	×	×
31	EGNFRTPMLP	×	×
32	EILWSAPLGP	×	×
33	EWKYLKPRTP	×	×
34	FALMGSPKPP	×	×
35	FPNWTYPVGP	×	×
36	FSHYPQPSPP	×	×
37	FSIFDSPSNP	×	×
38	GAGYPPPTMP	×	×
39	GGRFKRPTTP	×	×
40	GGRFNLPHAP	×	×
41	GLPYSHPPQP	×	×
42	GVVMPSPVKP	×	×

43	ILEYLHPRLP	×	×
44	IMQYTGPMLP	×	×
45	IMRYLAPEGP	×	×
46	ISSFSPPEKP	×	×
47	IVQWEEPVEP	×	×
48	KCGFPLPGVP	×	×
49	KPKFTTPEYP	×	×
50	KRMFPLPEVP	×	×
51	KSPFGVPGMP	×	×
52	KSQFRLPFKP	×	×
53	KSQYLQPKQP	×	×
54	LAAYAAPGYP	×	×
55	LAEYEMPIQP	×	×
56	LEDYKKPLPP	×	×
57	LEWMQNPEAP	×	×
58	LLNYIAPGEP	×	×
59	LSPFMIPLFP	×	×
60	MARWNQPQPP	×	√
61	MSFWLIPSRP	×	×
62	MTEMNPPTQP	×	×
63	NARFKRPVLP	×	×
64	PDPFKAPSRP	×	×
65	PGPYGLPGFP	×	×
66	PKKFHVPGLP	×	×
67	PLRWGPPEAP	×	×
68	PPPWAPPCSP	×	×
69	PPWMQPPPPP	√	√
70	QSIYGSPLSP	×	×
71	QSNYSYPQVP	×	×
72	RDSYPGPEDP	×	×
73	RFAFDRPGLP	×	√
74	RPAFGGPAIP	×	×
75	RWNFSPPEFP	×	×
76	SAPMPEPGAP	×	×
77	SEGWIEPSYP	×	×
78	SGHYSVPKLP	×	×
79	SLNFSSPDPP	×	×
80	SMPFAPPTLP	×	×
81	SNQFVGPIPP	×	×
82	SSLFYSPSSP	×	×
83	STKYNGPPFP	×	×
84	TALFTHPEGP	×	×
85	TFDMNRPLLP	×	×
86	TNSFYNPNSP	×	×

87	TPGFQNPQLP	×	×
88	TPTWESPARP	×	×
89	TRLYYTPEDP	×	×
90	TTTFDKPTVP	×	×
91	TWEFTQPGLP	×	×
92	TWVYYLPLLP	×	×
93	VGEWYKPDRP	×	×
94	VTAWQQPQPP	×	×
95	VVSMTPPHSP	×	×
96	WDNMVGPGKP	×	×
97	WPMFSAPSSP	×	×
98	YGDYTLPDVP	×	×
99	YITFIGPSWP	×	×
100	YLKYKDPQSP	×	×
101	YRMFISPLYP	×	×
102	AASLALPPQP	×	×
103	AAVEAYPEIP	×	×
104	ADIPKSPTKP	×	×
105	AFHSITPAPP	×	×
106	ALDQSMPVTP	×	×
107	ANFGLFPELP	×	×
108	ANTGGAPLNP	×	×
109	APCIKIPAAP	×	×
110	APRQQRPQQP	×	√
111	ATPVSGPTTP	×	×
112	AVPLIFPERP	×	×
113	CEDLPQPESP	×	×
114	CGPKPPPGFP	×	×
115	CIVKLVPSKP	×	×
116	CMIASTPPAP	×	×
117	DGLRFVPSLP	×	×
118	DRVHSFPTQP	×	×
119	DTFGDEPNNP	×	×
120	EIGKVPPPIP	×	×
121	EPAPRSPVPP	×	×
122	EREEGAPETP	×	×
123	ESLDDAPVAP	×	×
124	EVERNVPDPP	×	×
125	FCPCDTPYIP	×	×
126	FRGQGCPSTP	×	×
127	GALCSNPSCP	×	×
128	GEEQRPPETP	×	×
129	GIAVAQPILP	×	×
130	GNMGVVPPGP	×	×

131	GPSLPGPFSP	×	×
132	GPSPGGPQPP	×	×
133	GPSSLGPSNP	×	×
134	HSAGVIPIKP	×	×
135	HYGTMDPNIP	×	×
136	IIEETEPTVP	×	×
137	ISCDSSPVLP	×	×
138	KARSGPPTIP	×	×
139	KERIKQPPSP	×	×
140	KGDAGPPGIP	×	×
141	KVALGIPNLP	×	×
142	LADLGIPVMP	×	×
143	LAKQVDPYIP	×	×
144	LELQRLPERP	×	×
145	LFRISLPVAP	×	×
146	LFRLGPPKPP	×	×
147	LHLGDLPAEP	×	×
148	MVAEEAPPPP	×	×
149	NEPPPPPPPP	×	×
150	NGVLVRPRDP	×	×
151	NKCPAGPSGP	×	×
152	NMDVTFPSMP	×	×
153	PAAHGTPGAP	×	×
154	PANGEYPLNP	×	×
155	PDGSEIPLPP	×	×
156	PDWPLPPDWP	×	×
157	PFVHPKPPPP	×	×
158	PGTKGFPGSP	×	×
159	PKHSPPPPTP	√	×
160	PLRDPHPTPP	×	×
161	PPHQAIPLLP	×	×
162	PQLPLTPSTP	×	×
163	PSRGWAPPGP	×	√
164	PSSLMSPSTP	×	×
165	QPPPPPPQGP	×	×
166	QRLGPGPALP	×	×
167	RAALNLPLLP	×	×
168	RDFPGPPHAP	×	×
169	RKAVYLPGVP	×	×
170	RLSPTLPLLP	×	×
171	RNCLLRPGSP	×	×
172	RQLKLSPQVP	×	×
173	RQRLITPSPP	×	×
174	RTPPSPPGCP	×	×

175	RTPTTPPVFP	×	×
176	SERRDAPPPP	×	×
177	SETSPIPKTP	×	×
178	SHPSPAPVLP	×	×
179	SIDSGPPPLP	×	×
180	SKRAFEPRTP	×	×
181	SMVNSLPTFP	×	×
182	SQLVEFPLGP	×	×
183	SRDAASPDKP	×	×
184	SRHGLSPATP	×	×
185	SSDSLGPPRP	×	×
186	SSDSMFPYIP	×	×
187	SSLILPPKTP	×	×
188	SSSQLVPWKP	×	×
189	SSTKSKPGSP	×	×
190	STDPPKPPLP	×	×
191	STPCGEPNAP	×	×
192	SWPDDVPKIP	×	×
193	TGKKQVPLNP	×	×
194	TIEPVGPSDP	×	×
195	TKKRPAPRAP	×	×
196	TLWKAKPDEP	×	×
197	TSDGHCPLHP	×	×
198	TSFKIVPIVP	×	×
199	TTTTTTPDKP	×	×
200	VAPSSLPPPP	×	×
201	VDGDFIPDDP	×	×
202	VDPKYVPVKP	×	×
203	VEKVIYPGLP	×	×
204	VENHPNPAAP	×	×
205	VEPAAVPGEP	×	×
206	VEPNTVPHTP	×	×
207	VHQSTIPSNP	×	×
208	VSLGWEPVRP	×	×
209	WPPICDPPQP	×	×
210	YVRRRRPHKP	×	×

[a]√: binder; ×: non-binder

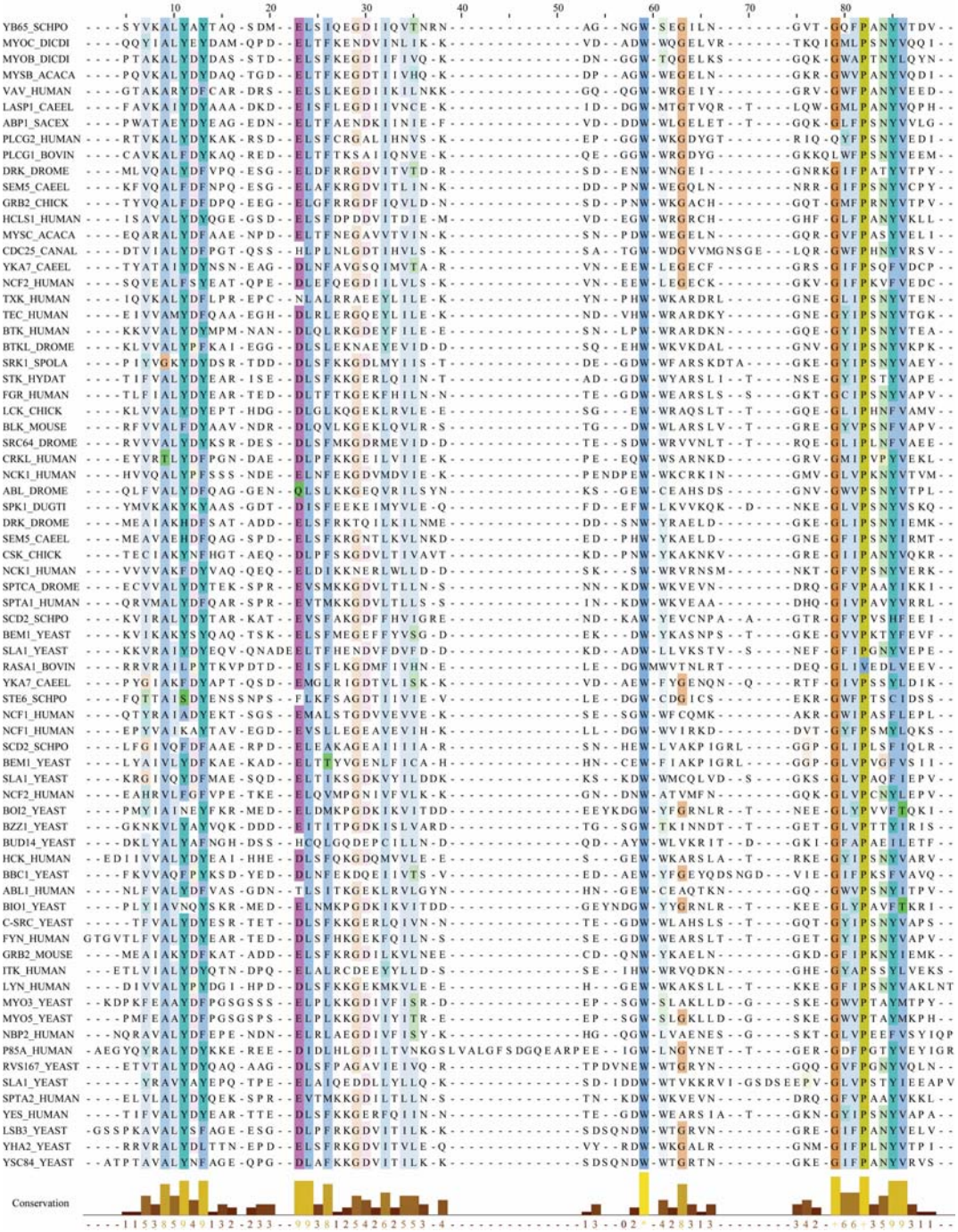


Figure S1. The multiple sequence alignment of the SH3 domains generated using MUSCLE. The sequences of the eighteen SH3 domains studied in this work are shown at the bottom of the multiple sequence alignment. The alignment is colored based on the sequence conservation (conservation larger than 25%) using the ClustalX coloring scheme. The conservation score reflecting the conservation of physico-chemical properties in the alignment is shown at the bottom of the figure. The figure was generated by Jalview.

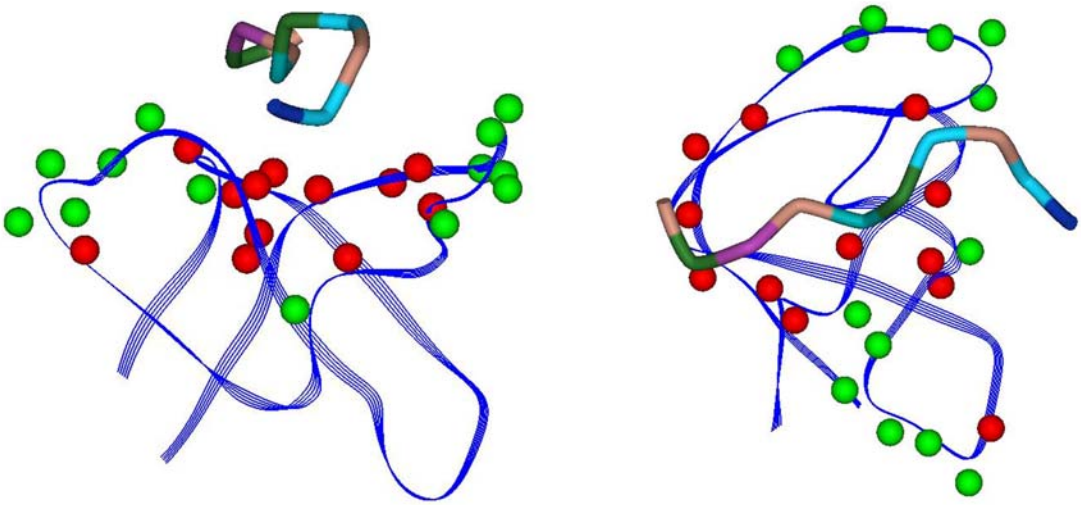


Figure S2. The spatial distributions of residues at the twenty-five important positions of the Bio1 SH3 domains (two different orientations). The SH3 domain is shown using the strand model; the peptide is shown using the stick model; the C_α carbons of the important residues are shown using the CPK model. The twelve conserved residues are colored in red and the others in green.

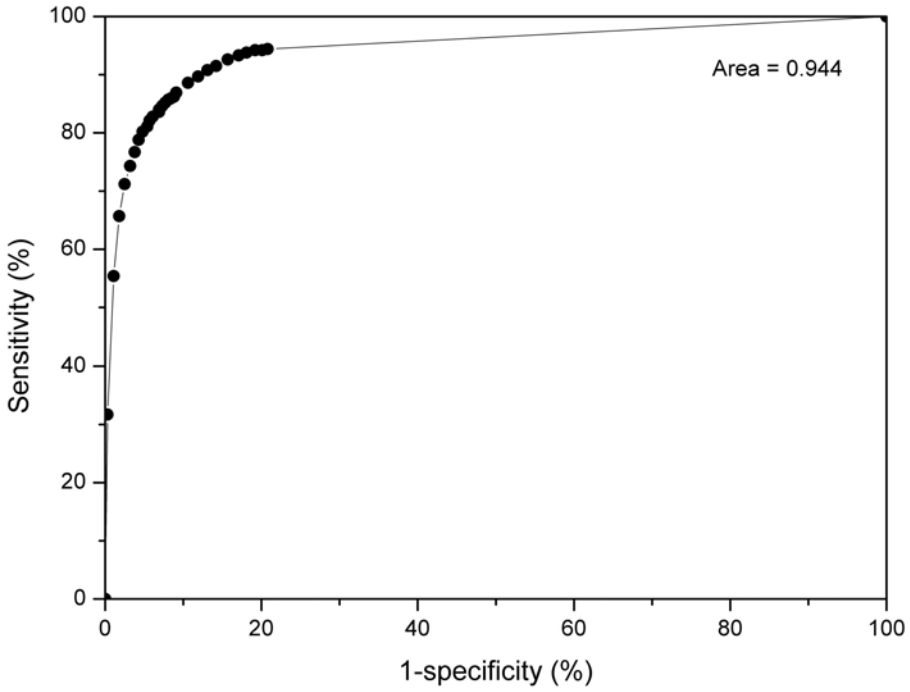


Figure S3. The ROC curve for the unified MIEC-SVM model.

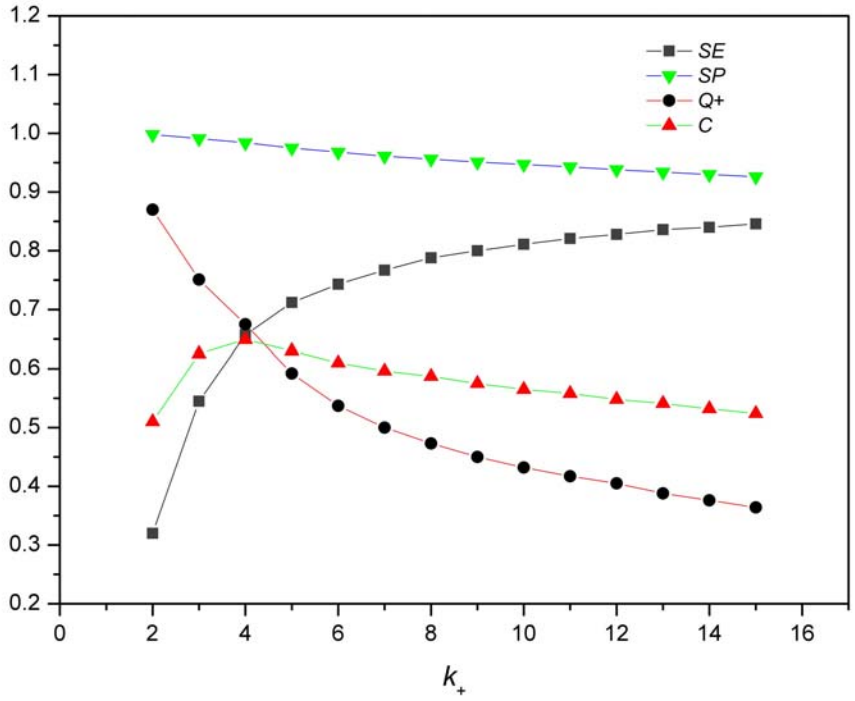


Figure S4. Sensitivity (*SE*), specificity (*SP*), accuracy of the binder class (Q_+) and Matthews correlation coefficient (*C*) of the MIEC-SVM versus the weight parameter (k_+) for the binder class in cross validations.

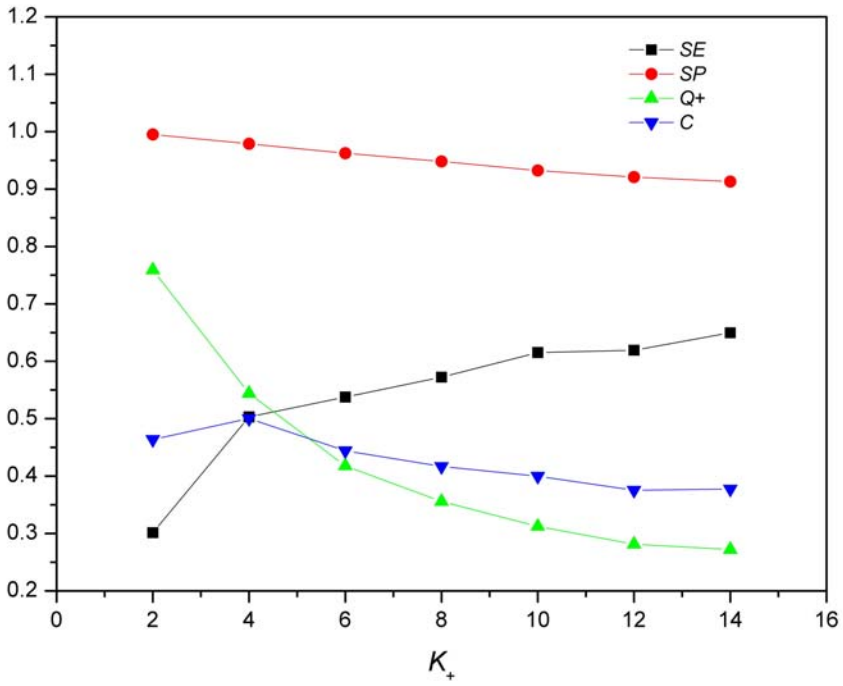
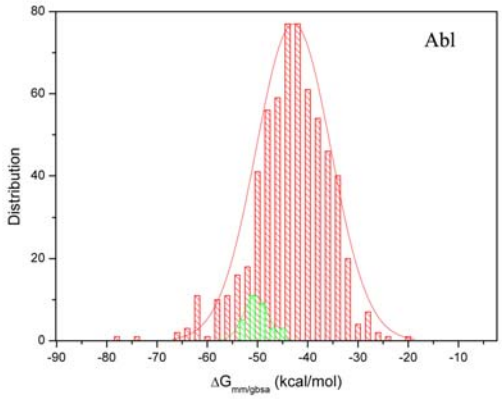
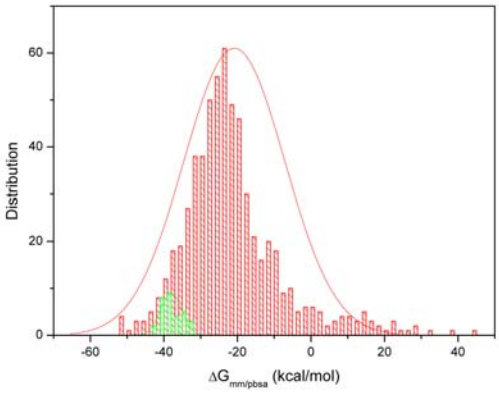
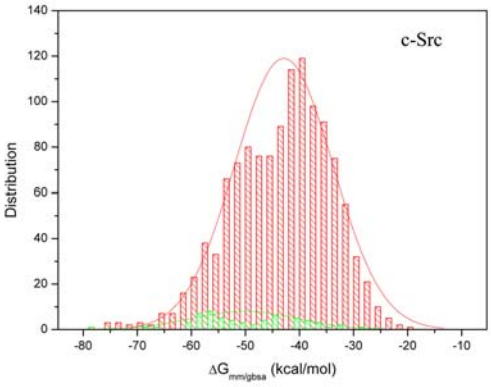
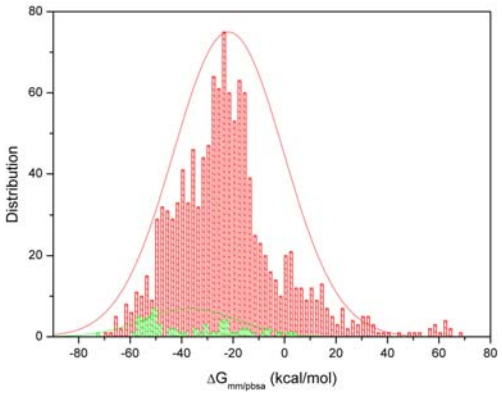


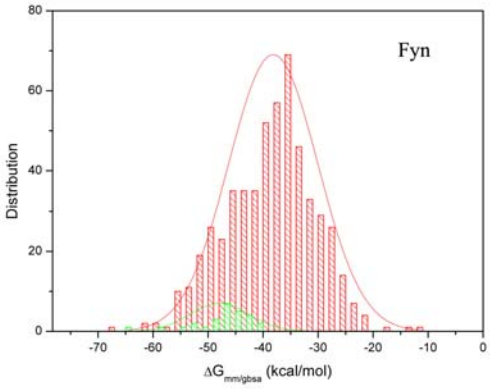
Figure S5. Sensitivity (*SE*), specificity (*SP*), accuracy of the binder class (Q_+) and Matthews correlation coefficient (*C*) versus the weight parameter (k_+) for the binder class in leave-one-SH3-out cross validations.

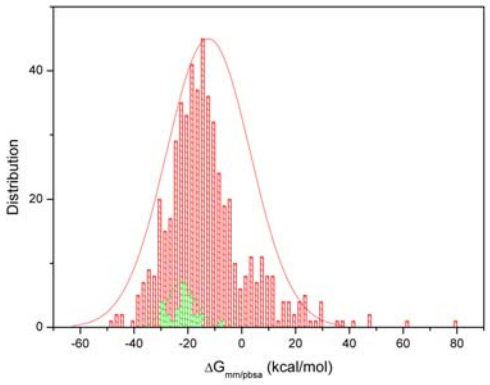

Abl
Distribution
ΔG mm/gbsa (kcal/mol)

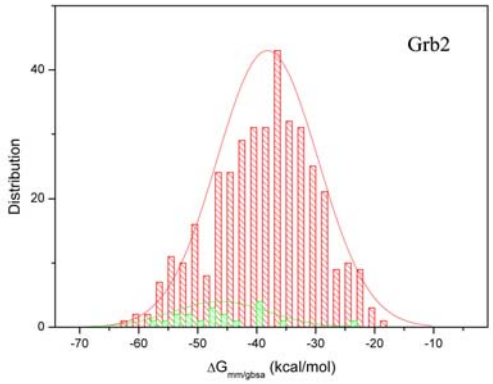

Distribution
ΔG mm/pbsa (kcal/mol)

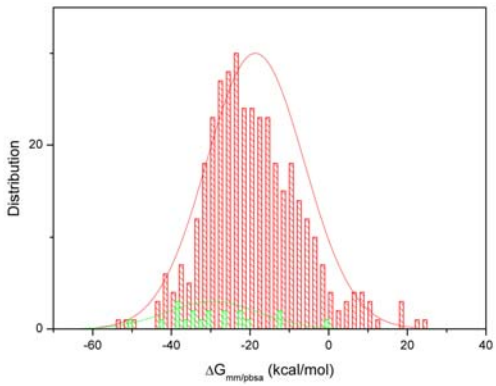

c-Src
Distribution
ΔG mm/gbsa (kcal/mol)

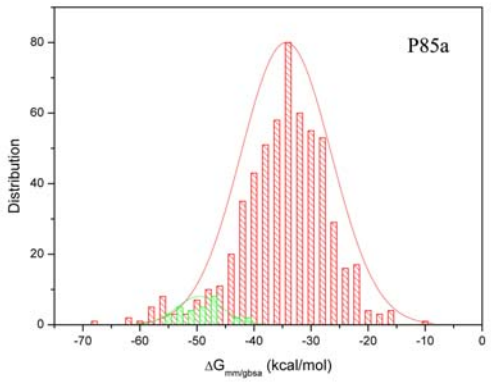

Distribution
ΔG mm/pbsa (kcal/mol)

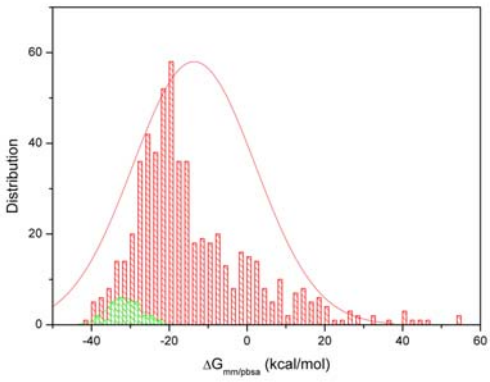

Fyn
Distribution
ΔG mm/gbsa (kcal/mol)

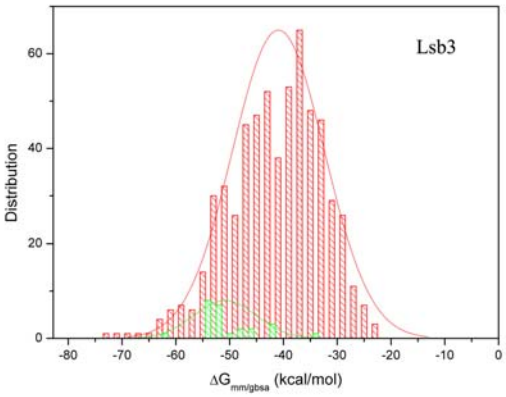

Distribution
ΔG mm/pbsa (kcal/mol)

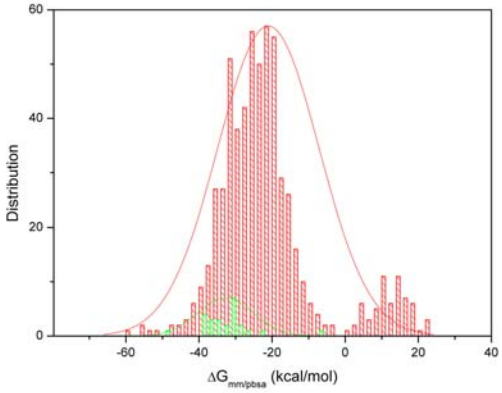

Grb2
Distribution
ΔG mm/gbsa (kcal/mol)

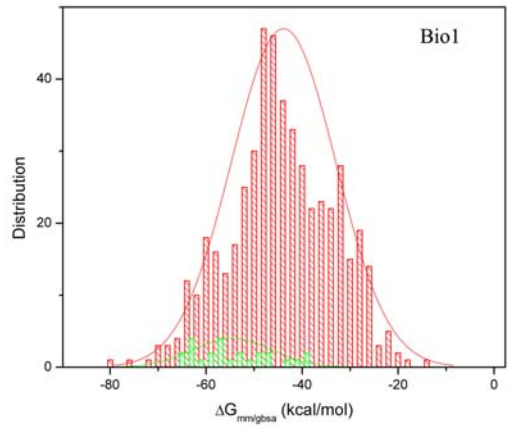

Distribution
ΔG mm/pbsa (kcal/mol)

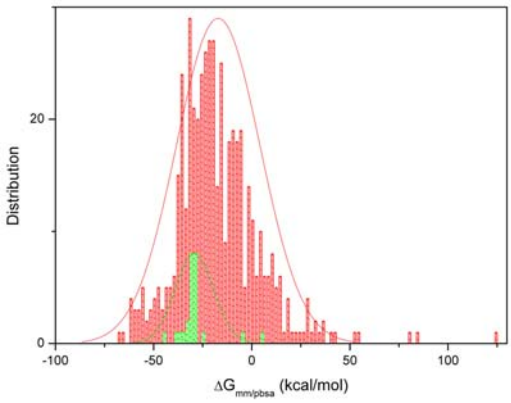

P85a
Distribution
ΔG mm/gbsa (kcal/mol)

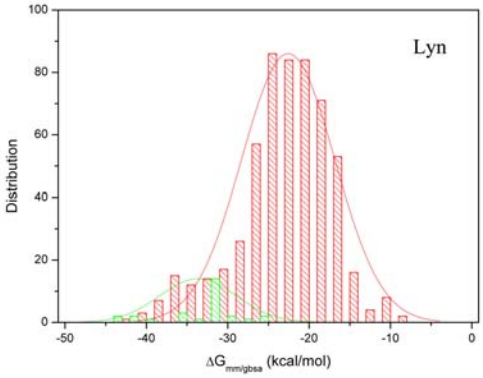

Distribution
ΔG mm/pbsa (kcal/mol)

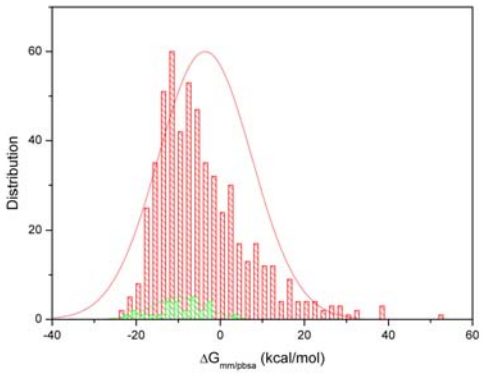

Lsb3
Distribution
ΔG mm/gbsa (kcal/mol)

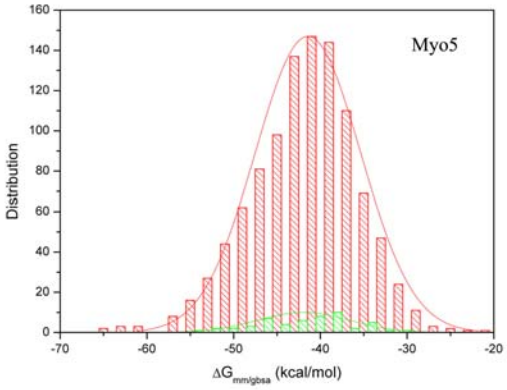

Distribution
ΔG mm/pbsa (kcal/mol)

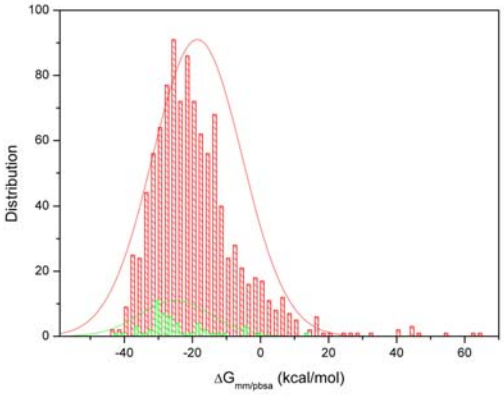

Bio1
Distribution
ΔG mm/gbsa (kcal/mol)

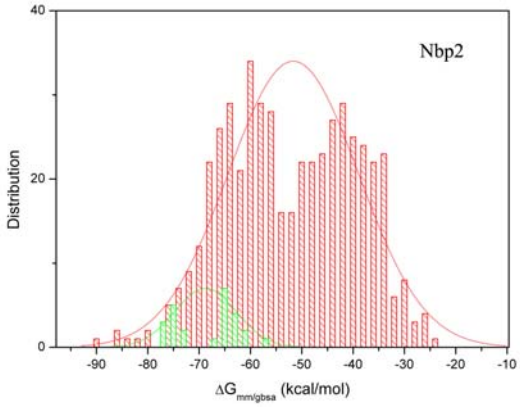

Distribution
ΔG mm/pbsa (kcal/mol)

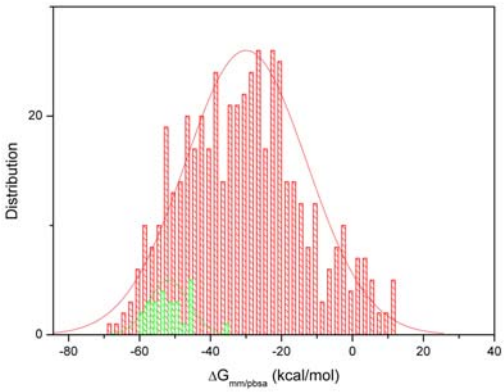

Lyn
Distribution
ΔG mm/gbsa (kcal/mol)

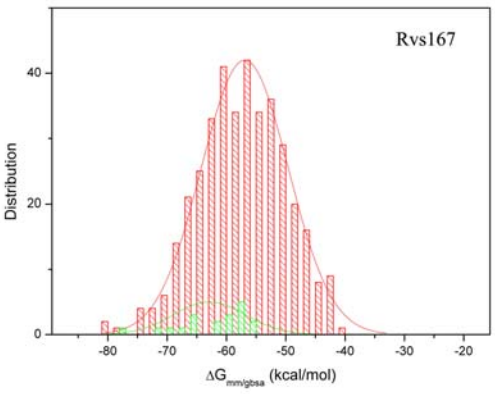

Distribution
ΔG mm/pbsa (kcal/mol)

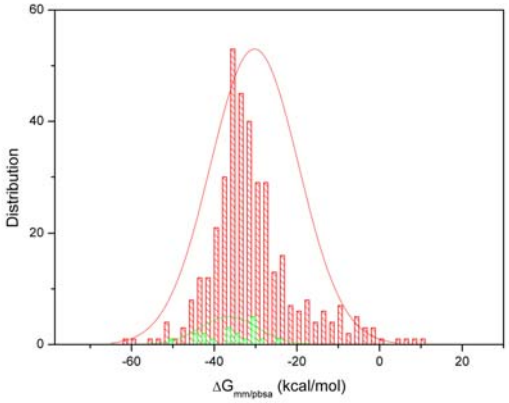

Myo5
Distribution
ΔG mm/gbsa (kcal/mol)

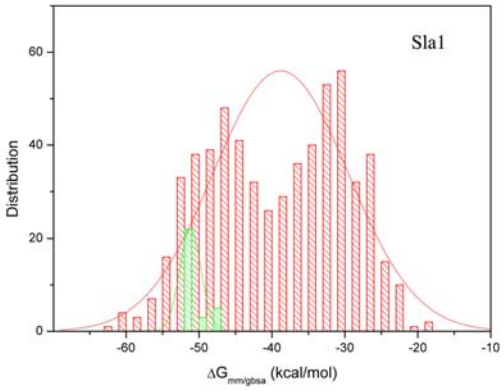

Distribution
ΔG mm/pbsa (kcal/mol)

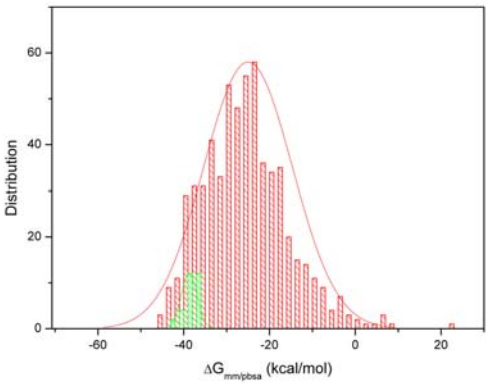

Nbp2
Distribution
ΔG mm/gbsa (kcal/mol)

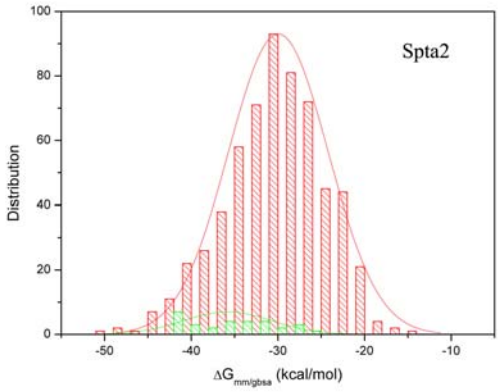

Distribution
ΔG mm/pbsa (kcal/mol)

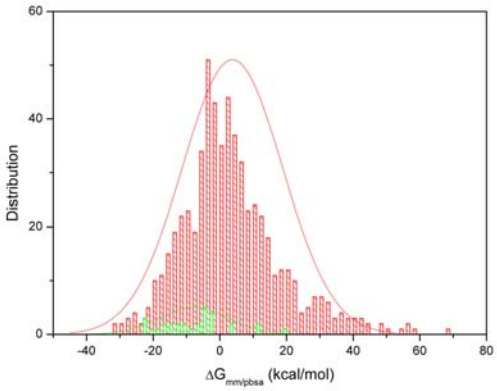

Rvs167
Distribution
ΔG mm/gbsa (kcal/mol)

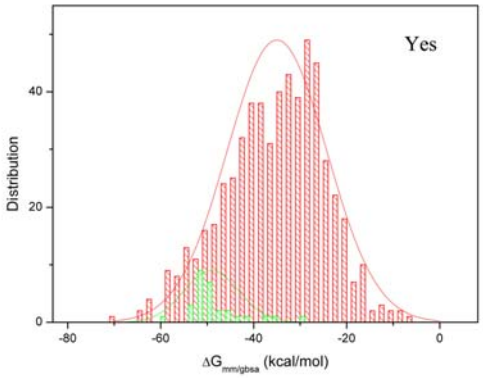

Distribution
ΔG mm/pbsa (kcal/mol)

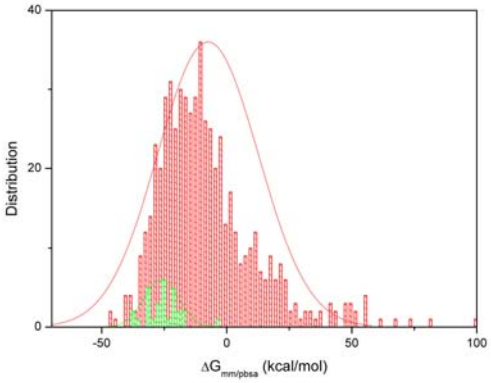

Sla1
Distribution
ΔG mm/gbsa (kcal/mol)

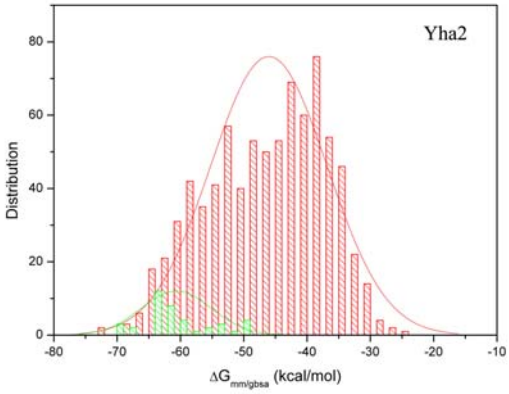

Distribution
ΔG mm/pbsa (kcal/mol)

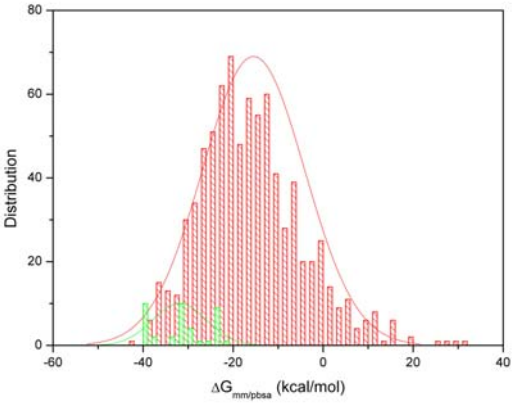

Spta2
Distribution
ΔG mm/gbsa (kcal/mol)


Distribution
ΔG mm/pbsa (kcal/mol)


Yes
Distribution
ΔG mm/gbsa (kcal/mol)


Distribution
ΔG mm/pbsa (kcal/mol)


Yha2
Distribution
ΔG mm/gbsa (kcal/mol)


Distribution
ΔG mm/pbsa (kcal/mol)

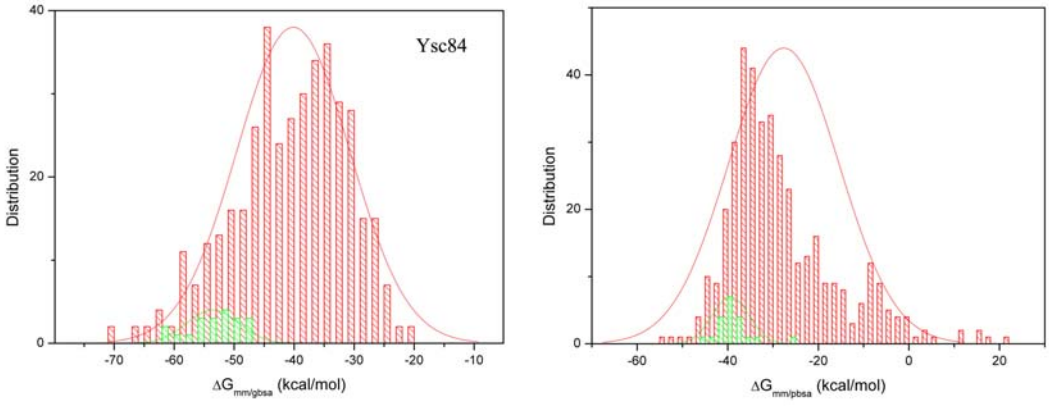


Figure S6. The distributions of the binding free energies of binders and non-binders for the 16 SH3 domains calculated by MM/GBSA (left) and MM/PBSA (right). Red and green bars represent non-binders and binders, respectively.