Supplementary Material

Predicting Ca^{2+}-binding Sites Using Refined Carbon Clusters

Kun Zhao[1], Xue Wang[2], Hing C. Wong[2], Robert Wohlhueter[2], Michael Kirberger[2], Guantao Chen[1] and Jenny J. Yang[2]*

[1]Department of Mathematics and Statistics, Georgia State University, Atlanta, GA, 30303

[2]Department of Chemistry, Center for Drug Design and Biotechnology, Georgia State University,
Atlanta, GA, 30303

Table S1. Adjacent matrix for carbon atoms graph representing binding loop of D20-E31 from calmodulin (3CLN.pdb)

C ID	D20C	D20CG	K21C	D22C	D22CG	G23C	D24C	D24CG	G25C	T26C	T26CB	I27C	T28C	T28CB	T29C	T29CB	K30C	E31C	E31CD
D20C	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	1	1
D20CG	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1
K21C	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
D22C	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1
D22CG	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	1
G23C	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1
D24C	0	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
D24CG	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	1
G25C	0	1	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0
T26C	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1
T26CB	0	1	0	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	1
I27C	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
T28C	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1
T28CB	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1
T29C	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
T29CB	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
K30C	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
E31C	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
E31CD	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	0	1	1	1

Binding loop of D20-E31 from calmodulin (3CLN.pdb) is used as example to illustrate how the adjacent matrix is constructed for the Figure 2d in the Methods section. In the Table S1, “D20C” represents the mainchain carbon from the 20th Asp while the “D20CG” represents the sidechain gamma carbon from the same residues. “1” represents the distance between two carbon atoms is smaller than 7.5Å; “0” otherwise. In this example, the distance cutoff of C-C is 7.5Å.

Table S2. The parameters used in the dataset for MUG^C in X-ray and NMR.

	X-ray	NMR
Cutoff of maximum clique	7.5Å	8.3Å
dist(ca,c1)	(2.5Å ,4.5Å)	(1.74Å ,4.9Å)
dist(ca,c2)	> dist(ca,c1)	> dist(ca,c1)-0.5
angle(ca,c1,c2)	(>90)	(>70)
Center of Mass		
R-Ca (sidechain O)	(<4.3)	(<4.5)
Ca-O (mainchain O)	(<Ca-R)	(<Ca-R)
Clash	Van der waals radius	Van der waals radius
Ca-N	(>2.55)	(>2.55)
Ca-C	(>1.74 [a]/2.7 [b])	(>1.74 [a]/2.7 [b])
Ca-O	(>1.6)	(>1.6)

[a, b]: 1.74 for monodentate and 2.7 for bidentate

Table S3. Summary of X-ray training dataset.

PDB[a]	Res[b]	Protein[c]	Chain[d]	Size[e]
1ALA	2.25	Annexin V	A	316
1ALV	1.90	Calpain A	A	173
1AUI	2.10	Calcineurin	A, B	635
1AVS	1.75	TroponinC	A, B	163
1B9O	1.15	Lactalbumin	A	123
1EXR	1.00	Calmodulin	A	146
1FJ3	2.00	Thermolysin	A	317
1GLG	2.00	Galactosebinding	A	310
1NLS	0.94	Concanavalin A	A	237
1PSH	2.30	Phospholipase A2	A,B,C	357
1SCD	2.30	Subtilisin Carlsberg lightly cross-linked with glutaraldehyde	A	275
1SNC	1.65	Staphyloccoal Nuclease	A	135
1THM	1.37	Thermitase	A	279
2PRK	1.50	Proteinase K	A	279
3EST	1.65	Elastase	A	230
4ICB	1.60	Calbindin	A	76
5PAL	1.54	Parvalbumin	A	109

[a] PDB code. [b] PDB resolution. [c] Protein name. [d] Chain number. [e] Number of residues.

Table S4. Summary of X-ray testing dataset.

PDB[a]	Res[b]	Protein[c]	Chain[d]	Size[e]
1BJR	2.44	Complex:lactoferrin fragment and proteinase K	E, I	289
1JDA	2.20	Maltotetraose forming exo-amylase	A	418
1FZC	2.30	Fibrin	A,B,C,D,E,F,G,H,I,J	1382
1SBH	1.80	Subtilisin	A	275
1OBR	2.30	Carboxypeptidase T	A	323
1EGZ	2.30	Cellulase	A, B, C	873
1ESL	2.00	E-Selectin	A	157
1AI4	2.35	Penicillin acylase	A, B	763
1ATL	1.80	Atrolysin C	A, B	500
1AX0	1.90	Lectin	A	239
1B9Z	2.10	β-Amylase	A	516
1BF2	2.00	Pseudomonas Isoamylase	A	750
1CE5	1.90	β-Trypsin	A	230
1CLX	1.80	Xylanase	A, B, C, C	1380
1GCG	1.90	Galactose binding protein	A	309
1HYT	1.70	Thermolysin	A	316
1IAG	2.00	Adamalysin (II)	A	201
1IRB	1.70	Carboxylic ester hydrolase	A	123
1JS4	2.00	Endoexocellulase E4	A, B	1210
1KBC	1.80	Neutrophil collagenase	A, B	328
1KIT	2.30	Hydrolase	A	757
1KVX	1.90	Carboxylic ester hydrolase	A	123
1MMQ	1.90	Matrilysin	A	165
1NBC	1.75	Cellusomal scaffolding Protein A	A, B	310
1OAC	2.00	Amine oxidase	A, B	1443
1OIL	2.10	Lipase	A, B	640
1SBF	2.43	Soybean agglutinin	A	234
1SRA	2.00	Calcium biniding protein	A	151
1TCM	2.20	Cyclodextrin glycosyl Transferase	A, B	1372
1TN3	2.00	Tetranectin	A	137
2FIB	2.01	Fibrogen	A, B	254
2TEP	2.50	Peanut lectin	A, B, C, E	928
4LIP	1.75	Lipase	D, E	638
1C9M	1.67	Bacillus lentus subtilisin	A	269
2SCP	2.00	Sarcoplasmic Ca(2+)-binding protein (SCP)	A, B	174
1TVG	1.6	HSPC034	A	153
2GGM	2.35	Human centrin 2 xeroderma pigmentosum group C protein complex	A,B,C,D	172
3FIA	1.45	Human intersectin-1 protein	A	121
1K9K	1.76	calcium bound human S100A6	A, B	90
1DAN	2.00	Complex of active site inhibited human blood coagulation factor via with human recombinant soluble tissue factor	L, H, T, U, C	152, 254, 80, 121, 4
1MHO	2.00	S100B from bovine brain	A	88
1EDH	2.00	E-cadherin domains 1 and 2 in complex with calcium	A, B	226
2EGD	1.8	human S100A13	A, B	98

[a] PDB code. [b] PDB resolution. [c] Protein name. [d] Chain number. [e] Number of residues.

Table S5. Prediction results on the X-ray training dataset.

PDB ID	Ca# [a]	Documented Ligands	Predicted Ligands	R[b]
1ALA	2505	M28, G32, T37, E72	M28, G32, E72	3/4
	2506	I100, G102, G104, E144	I100, G102, G104, E144	4/4
	2507	M259, G261, G263, D303	M259, G261, G263, D303	4/4
1ALV	3425	A107, D110, E112, E117	A107, D110, E112, E117	4/4
	3426	D150, D152, T154, K156, E161	D150, D152, T154, K156, E161	5/5
	3427	D180, D182, S184, T186, E191	D180, D182, S184, T186, E191	5/5
	3428	D135, D223, D225, N226	D135, D223, D225, N226	4/4
1AUI	4393	D30, D32, S34, S36, E41	D30, D32, S34, S36, E41	5/5
	4394	D62, D64, N66, E68, E73	D62, D64, N66, E68, E73	5/5
	4395	D99, D101, D103, Y105, E110	D99, D101, D103, Y105, E110	5/5
	4396	D140, D142, D144, R146, E151	D140, D142, D144, R146, E151	4/4
1AVS	1266	D30, D32, D36, E41	D30, D32, D36, E41	4/4
	1267	D66, D68, S70, T72, E77	D66, D68, S70, T72, E77	3/3
	1268	D30, D32, D36, E41	D30, D32, D36, E41	3/3
	1269	D66, D68, S70, T72, E77	D66, D68, S70, T72, E77	4/4
1B9O	1032	K79, D82, D84, D87	K79, D82, D84, D87	4/4
1EXR	1469	D20, D22, D24, T26, E31	D20, D22, D24, T26, E31	5/5
	1470	D56, D58, N60, T62, E67	D56, D58, N60, T62, E67	5/5
	1471	N129, D131, D133, H135, E140	N129, D131, D133, H135, E140	5/5
	1472	E47	-	
	1473	D93, D95, N97, L99, E104	D93, D95, N97, L99, E104	5/5
1FJ3	2461	D138, E177, D185, E187, E190	D138, E177, D185, E187, E190	4/4
	2462	E177, D185, E190	E177, D185, E190	3/3
	2463	D57, D59, G61	-	0/3
	2464	Y193, T194, I197, D200	Y193, T194, I197, D200	4/4
1GLG	2362	D134, N136, D138, Q140, Q142, E205	D134, N136, D138, Q140, Q142, E205	6/6
1K96	715	S20, E23, D25, T28, E33	S20, E23, D25, T28, E33	5/5
	716	D61, N65, D65, E67, E72	D61, N65, D65, E67, E72	5/5
1NLS	1875	D10, Y12, N14, D19	D10, Y12, N14, D19	4/4
1PSH	2788	Y27, G29, G31, D48	Y27, G29, G31, D48	4/4
	2789	Y27, G29, G31, D48	Y27, G29, G31, D48	4/4
	2790	Y27, G29, G31, D48	Y27, G29, G31, D48	4/4
1SCD	1922	D41, L75, N77, T79, V81	D41, L75, N77, T79, V81	5/5
	1923	A169, Y171, V174	-	4/4
1SNC	1184	D21, D40, T41	D21, T41	2/3
1THM	2005	D5, D47, V82, N85, T87, I89	D5, D47, V82, N85, T87, I89	4/4
	2006	D57, D62, T64, Q66	D57, D62, Q66	3/4
2PRK	2019	P175, V177, D200	P175, V177, D200	3/3
	2020	T16, D260	-	2/2
3EST	1824	E70, N72, Q75, N77, E80	E70, N72, Q75, N77, E80	5/5
4ICB	641	A14, E17, D19, Q22, E27	A14, E17, D19, Q22, E27	5/5
	642	D54, N56, D58, E60, E65	D54, N56, D58, E60, E65	5/5
5PAL	843	D90, D92, D94, K96, E101	D90, D92, D94, K96, E101	4/4
	844	D51, D53, S55, F57, E59, E62	D51, D53, S55, F57, E59, E62	6/6

[a]: metal identification number in PDB file. [b]: the correctly predicted ligands over documented ligands

Table S6. Prediction results on the X-ray testing dataset.

PDB ID	Ca# [a]	Documented Ligands	Predicted Ligands	R[b]
1BJR	2090	P175, V177, D200	P175, V177, D200	3/3
	2089	R12, S15, N257, A273	-	0/4
1JDA	3299	N116, D151, D154, D162, G197	N116, D151, D154, D162, G197	5/5
	3300	D1, Q2, H13, D16, E17	D1, Q2, H13, D16, E17	5/5
1FZC	11170	D381, D383, W385	D381, D383, W385	3/3
	11171	D318, D320, F322, G324	D318, D320, F322, G324	4/4
	11172	D381, D383, W385	D381, D383, W385	3/3
	11173	D318, D320, F322, G324	D318, D320, F322, G324	4/4
1SBH	1942	Q2, D41, L75, N77, V81	Q2, D41, L75, N77, V81	5/5
	1943	A169, Y171, V174	-	0/3
1OBR	2584	D56, E57, E61, E104	D56, E57, E61, E104	4/4
	2585	S50, D51,E57,E59	S50, D51,E57,E59	4/4
	2586	D51,E59, N101	D51,E59, N101	3/3
	2587	S7, Y9, E14	S7, Y9, E14	3/3
1EGZ	6810	G121, D158, D160, N161	G121, D158, D160, N161	4/4
	6811	G121, D158, D160, N161	G121, D158, D160, N161	4/4
	6812	G121, D158, D160, N161	G121, D158, D160, N161	4/4
1ESL	1267	E80, N82, N105, D106	E80, N82, N105, D106	4/4
	1268	E33, E36	-	0/2
	1270	Q20, Y23	-	0/2
1AI4	6074	E152, D73, V75, D76, P205, D252	E152, D73, V75, D76, P205, D252	6/6
1ATL	3260	E9, D93, C197, N200	E9, D93, C197, N200	4/4
	3262	E9, D93, C197, N200	E9, D93, C197, N200	4/4
1AX0	1996	D129, F131, N133, D136	D129, F131, N133, D136	4/4
1B9Z	4310	D56, D60, Q61, E141, E144	D56, D60, Q61, E141, E144	5/5
1BF2	5737	D128, E229, T230, N232, D259	D128, E229, T230, N232, D259	5/5
1CE5	1631	E70, N72, V75, E80	E70, N72, V75, E80	4/4
1CLX	10801	N253, D256, N258, N261, D262	N253, D256, N258, N261, D262	5/5
	10802	N253, D256, N258, N261, D262	N253, D256, N258, N261, D262	5/5
	10803	N253, D256, N258, N261, D262	N253, D256, N258, N261, D262	5/5
	10804	N253, D256, N258, N261, D262	N253, D256, N258, N261, D262	5/5
1GCG	2895	D134, N136, D138, K140, Q142, E205	D134, N136, D138, K140, Q142, E205	6/6
1HYT	2440	D138, E177, D185, E187, E190	D138, E177, D185, E187, E190	5/5
	2441	E177, N183, D185, E190	E177, N183, D185, E190	4/4
	2442	D57, D59, N61	-	0/3
	2443	Y193, T194, I197, D200	Y193, T194, I197, D200	4/4
1IAG	1623	E9, D93, C197, N200	E9, D93, C197, N200	4/4
1IRB	951	Y28, G30, G32, D49	Y28, G30, G32, D49	4/4
1JS4	9586	S210, G211, D214, E215, D261	S210, G211, D214, E215, D261	5/5
	9587	T504, D506, D571, N574, D575	T504, D506, D571, N574, D575	5/5
	9588	S210, G211, D214, E215, D261,	S210, G211, D214, E215, D261,	5/5
	9589	T504, D506, D571, N574, D 575	T504, D506, D571, N574, D 575	5/5
1KBC	2591	D137, G169, G171, D173,	D137, G169, G171, D173,	4/4
	2592	D154, G155, N157, I159, D177, E180	D154, G155, N157, I159, D177, E180	6/6
	2595	D137, G169, G171, D173	D137, G169, G171, D173	4/4
	2596	D154, G155, N157, I159, D177, E180	D154, G155, N157, I159, D177, E180	6/6
1KIT	5861	A253, N256, D289, T313	A253, N256, D289, T313	4/4
	5862	D621, D682, A683	D621, D682, A683	3/3
1KVX	956	Y28, G30, G32, D49	Y28, G30, G32, D49	4/4
1MMQ	1272	D175, G176, G178, T180, D198, E201	D175, G176, G178, T180, D198, E201	6/6
	1273	D158, G190, G192, D194	D158, G190, G192, D194	4/4
1NBC	2437	T44, D46, T122, N125, D126	T44, D46, T122, N125, D126	5/5

	2438	T44, D46, T122, N125, D126	T44, D46, T122, N125, D126	5/5
1OAC	11388	D533, L534, D535, D678, A679	D533, L534, D535, D678, A679	5/5
	11389	E573, Y667, D670, E672	E573, Y667, D670, E672	4/4
	11391	D533, L534, D535, D678, A679	D533, L534, D535, D678, A679	5/5
	11392	E573, Y667, D670, E672	E573, Y667, D670, E672	4/4
1OIL	4677	D242, D288, Q292, V296	D242, D288, Q292, V296	4/4
	4678	D242, D288, Q292, V296	D242, D288, Q292, V296	4/4
1SBF	1735	D126, F128, N130, D133	D126, F128, N130, D133	4/4
1SRA	1264	D222, P225, D227, Y229, E234	D222, P225, D227, Y229, E234	4/4
	1265	D257, D259, D261, Y263, E268	D257, D259, D261, Y263, E268	5/5
	1266	P241, I243, E246	-	0/3
1TCM	10513	D27, N29, N32, N33, G51, D53	D27, N29, N32, N33, G51, D53	4/4
	10514	N139, I190, D199, H233	N139, I190, D199, H233	4/4
	10515	D27, N29, N32, N33, G51, D53	D27, N29, N32, N33, G51, D53	5/5
	10616	N139, I190, D199, H233	N139, I190, D199, H233	4/4
1TN3	1068	D116, E120, G147, E150, N151,	D116, E120, G147, E150, N151,	5/5
	1069	Q143, D145, E150, D165	Q143, D145, E150, D165	4/4
2FIB	2036	D318, D320, F322, G324	D318, D320, F322, G324	4/4
2TEP	7081	D123, Y125, N127, D132	D123, Y125, N127, D132	4/4
	7083	D123, Y125, N127, D132	D123, Y125, N127, D132	4/4
	7085	D123, Y125, N127, D132	D123, Y125, N127, D132	4/4
	7087	D123, Y125, N127, D132	D123, Y125, N127, D132	4/4
4LIP	4669	D242, D288, Q292, V296	D242, D288, Q292, V296	4/4
	4670	D242, D288, Q292, V296	D242, D288, Q292, V296	4/4
1C9M	1897	G2, D41, L75, N77, I79, V81	G2, D41, L75, N77, I79, V81	6/6
	1898	A169, Y171, A174, G195 D197	A169, A174, D197	3/5
2SCP	2739	D16, D18, D20, A22, D27	D16, D18, D20, A22, D27	5/5
	2740	D104,N106, D108, N110, E115	D104,N106, D108, N110, E115	5/5
	2741	D138, N140, D142, L144, E149	D138, N140, D142, L144, E149	5/5
	2742	D16, D18, D20, A22, D27	D16, D18, D20, A22, D27	5/5
	2743	D104,N106, D108, N110, E115	D104,N106, D108, N110, E115	5/5
	2744	D138, N140, D142, L144, E149	D138, N140, D142, L144, E149	5/5
1TVG	1086	N29, D32, N34, T37, H130	N29, D32, N34, T37, H130	5/5
2GGM	2682	D114, D116, T118, K120, N125	D114, D116, T118, K120, N125	5/5
	2683	D150, D152, D154, E156, E161	D150, D152, D154, E156, E161	5/5
	2684	D114, D116, T118, K120, N125	D114, D116, T118, K120, N125	5/5
	2685	D150, D152, D154, E156, E161	D150, D152, D154, E156, E161	5/5
3FIA	780	D66, N68, D70, R72, E77	D66, N68, D70, R72, E77	5/5
1K9K	1417	S20, E23, D25, T28, E33	S20, E23, D25, T28, E33	5/5
	1418	D61, D63, D65, E67, E72	D61, D63, D65, E67, E72	5/5
	1423	S20, E23, D25, T28, E33	S20, E23, D25, T28, E33	5/5
	1424	D61, D63, D65, E67, E72	D61, D63, D65, E67, E72	5/5
1DAN*	4723	D46, G47, N49, D63, N64	D46, G47, N49, D63, N64	5/5
	4724	CGU	-	
	4725	CGU	-	
	4726	CGU	-	
	4727	CGU	-	
	4728	CGU	-	
	4729	CGU	-	
	4730	CGU	-	
	4731	D70, D72, E75, E80	D70, D72, E75, E80	4/4
1MHO	713	S18, E21, D23, K26, E31	S18, E21, D23, K26, E31	5/5
	714	D61, D63, D65, E67, E72	D61, D63, D65, E67, E72	5/5
1EDH	3230	E11, E69, D100, Q101, D103	E11, E69, D100, Q101, D103	5/5
	3231	E11, D67, E69, D103	E11, D67, E69, D103	4/4

2EGD	3232	E11, N12, D67, E69, D103	E11, N12, D67, E69, D103	5/5
	3234	E11, E69, D100, Q101, D103	E11, E69, D100, Q101, D103	5/5
	3235	E11, D67, E69, D103	E11, D67, E69, D103	4/4
	3236	E11, N12, D67, E69, D103	E11, N12, D67, E69, D103	5/5
	1386	A24, E27, R29, S32, E37	A24, E27, R29, S32, E37	5/5
	1387	D64, N66, D68, E70, E75	D64, N66, D68, E70, E75	5/5
	1388	A24, E27, R29, S32, E37	A24, E27, R29, S32, E37	5/5
	1389	D64, N66, D68, E70, E75	D64, N66, D68, E70, E75	5/5

[a]: metal identification number in PDB file. [b]: the correctly predicted ligands over documented ligands

Table S7. Prediction results on the NMR training dataset.

Protein	ID	M[a]	L[b]	Real Ligands	Predicted Ligands	R[c]
Epidermal growth factor receptor pathway substrate 15	1C07	20	95	D28, D30, D32, F34, E39	D28, D30, D32, F34, S36, E39	5/5
Calcium-binding protein NCS-1	1FPW	20	190	D73, D75, N77, F79, E84	D73, D75, N77, F79, E84	5/5
				D109, N111, D113, Y115, E120	D109, N111, D113, Y115, E120	5/5
				D157, N159, D161, Y163, E168	D157, N159, D161, Y163, E168	5/5
Troponin C	1TNW	23	162	D30, D32, G34, D36, E41	D30, G33, G34, D36, E41	5/5
				D66, D68, S70, T72, E77	D66, T72, D74, E77	3/5
				D106, N108, D110, F112, E117	D106, N108, D114, E117	4/5
				D142, N144, D146, R148, E153	D142, D146, R148, E150	4/5
Calmodulin	2BBM	1	148	D20, D22, D24, T26, E31	D24, T26, T28, E31	4/5
				D56, D58, N60, T62, E67	D56, D58, T62, E67	4/5
				D93, D95, N97, Y99, E104	D93, D95, N97, Y99, E104	5/5
				N129, D131, D133, D135, E140	N129, D131, D133, E140	4/5
Calbindin D9K	2BCB	32	75	A14, E17, D19, Q22, E27	A14, E17, D19, Q22, E27	5/5
				D54, N56, D58, E60, E65	D56, N58, E60, E65	4/5
Parvalbumin	2PAS	9	109	D51, D53, S55, F57, E62	D51, D53, S55, F57, E62	5/5
				D90, D92, D94, K96, E101	D90, D92, D94, K96, E101	5/5

[a]: number of structures in the ensembles. [b]: number of the residues in proteins. [c]: number of correctly predicted ligands over number of documented ligands.

Table S8. Prediction results on the NMR testing dataset.

Protein	X-ray	Chain	NMR	Chain	Identity	M[a]	L[b]	Real Ligands	Predicted Ligands	R[c]
SERINE PROTEASE PB92	1C9M	A	1AH2	A	98%	18	269	G2, D40, L75, N77, I79, V81	S3, D40, L75, G78, I79	4/6
								A163, Y165, A168, G189, D191	A163, R164, Y165, A168, D191	4/5
Nereis diversicolor sarcoplasmic calcium-binding protein (NSCP)	2SCP	B	1Q80	A	100%	17	174	D16, D18, D20, A22, D27	D16, D18, A22, D27	4/5
								D104, N106, D108, N110, E115	D104, N106, D108, N110, E115	5/5
								D138, N140, D142, L144, E149	D138, N140, L144, E149	4/5
human protein HSPCO34	1TVG	A	1XPW	A	100%	20	153	N29, D32, N34, T37, H129	N29, D32, N34, T37, H129	5/5
the human centrin 2 in complex with a 17 residues peptide (P1-XPC) from xeroderma pigmentosum group C protein	2GGM	A	2A4J	A	100%	20	96	D114, D116, T118, K120, N125	D114, T118, K120	3/5
								D150, D152, D154, E156, E161	-	0/5
human intersectin-1 protein**	3FIA	A	2KHN	A	99%	20	121	D76, N78, D80, R82, E87	D76, N78, R82, E87	4/5
Staphylococcal nuclease	1SNC	A	1JOQ	A	99%	30	149	D21, D40, T41, THP151	D21, D40, T41	3/3
Human blood coagulation FVII	1DAN	L	1F7E	A	100%	20	46	D46, G47, N49, D63, N64	D46, N49, D63	3/5
S100B	1MHO	A	1UWO	A, B	96%	20	91	S18, E21, D23, K26, E31	S18, E21, D23, K26	4/5
								D61, D63, D65, E67, E72	D61, D63, E67, E72	4/5
								S18, E21, D23, K26, E31	E21, D23, K26, E31	4/5
								D61, D63, D65, E67, E72	D61, D63, E67, E72	4/5
Epithelial cadherin	1EDH	A	1SUH	A	100%	20	146	E11, E69, D100, Q101, D103	E11, E69, Q101, D103	4/5
								E11, D67, E69, D103	E11, N12, D67, E69	3/4
S100A13	2EGD	A	2K8M	B,C	100%	20	98	A24, E27, R29, S32, E37	A24, G28, R29, S32, E37	5/5
								D64, N66, D68, E70, E75	D64, D68, E70, E75	4/5
								A24, E27, R29, S32, E37	A24, G28, R29, S32, E37	5/5
								D64, N66, D68, E70, E75	D64, D68, E70, E75	4/5

[a]: number of structures in the ensembles. [b]: number of the residues in proteins. [c]: number of correctly predicted ligands over number of documented ligand.

We modeled 199 equilibrium structures of calcium-binding protein (3ICB.pdb, 75 residues) by molecular dynamics using AMBER software package as follows. The two documented calcium ions were removed from the original protein structure. The resulting structure was solvated in a water box, followed by a short minimization, 60 ps of heating, 60 ps of density equilibration, and 500 ps of constant pressure equilibration at 300K. In production phase, 199 structures were generated in 2.4 ns. The strategy follows Walker and Steinbrecher http://ambermd.org/tutorials/advanced/tutorial3/. The RMSD below is calculated by aligning each modeled structure to original structure using CLICK.

Table S9. Testing on 199 conformations (PDB: 3ICB) modeled by molecular simulation.

Model[a]	RMSD[b]	R[c]
1	1.67	2/2
2	1.77	2/2
3	1.61	2/2
4	1.69	2/2
5	1.56	2/2
6	1.71	2/2
7	1.68	2/2
8	1.77	2/2
9	1.89	2/2
10	1.89	2/2
11	1.72	2/2
12	1.91	2/2
13	1.70	2/2
14	1.84	2/2
15	2.00	2/2
16	1.77	2/2
17	1.70	2/2
18	1.75	2/2
19	1.79	2/2
20	1.68	2/2
21	1.83	2/2
22	1.61	2/2
23	1.73	2/2
24	1.74	2/2
25	1.70	2/2
26	1.83	2/2
27	1.76	2/2
28	1.83	2/2
29	1.90	2/2
30	1.80	2/2
31	1.58	2/2
32	1.76	2/2
33	1.91	2/2
34	1.74	2/2
35	1.66	2/2
36	1.69	2/2
37	1.89	2/2
38	1.77	2/2
39	1.71	2/2

40	1.69	2/2
41	1.86	2/2
42	1.74	2/2
43	1.68	2/2
44	1.64	2/2
45	1.75	2/2
46	1.58	2/2
47	1.54	2/2
48	1.65	2/2
49	1.67	2/2
50	1.70	2/2
51	1.68	2/2
52	1.86	2/2
53	1.74	2/2
54	1.53	2/2
55	1.48	1/2
56	1.56	2/2
57	1.79	2/2
58	1.82	2/2
59	1.60	2/2
60	1.67	2/2
61	1.62	2/2
62	1.78	2/2
63	1.60	2/2
64	1.59	2/2
65	1.68	2/2
66	1.81	2/2
67	1.83	2/2
68	1.61	2/2
69	1.68	2/2
70	1.76	2/2
71	1.74	2/2
72	1.79	2/2
73	1.70	2/2
74	1.57	2/2
75	1.64	2/2
76	1.71	2/2
77	1.69	2/2
78	1.60	2/2
79	1.63	2/2
80	1.84	2/2
81	1.53	2/2
82	1.64	2/2
83	1.62	2/2
84	1.73	2/2
85	1.80	2/2
86	1.68	2/2
87	1.52	2/2
88	1.55	1/2
89	1.73	2/2
90	1.69	2/2
91	1.66	2/2
92	1.70	2/2
93	1.79	2/2
94	1.66	2/2
95	1.70	2/2
96	1.77	2/2
97	1.78	2/2
98	1.59	2/2
99	1.54	2/2
100	1.57	2/2
101	1.55	2/2
102	1.52	2/2
103	1.52	2/2
104	1.45	2/2

105	1.58	2/2
106	1.53	2/2
107	1.64	2/2
108	1.64	2/2
109	1.61	2/2
110	1.78	2/2
111	1.67	2/2
112	1.72	2/2
113	1.50	2/2
114	1.44	2/2
115	1.54	2/2
116	1.51	2/2
117	1.66	2/2
118	1.85	2/2
119	1.84	2/2
120	1.73	2/2
121	1.78	2/2
122	1.66	1/2
123	1.81	1/2
124	1.70	1/2
125	1.55	2/2
126	1.61	2/2
127	1.60	2/2
128	1.52	2/2
129	1.57	2/2
130	1.64	2/2
131	1.54	2/2
132	1.55	2/2
133	1.45	2/2
134	1.61	2/2
135	1.58	2/2
136	1.49	2/2
137	1.58	2/2
138	1.53	2/2
139	1.63	2/2
140	1.52	2/2
141	1.68	2/2
142	1.61	2/2
143	1.51	2/2
144	1.62	1/2
145	1.54	2/2
146	1.50	2/2
147	1.47	2/2
148	1.61	2/2
149	1.61	2/2
150	1.58	2/2
151	1.59	2/2
152	1.55	2/2
153	1.51	2/2
154	1.43	2/2
155	1.43	1/2
156	1.45	2/2
157	1.45	2/2
158	1.48	2/2
159	1.54	2/2
160	1.64	2/2
161	1.61	2/2
162	1.52	2/2
163	1.53	2/2
164	1.46	2/2
165	1.52	2/2
166	1.54	1/2
167	1.70	2/2
168	1.78	2/2
169	1.76	1/2

170	1.77	1/2
171	1.68	1/2
172	1.75	2/2
173	1.65	2/2
174	1.61	2/2
175	1.59	2/2
176	1.54	2/2
177	1.51	1/2
178	1.52	2/2
179	1.50	2/2
180	1.58	2/2
181	1.70	2/2
182	1.65	2/2
183	1.62	2/2
184	1.61	1/2
185	1.62	1/2
186	1.70	2/2
187	1.67	2/2
188	1.53	2/2
189	1.73	2/2
190	1.58	2/2
191	1.62	2/2
192	1.75	2/2
193	1.69	2/2
194	1.65	2/2
195	1.59	2/2
196	1.61	2/2
197	1.61	2/2
198	1.50	2/2
199	1.57	2/2

[a] Model number. [b] RMSD between model and experimental X-ray structure. [c] the correctly predicted sites over total predicted sites.

Non-redundant metal binding proteins resources:

Protein structures in Tables S9 and S10 are from the published dataset .[35]

Protein structures in Table S11 are from a survey of PDB with the following criteria.

1. X-ray resolution is better than 2.5 Å.

2. There is a Protein chain but not any DNA or RNA or Hybrid.

3. Structures contain desired metal.

4. Sequences with more than 30% similarity are removed.

Mis-classification criteria:

We consider that the algorithm mis-classified a binding site if it predicted a Ca^{2+} within 3.0 Å (usually the metal oxygen distance is less than this value) of the documented non-Ca^{2+} metal. [32,39]

Classification as “Other” indicates that the predicted Ca^{2+}-binding site may represent true binding sites that have not yet been identified as such, or have been identified as binding sites for other metal divalent ions that may bind different metals, which may include Ca^{2+}.

Table S10. Testing on Mg^{2+}-binding proteins (X-ray structures).

PDB[a]	Res[b]	Protein[c]	Chain[d]	Mg#[e]	Mis-classified[f]
1CMC	1.8	Met repressor (metj)	A,B	1693	No
				1721	No
1EBH	2.2	Enolase	A,B	6631	No
				6633	No
1XLB	2.3	D-xylose isomerase	A	6055	No
1CHN	1.6	Chey	A	968	No
1EO3	1.9	Restriction enzyme ecoRV	A,B	4229	No
				4230	No
				4235	No
				4236	No
1VSD	1.9	Integrase	A	1129	No
1MUS	2.5	Adenine phosphoribosyltransferase	A,B	4435	No
				4436	No
1QB7	1.9	Xanthine-guanine phosphoribosyltransferase	A	1857	No
1EYJ	2.1	Fructose-1,6-bisphosphatase	A,B	5011	No
				5056	No
2UAG	1.7	D-glutamate ligase	A	3247	No
			A	3248	No
3PRN	1.9	Porin	A	2203	No
1HBN	1.1	Methyl-coenzyme m reductase	B	19432	No
			D	19557	No
			E	19570	No
2TCT	2.1	Tetracycline repressor	A	1575	No
1LUC	1.5	Bacterial luciferase	A	5096	No
			B	5106	No
1KQP	1.0	Nh(3)-dependent nad(+) synthetase	B	8739	No
1NG1	2.0	Signal sequence recognition protein FFH	A	2279	No
1BL3	2.0	Integrase	B	3445	No
1NUL	1.8	Xanthine-guanine phosphoribosyltransferase	A	2159	No
2UAG	1.7	D-glutamate ligase	A	3247	No
1IDE	2.5	Isocitrate dehydrogenase	A	3881	No
1JIV	2.0	DNA beta-glucosyltransferase	A	2871	No
			A	2872	No
1DOZ	1.8	Ferrochelatase	A	2490	No
1G8T	1.1	Nuclease sm2 isoform	A	3878	No
1A73	1.8	Intron 3 (i-ppo) encoded endonuclease	A	3353	No
1FWK	2.1	Homoserine kinase	D	9162	No
1JKK	2.4	Death-associated protein kinase	A	2247	No
1LDF	2.1	Glycerol uptake facilitator	A	1936	No
1OBW	2.1	Inorganic pyrophosphatase	A,B,C	4141	Other
				4142	No
				4143	No
				4144	Other
				4145	No
				4146	No
				4147	Other
1KCZ	1.9	Beta-methylaspartase	A,B	6431	Other
				6440	Other
1RK2	1.8	Ribokinase	A,B,C,D	8992	Other
				9035	Other
				9078	Other
				9121	Other

[a] PDB code. [b] PDB resolution. [c] Protein name. [d] Chain number. [e] metal identification number in PDB file. [f] mistakenly classified Mg^{2+}-binding site as Ca^{2+}-binding site.

Table S11. Testing on Zn^{2+}-binding proteins (X-ray structures).

PDB[a]	Res[b]	Protein[c]	Chain[d]	Zn#[e]	Mis-classified[f]
1FWZ	2.3	Diphtheria toxin repressor	A	1588	No
1CY5	1.3	Apoptotic protease activating factor 1	A	749	No
				750	No
				752	No
1WEJ	1.8	E8 antibody	A	4170	No
1E67	2.1	Azurin	A,B,C,D	3901	No
				3906	No
				3907	No
				3908	No
1GS8	1.9	Nitrite reductase	A	2590	No
				2591	No
1F5F	1.7	Sex hormone-binding globulin	A	1369	No
				1370	No
1GI4	1.3	Beta-trypsin	A	3369	No
2CBA	1.5	Carbonic anhydrase	A	2081	No
1F3Z	1.9	Glucose-specific phosphocarrier	A	1109	No
1C8Y	2.0	Endo-beta-n-acetyl-glucosaminidase H	A	2015	No
4ENL	1.9	Enolase	A	3291	No
1I6N	1.8	Loli protein	A	2231	No
1IM5	1.6	Pyrazinamidase	A	1439	No
1VSH	1.9	Integrase	A	1129	No
				1130	No
				1131	No
1NOY	2.2	DNA polymerase	A	5953	No
2CTB	1.5	Carboxypeptidase A	A	2452	No
1TOA	1.8	Periplasmic binding protein	A,B	4295	No
				4302	No
1A2P	1.5	Barnase	A,B,C	2628	No
				2629	No
				2630	No
1EU3	1.6	Superantigen Smez-2	A,B	3419	No
				3436	No
1EWC	1.9	Enterotoxin H	A	1733	No
1EU4	2.5	Superantigen spe-H	A	1668	No
1AST	1.8	Astacin	A	1593	No
1ZFP	1.8	Growth factor receptor binding protein	E	870	No
1K4P	1.0	3,4-dihydroxy-2-butanone 4-phosphate synthase	A	1643	No
1K9Z	1.5	Halotolerance protein HAL2	A	2731	No
				2732	No
				2735	No
				2733	No
1CNQ	2.2	Fructose-1,6-bisphosphatase	A	2572	No
1KSP	2.3	Klenow fragment	A	4817	No
3IVE	2.0	Immunoglobulin	A	893	No
1M5E	1.4	Glutamate receptor 2	A	6150	No
1L7O	2.2	Phosphoserine phosphatase	B	3208	No
8RNT	1.8	Ribonuclease T1	A	779	No
1XLL	2.5	D-xylose isomerase	A,B	6057	Other
				6058	No
				6059	Other
				6060	No

[a] PDB code. [b] PDB resolution. [c] Protein name. [d] Chain number. [e] metal identification number in PDB file. [f] mistakenly classified Zn^{2+}-binding site as Ca^{2+}-binding site.

Table S12. Testing on Pb^{2+}-binding proteins (X-ray structures).

PDB[a]	Res[b]	Protein[c]	Chain[d]	Pb#[e]	Mis-classified[f]
1E9N	2.20	DNA-lyase	A,B	4339	No
				4340	No
				4341	No
				4342	No
1FJR	2.3	Methuselah ectodomain	A,B	3120	No
				3121	No
				3169	No
				3170	No
1NA0	1.60	Designed protein CTPR3	A,B	1969	No
				1970	No
				1975	No
				1976	No
				1977	No
1QNV	2.5	5-aminolaevulinic acid dehydratase	A	2548	No
				2549	No
1SN8	2.00	Ribonuclease E	A,B	1330	No
				1331	No
1SYY	1.7	Ribonucleoside-diphosphate reductase	A	2617	No
1XXA	2.20	Arginine repressor	A~F	3245	No
				3246	No
				3259	No
				3284	No
1ZHY	1.6	KES1 protein	A	3516	No
				3517	No
2CH7	2.5	Methyl-accepting chemotaxis protein	A,B	4633	No
				4634	No
2FJ9	1.6	Acyl-CoA-Binding protein	A	710	No
2FP1	1.55	Chorismate mutase	A,B	2705	No
				2706	No
2OQ1	1.9	Tyrosine-protein kinase	A,B	2200	No
2QD5	2.3	Ferrochelatase	A,B	5845	No
				5846	No
				5965	No
				5966	No
2QKL	2.3	Hydrolase	A,B	1772	No
3EC8	2.6	FLJ10324	A	1083	No
				1084	No
3FHH	2.6	Outer membrane heme receptor ShuA	A	4778	No
				4779	No
				4780	No
				4781	No
1HD7	1.95	DNA-lyase	A	2072	Other
2G0A	2.35	Cytosolic 5’-nucleotidase III	A,B	4655	Other
				4671	Other
2O3C	2.30	APEX nuclease 1	A,B	6648	Other
				6649	No
				6650	Other

[a] PDB code. [b] PDB resolution. [c] Protein name. [d] Chain number. [e] metal identification number in PDB file. [f] mistakenly classified Pb^{2+}-binding site as Ca^{2+}-binding site.

Table S13. Testing on a negative control dataset (X-ray structures).

PDB[a]	Protein[b]	FN[c]
1DTS	Dethiobiotin synthase	0
1L68	Lysozyme	0
1PTX	Scorpion toxin II	0
1VCC	DNA topoisomerase I	0
1WBA	Winged bean albumin 1	0
2ENG	Endoglucanase V	0
2YLE	Human spir-1 kind fsi domain in complex with the fsi peptide	0
3O5F	Fk1 domain of FKBP51	0
3OQ7	Multidrug-Resistant Clinical Isolate 769 HIV-1 Protease Variants	0
1IQR	DNA photolyase	0
1IUG	Aspartate aminotransferase which belongs to subgroup IV	0
1IZ0	Quinone Oxidoreductase	0
1J27	Hypothetical protein, TT1725	0
1J3M	Conserved hypothetical protein TT1751	0
1JJF	Feruloyl esterase domain of the cellulosomal xylanase z of clostridium thermocellum	0
1TCA	Lipase	2
2OLB	Oligo-peptide binding protein	4
1TTB	Transthyretin	2
1BDM	Malate Dehydrogenase	5
1K4N	Protein EC4020	3
2AQJ	Tryptophan 7-halogenase (PrnA)	5
1ISO	Isocitrate dehydrogenase	2
1SGV	Trna psi55 pseudouridine synthase (trub)	3

[a] PDB code. [b] Protein name. [c] Number of False Negative predictions.