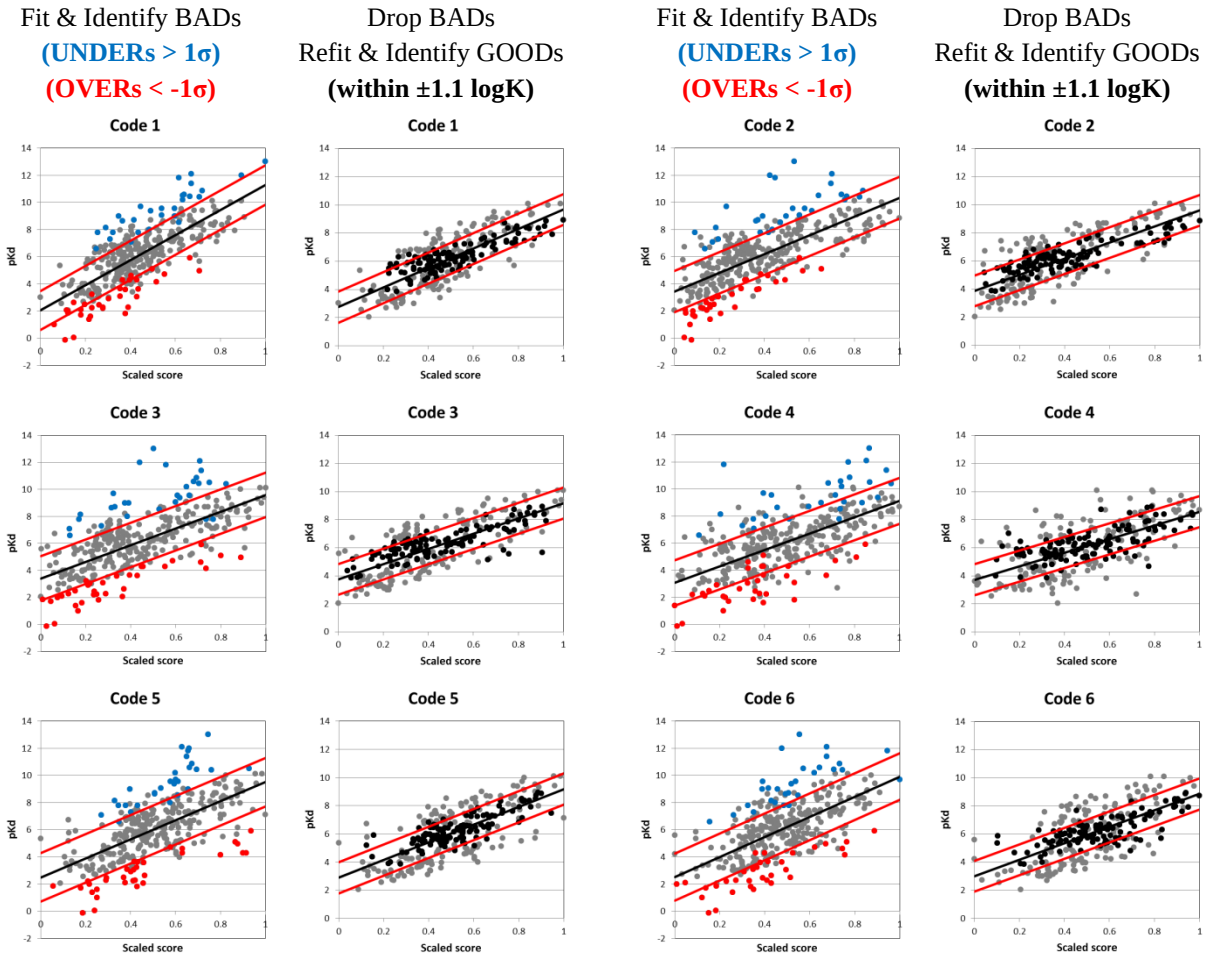
Supplemental Information for "*CSAR Benchmark Exercise of 2010: Combined evaluation across all submitted scoring functions*" by Smith, Dunbar, Ung, Esposito, Yang, Wang, and Carlson.

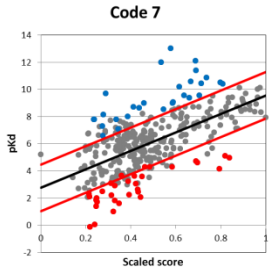
1. Table of 10 proteins with ligand series and the performance of each of the 17 core codes on ranking
2. Discussion of methods and metrics for identifying the GOOD complexes, including figures S1 and S2
3. A complete listing of GOOD and BAD complexes.

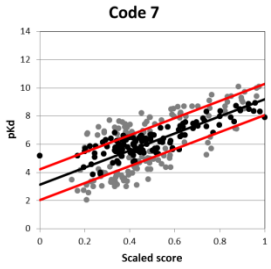
Table S1. Performance of each code to estimate the relative ranking across individual protein families ordered by range in $pK_{d/i}$. The normalization of the scores results in correlations having positive numbers for Pearson R, Spearman ρ, and Kendall τ. Anti-correlation results in negative numbers (ie, a -1 value is *not* good). Columns shaded gray have a significant number of outliers in the protein family.

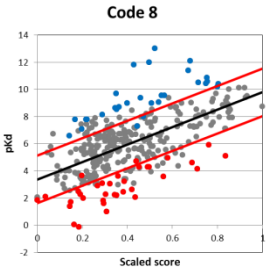
Protein Family	HIVp (wt)			PTP1B			HIVp (L63P)			GluRζ1			LeuT			TGT			FXa			CA II			uPA			GluR6		
Range of pKd/i	4.86			4.25			3.89			3.50			3.14			2.62			2.19			1.75			1..74			1.34		
# complexes	11			8			11			3			3			6			7			4			3			4		
	R	**ρ**	**τ**	**R**	**ρ**	**τ**	**R**	**ρ**	**τ**	**R**	**ρ**	**τ**	**R**	**ρ**	**τ**	**R**	**ρ**	**τ**	**R**	**ρ**	**T**	**R**	**ρ**	**τ**	**R**	**ρ**	**τ**	**R**	**ρ**	**τ**
Code 1	-0.01	0.08	0.09	0.85	0.46	0.25	0.62	0.66	0.56	-0.99	-1	-1	0.97	1	1	0.55	-0.12	-0.28	0	0.21	0.14	0.25	0.80	0.67	0.45	0.50	0.33	0.86	0.78	0.60
Code 2	-0.17	-0.14	-0.13	0.97	0.81	0.64	0.66	0.60	0.45	-0.99	-1	-1	0.94	1	1	-0.24	-0.03	-0.14	0.09	0	-0.05	0.80	1	1	-0.45	-0.50	-0.33	0.86	0.95	0.91
Code 3	0.34	0.45	0.38	0.93	0.60	0.43	0.49	0.41	0.35	-0.83	-0.50	-0.33	0.96	1	1	-0.10	-0.03	-0.14	0.70	0.75	0.62	-0.59	-0.20	0	-0.72	-1	-1	1	0.95	0.91
Code 4	0.21	0.22	0.15	0.70	0.76	0.57	-0.25	-0.13	-0.07	-0.92	-0.50	-0.33	0.87	1	1	-0.38	0.06	0.14	-0.02	0.39	0.24	0.19	0.74	0.55	0.33	-0.50	-0.33	0.83	0.74	0.55
Code 5	-0.20	0	-0.05	0.39	0.10	0	0.65	0.53	0.42	-1	-1	-1	0.99	1	1	0.17	0.32	0.28	0.35	0.43	0.33	0.39	0.80	0.67	0.20	0.50	0.33	1	0.95	0.91
Code 6	0.25	0.35	0.24	0.92	0.57	0.5	0.61	0.35	0.24	-1	-1	-1	0.99	1	1	-0.09	0.06	0	0.52	0.57	0.43	-0.19	0.40	0.33	-0.51	-0.50	-0.33	0.95	0.95	0.91
Code 7	0.59	0.69	0.53	0.95	0.86	0.71	0.54	0.52	0.42	0.30	0.50	0.33	0.86	1	1	0.11	0.12	0	0.24	-0.14	-0.24	0.31	-0.20	0	-0.95	-1	-1	0.51	0.74	0.55
Code 8	0.50	0.28	0.24	0.70	0.55	0.36	0.77	0.44	0.35	-0.95	-1	-1	0.97	1	1	0.14	0.03	0	0.15	0.46	0.33	-0.86	-0.40	-0.33	-0.44	-0.50	-0.33	1	0.95	0.91
Code 9	0.13	0.10	0.02	0.67	0.62	0.43	0.85	0.80	0.64	-0.79	-0.50	-0.33	0.95	1	1	-0.33	-0.12	-0.28	0.67	0.71	0.52	-0.68	-0.40	-0.33	0.55	0.50	0.33	0.99	0.95	0.91
Code 10	-0.15	-0.31	-0.27	0.94	0.79	0.57	0.74	0.41	0.31	-0.98	-1	-1	0.98	1	1	-0.43	-0.23	-0.28	0.38	0.21	0.14	0.13	0.80	0.67	-0.91	-0.50	-0.33	0.98	0.95	0.91
Code 11	0.10	0.15	0.09	-0.15	-0.05	-0.04	0.66	0.45	0.35	-0.86	-0.50	-0.33	0.94	1	1	-0.33	-0.03	-0.14	0.47	0.60	0.47	0.78	1	1	-0.93	-0.50	-0.33	0.86	0.74	0.55
Code 12	0.56	0.41	0.27	0.90	0.76	0.57	-0.24	-0.35	-0.27	-1	-1	-1	0.98	1	1	0.20	-0.03	-0.14	0.67	0.71	0.52	-0.05	0.40	0.33	-0.91	-1	-1	0.88	0.74	0.55
Code 13	-0.21	-0.17	-0.09	0.93	0.69	0.57	0.62	0.69	0.49	-0.92	-1	-1	0.93	1	1	0.05	0.06	0	0.33	0.43	0.24	0.17	0.80	0.67	0.96	0.50	0.33	0.95	0.95	0.91
Code 14	0.18	0.12	0.13	0.23	0.19	0.14	0.54	0.47	0.31	0.16	0	0	0.96	1	1	0.12	0.06	0	0.41	0.64	0.43	0.72	0.80	0.67	-1	-1	-1	0.74	0.32	0.18
Code 15	0.11	0.33	0.27	0.80	0.67	0.50	0.52	0.61	0.49	-1	-1	-1	0.92	1	1	-0.75	-0.84	-0.69	0.52	0.32	0.24	-0.35	-0.40	-0.33	-0.99	-0.50	-0.33	0.98	0.95	0.91
Code 16	-0.27	0.05	0.02	0.87	0.69	0.57	0.68	0.46	0.35	0.32	0.50	0.33	0.91	1	1	-0.26	-0.17	-0.14	0.07	0.21	0.14	0.13	0.60	0.33	0.50	0.50	0.33	0.47	0.74	0.55
Code 17	0	0.28	0.16	0.87	0.69	0.50	0.13	0.25	0.20	-0.95	-1	-1	0.90	1	1	-0.26	-0.12	-0.28	0.22	-0.11	-0.14	0.50	0.60	0.66	-0.82	-1	-1	0.01	0.32	0.18
median	**0.11**	**0.15**	**0.13**	**0.87**	**0.67**	**0.50**	**0.62**	**0.46**	**0.35**	**-0.95**	**-1**	**-1**	**0.95**	**1**	**1**	**-0.1**	**-0.03**	**-0.14**	**0.35**	**0.41**	**0.24**	**0.17**	**0.60**	**0.55**	**-0.62**	**-0.50**	**-0.33**	**0.88**	**0.95**	**0.91**
Codes Retrained on the CSAR-NRC Set (correlation to full set is noted)																														
Code 4 (R^2=0.86)	0.70	0.64	0.38	0.87	0.62	0.57	0.92	0.89	0.71	0.84	0.50	0.33	0.98	1	1	0.83	0.84	0.69	0.68	0.71	0.52	0.73	0.80	0.67	1	1	1	1	0.95	0.91
Code 6 (R^2=0.42)	0.59	0.63	0.49	0.95	0.67	0.50	0.70	0.58	0.48	-1	-1	-1	0.89	1	1	-0.14	-0.12	-0.28	0.27	0	-0.05	-0.52	-0.40	-0.33	-0.99	-1	-1	0.84	0.74	0.55
Code 16 (R^2=0.47)	-0.37	-0.19	-0.20	0.93	0.60	0.50	0.62	0.58	0.42	-0.59	-0.50	-0.33	0.89	1	1	0.44	0.81	0.69	-0.17	0	-0.05	0.42	0.60	0.33	0.60	0.50	0.33	0.50	0.74	0.55
Code A (R^2=0.54)	-0.02	-0.06	-0.09	0.84	0.69	0.57	-0.06	0.11	0.09	-0.70	-0.50	-0.33	0.91	1	1	-0.77	-0.41	-0.41	0.62	0.71	0.52	0.15	0.60	0.33	0.99	1	1	0.96	0.95	0.91
Code B (R^2=0.47)	-0.02	-0.19	-0.20	0.36	0.31	0.21	0.40	0.40	0.35	0.93	1	1	0.87	1	1	-0.41	-0.14	-0.14	0.27	0.07	0.05	0.63	0.80	0.67	0.96	0.50	0.33	0.68	0.74	0.55
median	**-0.02**	**-0.06**	**-0.09**	**0.87**	**0.62**	**0.50**	**0.62**	**0.58**	**0.42**	**-0.59**	**-0.50**	**-0.33**	**0.89**	**1**	**1**	**-0.14**	**-0.12**	**-0.14**	**0.27**	**0.07**	**0.05**	**0.42**	**0.60**	**0.33**	**0.96**	**0.50**	**0.33**	**0.84**	**0.74**	**0.55**
Yardsticks																														
Heavy atoms	0.41	0.40	0.31	0.89	0.68	0.55	0.75	0.63	0.47	-0.99	-1	-1	0.91	1	1	-0.67	-0.69	-0.62	0.66	0.55	0.46	-0.66	-0.40	-0.33	0.81	0.50	0.33	1	1	1
SlogP	0.17	0.16	0.07	0.96	0.64	0.50	-0.55	-0.45	-0.29	-0.99	-1	-1	0.95	1	1	0.06	0.35	0.29	0.49	0.78	0.60	0.96	1	1	0.73	0.50	0.33	0.99	1	1

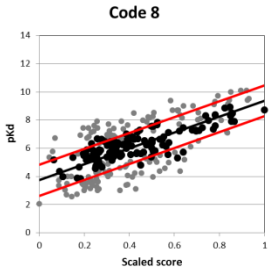
Figure S1. A graphical representation of how least-squares linear regression was used to identify BAD and GOOD complexes. In each frame, the black line is fit to the data points shown. For the identification of BADs, the red lines indicate ±σ, the standard deviation of the residuals from the linear regression. Blue points highlight complexes that were underscored by more than 1σ in at least 12 of 17 scoring functions, and red are for complexes overscored by more than 1σ in at least 12 of 17 scoring functions. For the frames showing the subsequent identification of GOODs, the data points for the 63 BAD complexes have been removed and the least-squares linear regression has been refit with the remaining 269 complexes. Black points are for complexes with residuals within 1.1 logK units (1.5 kcal/mol) in at least 12 of 17 scoring functions. Note that there are black data points outside the red lines in each GOODs graph, just like there are blue and/or red points between the red lines in each BADs graph.

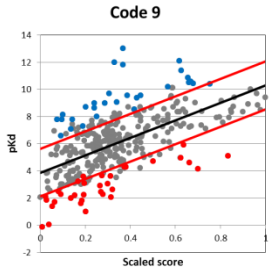


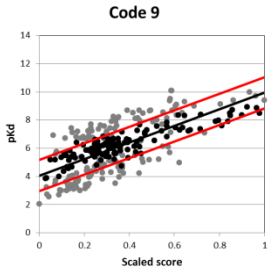

Code 7
pKd
Scaled score

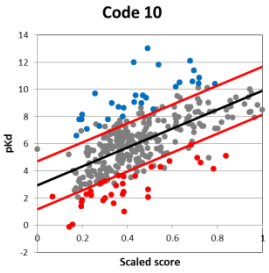

Code 7
pKd
Scaled score

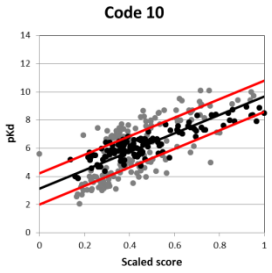

Code 8
pKd
Scaled score

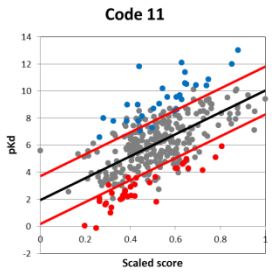

Code 8
pKd
Scaled score

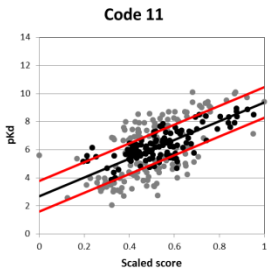

Code 9
pKd
Scaled score

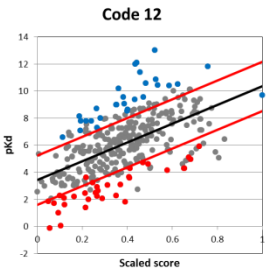

Code 9
pKd
Scaled score

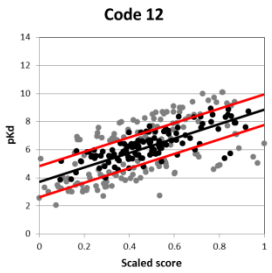

Code 10
pKd
Scaled score

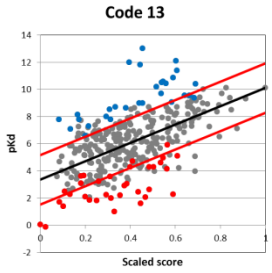

Code 10
pKd
Scaled score

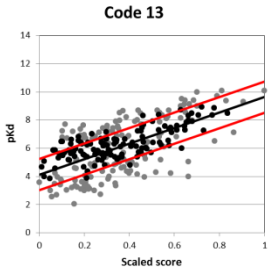

Code 11
pKd
Scaled score

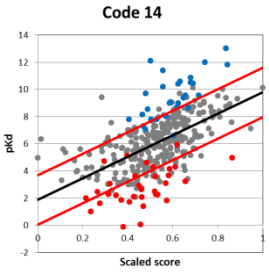

Code 11
pKd
Scaled score

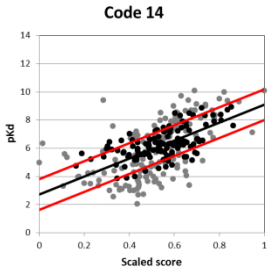

Code 12
pKd
Scaled score

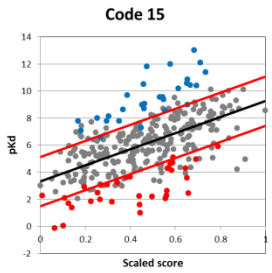

Code 12
pKd
Scaled score

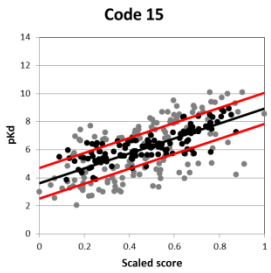

Code 13
pKd
Scaled score

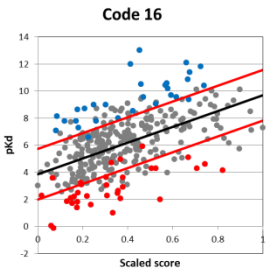

Code 13
pKd
Scaled score

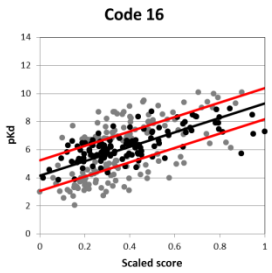

Code 14
pKd
Scaled score

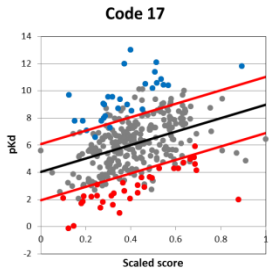

Code 14
pKd
Scaled score

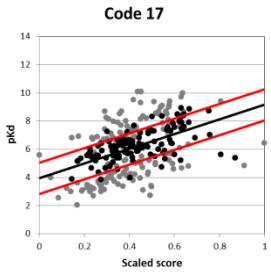

Code 15
pKd
Scaled score


Code 15
pKd
Scaled score


Code 16
pKd
Scaled score


Code 16
pKd
Scaled score


Code 17
pKd
Scaled score


Code 17
pKd
Scaled score

Other considerations behind our method for choosing GOOD complexes. We could have chosen GOOD complexes from the initial linear regression for each method. The number that could be considered well scored depended upon the cutoff used: 54 had residuals within ±0.73 logK units (1.0 kcal/mol) for ≥12 of 17 methods, 116 had residuals within ±1.1 logK (1.5 kcal/mol), and 220 were within ±1σ (ranging 1.4-2.1 logK, 1.9-2.9 kcal/mol, across the 17 methods). We did not choose the 0.73-logK cutoff because it would require more accuracy from the calculations than was possible for some of the experimental data. We chose not to base our identification of GOOD complexes upon σ from the initial fits because those cutoffs are too high; furthermore, it would include 220 complexes which are too many of the 269 complexes that did not meet the BAD criteria.

When we removed the 63 BAD complexes and refit the remaining 269 complexes for each method, the correlations show modest improvement as expected (0.01-0.17, see table below). This resulted in σ values decreasing ~0.6 logK units to range 1.0-1.5 logK (1.4-2.0 kcal/mol) for the 17 core methods. While those values of σ were more reasonable, the underlying experimental data was no longer normally distributed because many points in the tails were removed. The average (6.11 pK_d), median (6.15 pK_d), and skew (0.02) of the 269 complexes is very similar to the full set, but the maximum (10.1 pK_d), minimum (2.06 pK_d), st dev (1.71 pK_d), and – most importantly – kurtosis (-0.56!) highlight the effect. The fit linear models were still appropriate, but the statistics for a σ-based definition were no longer the same.

After the refitting, the number of systems with residuals within ±0.73 logK for ≥12 of 17 methods increased to 69 (54 in the initial fit), and those with residuals within ±1.1 logK increased to 123 (up from 116). More systems fit the linear models with less error after the refitting, leading to a more solid definition in our opinion.

Method	Initial Linear Regression of 332 Complexes R^2 (95%-confidence interval)	Remove BADs and Refit 269 Complexes R^2 (95%-confidence interval)
Code 1	0.58 (0.65-0.51)	0.62 (0.69-0.54)
Code 2	0.52 (0.59-0.44)	0.60 (0.67-0.52)
Code 3	0.45 (0.53-0.37)	0.57 (0.64-0.49)
Code 4	0.42 (0.50-0.34)	0.43 (0.52-0.34)
Code 5	0.40 (0.48-0.32)	0.53 (0.61-0.44)
Code 6	0.39 (0.47-0.31)	0.45 (0.54-0.36)
Code 7	0.38 (0.46-0.30)	0.51 (0.59-0.42)
Code 8	0.37 (0.45-0.29)	0.48 (0.56-0.39)
Code 9	0.37 (0.45-0.29)	0.50 (0.58-0.41)
Code 10	0.36 (0.44-0.28)	0.53 (0.61-0.44)
Code 11	0.35 (0.43-0.27)	0.40 (0.49-0.31)
Code 12	0.32 (0.40-0.24)	0.41 (0.50-0.32)
Code 13	0.32 (0.40-0.24)	0.42 (0.51-0.33)
Code 14	0.32 (0.40-0.24)	0.36 (0.45-0.27)
Code 15	0.31 (0.39-0.23)	0.43 (0.52-0.34)
Code 16	0.28 (0.36-0.20)	0.35 (0.44-0.26)
Code 17	0.12 (0.19-0.06)	0.21 (0.30-0.13)

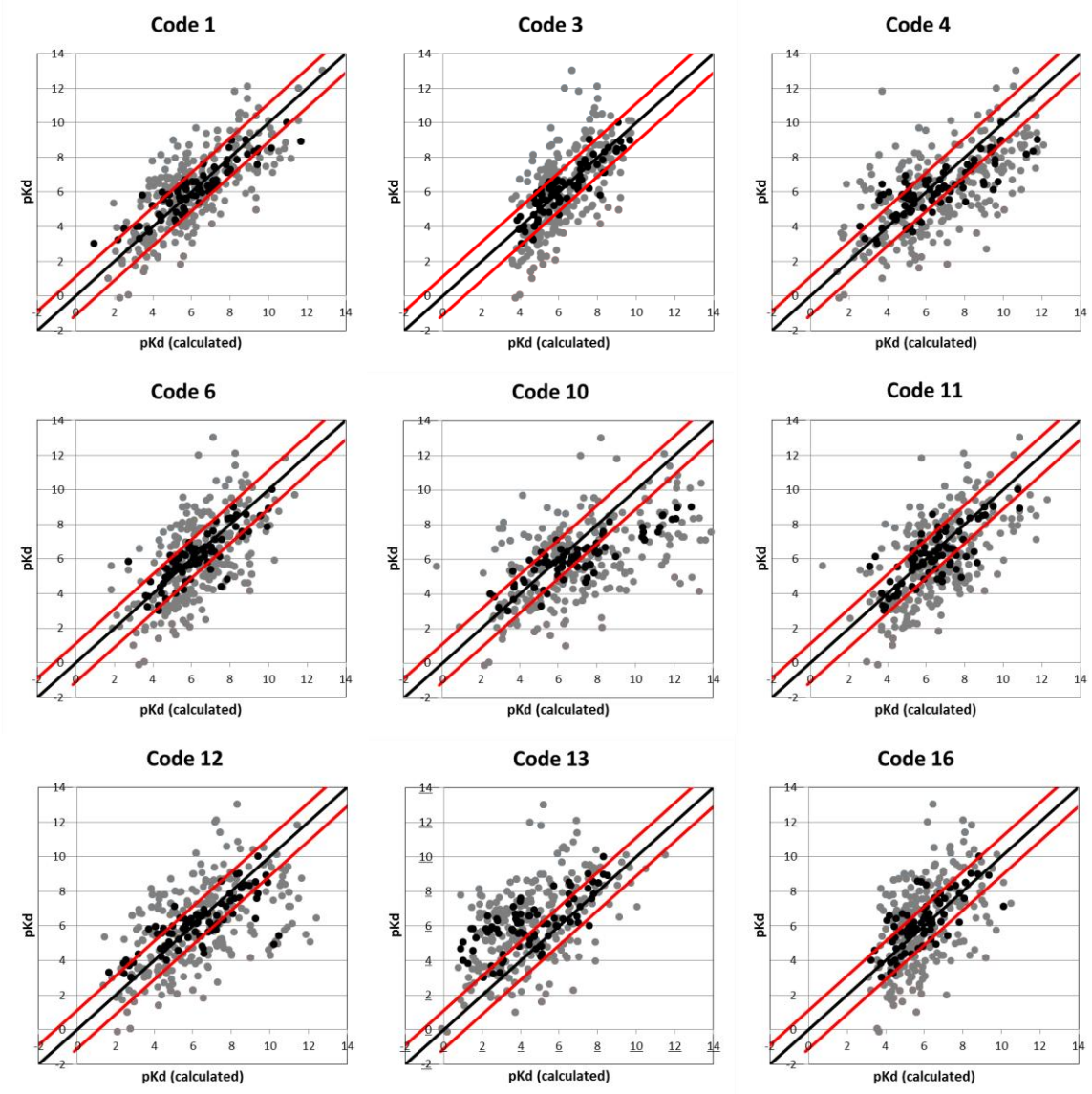


Figure S2. Scatter plots for the functions which predicted absolute binding affinity. The line y = x is in black and represents the perfect agreement between experimental and calculated affinities. The red lines indicate ±1.1 logK units from that line. Black points are GOOD complexes, defined as being within 1.1 logK (between the red lines) for at least 7 of the 9 methods. For this analysis, the identification of GOODs was based on error, and not a residual from linear regression. The BADs have not been removed from the graphs because they have no bearing on the identification of GOOD complexes.

Listing of the OVER, UNDER, and GOOD Complexes Determined from Linear Regression

34 OVER complexes from ≥12 of 17 methods determined by regression and σ of residuals
(The 13 italicized complexes below are also in the list of the 13 OVERs determined by RMSE)
Values at the end of each line are the pK_d for the complex.

1f74 N-ACETYL-NEURAMINATE LYASE 3.05
1gww N-ACETYLLACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE 4.28
1gx0 N-ACETYLLACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE 4.28
1h46 EXOGLUCANASE I 3.57
1ps3 Alpha-mannosidase II 2.28
1q6e beta-amylase 2.62
1q6g beta-amylase 2.06
1rdn MANNOSE-BINDING PROTEIN-C 1.84
1tog Aspartate aminotransferase 3.22
1tok Aspartate aminotransferase 2.47
1uto TRYPSINOGEN 2.27
1x8d Hypothetical protein yiiL 2.21
1y1m Glutamate [NMDA] receptor subunit zeta 1 1.82
1y1z Glutamate [NMDA] receptor subunit zeta 1 3.08
1y93 Macrophage metalloelastase 2.10
2aac ARAC 2.22
2add fructan 1-exohydrolase IIa 2.46
2azr Tyrosine-protein phosphatase, non-receptor type 1 3.64
2b1q hypothetical protein slr0953 1.59
2cli TRYPTOPHAN SYNTHASE ALPHA CHAIN 4.30
2d2v hypothetical protein slr0953 1.00
2fw6 N5-carboxyaminoimidazole ribonucleotide mutase 2.92
2hdq Beta-lactamase 1.40
2hdr Beta-lactamase 1.72
2i2c Probable inorganic polyphosphate/ATP-NAD kinase 1 4.70
2ihj Alpha-2,3/2,6-sialyltransferase/sialidase 4.59
2ihk Alpha-2,3/2,6-sialyltransferase/sialidase 4.16
2jff UDP-N-ACETYLMURAMOYLALANINE--D-GLUTAMATE LIGASE 3.62
2oiq Proto-oncogene tyrosine-protein kinase Src 4.94
2pjo Ricin (EC 3.2.2.22) 0.05
2r3d Ricin (EC 3.2.2.22) -0.15
2v7u 5'-FLUORO-5'-DEOXY ADENOSINE SYNTHETASE 5.90
2vhw ALANINE DEHYDROGENASE 5.09
3c2f Nicotinate-nucleotide pyrophosphorylase 2.00

29 UNDER complexes from ≥12 of 17 methods determined by regression and σ of residuals
(The 14 italicized complexes below are also in the list of the 15 UNDERs determined by RMSE)
Values at the end of each line are the pK_d for the complex.

1duv ORNITHINE TRANSCARBAMOYLASE 11.80

1kzk Protease (HIV-1 WT) 10.39
1q0y Fab 9B1, Light chain 9.00
1qkt Estrogen receptor alpha 9.04
1sw1 osmoprotection protein (proX) 7.30
1swk CORE-STREPTAVIDIN 12.00
1ui0 Uracil-DNA Glycosylase 7.06
1y0l Catalytic Antibody Fab 34E4 Light chain 9.00
2c1q BIOTIN BINDING PROTEIN A 13.00
2fv5 ADAM 17 10.52
2fvd Cell division protein kinase 2 8.52
2hd6 Carbonic anhydrase 2 7.80
2hzl TRAP-T family sorbitol/mannitol transporter, periplasmic binding protein, SmoM 6.57
2i0a HIV-1 Protease (WT) 11.40
2i0d HIV-1 Protease (WT) 12.10
2jbj GLUTAMATE CARBOXYPEPTIDASE 2 9.70
2jj3 Estrogen receptor beta 9.55
2pog Estrogen receptor alpha 9.54
2qeh D7R4 Protein 8.62
2qi3 HIV-1 Protease (WT) 10.20
2qi4 HIV-1 Protease (WT) 10.44
2qi5 HIV-1 Protease (WT) 10.85
2qi6 HIV-1 Protease (WT) 10.57
2rca Glutamate [NMDA] receptor subunit 3B 7.79
2z4b Estrogen receptor beta 9.36
3b92 ADAM 17 8.00
3brn Lipocalin 8.70
3bu1 Lipocalin 8.15
3czv Carbonic anhydrase 13 7.80

123 GOOD complexes from ≥12 of 17 methods determined by regression and 1.1 logK
(The 25 italicized complexes below are also in the list of the 34 GOODs determined separately for the nine methods that estimated absolute binding affinities) Values at the end of each line are the pK_d for the complex.

10gs GLUTATHIONE S-TRANSFERASE P1-1 6.40
1a8i GLYCOGEN PHOSPHORYLASE B 5.52
1a99 PUTRESCINE-BINDING PROTEIN 5.70
1b6j RETROPEPSIN 7.92
1b6l RETROPEPSIN 8.30
1b6m RETROPEPSIN 8.40
1bky VP39 3.84
1d4i HIV-1 Protease (L63P) 8.85
1d4j HIV-1 Protease (L63P) 8.36
1ec0 HIV-1 Protease (L63P) 8.49
1ec1 HIV-1 Protease (L63P) 8.92

1enu tRNA guanine transglycosylase 5.08
1fh9 BETA-1,4-XYLANASE 6.43
1g2k HIV-1 Protease (L63P) 7.96
1gi9 UROKINASE-TYPE PLASMINOGEN ACTIVATOR 5.96
1gja UROKINASE-TYPE PLASMINOGEN ACTIVATOR 5.42
1gpk ACETYLCHOLINESTERASE 5.37
1h23 ACETYLCHOLINESTERASE 8.35
1ha2 SERUM ALBUMIN 5.54
1hnn PHENYLETHANOLAMINE N-METHYLTRANSFERASE 6.24
1i7z CHIMERA OF IG KAPPA 6.40
1iy7 Carboxypeptidase A 6.19
1nc1 MTA/SAH nucleosidase 6.12
1o5a Urokinase-type plasminogen activator 6.42
1ow4 pheromone binding protein 5.68
1p1n Glutamate receptor 2 6.80
1p1o Glutamate receptor 2 5.76
1pxp Cell division protein kinase 2 6.66
1q4w tRNA guanine transglycosylase 6.46
1r5y tRNA guanine transglycosylase 6.46
1s38 tRNA guanine transglycosylase 5.15
1s50 Glutamate Receptor 6 5.85
1s7y Glutamate Receptor 6 5.85
1s9t Glutamate Receptor 6 6.60
1syh Glutamate receptor 2 6.41
1u1w Phenazine biosynthesis protein phzF 5.85
1ugx Agglutinin alpha chain 5.91
1urg MALTOSE-BINDING PROTEIN 5.82
1usk LEUCINE-SPECIFIC BINDING PROTEIN 6.40
1uwu BETA-GALACTOSIDASE 5.98
1uzv PSEUDOMONAS AERUGINOSA LECTIN II 5.18
1v0l ENDO-1,4-BETA-XYLANASE A 6.32
1v1j 3-DEHYDROQUINATE DEHYDRATASE 4.82
1vot ACETYLCHOLINESTERASE 6.60
1xge Dihydroorotase 4.68
1xl5 HIV-1 Protease (L63P) 7.35
1xw6 Glutathione S-transferase Mu 1 5.62
1y20 Glutamate [NMDA] receptor subunit zeta 1 5.32
1yc1 Heat shock protein HSP 90-alpha 6.17
1ydk Glutathione S-transferase A1 5.89
1yxd dihydrodipicolinate synthase 4.57
1z95 Androgen Receptor 7.12
1zhx KES1 protein 7.26
2afx Glutaminyl-peptide cyclotransferase 5.15
2b07 Tyrosine-protein phosphatase, non-receptor type 1 6.43
2b3f glucose-binding protein 6.03
2bvd ENDOGLUCANASE H 6.00

2c94 6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE 6.82
2cem HIV-1 Protease (L63P) 7.92
2cen HIV-1 Protease (L63P) 8.30
2d3z polyprotein 6.64
2e2p HEXOKINASE 4.74
2fai Estrogen receptor 6.24
2fdp Beta-secretase 1 7.59
2flr Coagulation factor VII 6.10
2fqt S-ribosylhomocysteine lyase 6.14
2fqw Membrane lipoprotein tmpC 6.68
2fqx Membrane lipoprotein tmpC 7.15
2fqy Membrane lipoprotein tmpC 6.57
2gkl Beta-lactamase 5.35
2gz2 Aspartate beta-semialdehyde dehydrogenase 5.22
2h15 Carbonic anhydrase 2 5.67
2h21 Ribulose-1,5 bisphosphate carboxylase/oxygenase 6.54
2h6b ChloroPhenol Reduction gene K 5.39
2h6t Candidapepsin-3 7.22
2hj4 Aromatic amine dehydrogenase; chain D, H 5.37
2ilz poliovirus polymerase 5.40
2j78 BETA-GLUCOSIDASE A 6.42
2jdy FUCOSE-BINDING LECTIN PA-IIL 4.37
2jgb EUKARYOTIC TRANSLATION INITIATION FACTOR 4E TYPE 2 5.61
2nmx Carbonic anhydrase 1 5.57
2nn1 Carbonic anhydrase 1 5.82
2ojg Mitogen-activated protein kinase 1 5.64
2p7g Estrogen-related receptor gamma 6.53
2p98 Methionine aminopeptidase 6.19
2pgz Soluble acetylcholine receptor 5.74
2pqc 3-phosphoshikimate 1-carboxyvinyltransferase 5.74
2psv HIV-1 Protease (WT) 7.24
2pvu ArtJ 6.31
2pwg Sucrose isomerase 4.82
2pzv Steroid Delta-isomerase 3.87
2q3c Cysteine synthase A 5.30
2q6m Cholix toxin 6.29
2q88 Putative ABC transporter amino acid-binding protein 5.80
2q89 Putative ABC transporter amino acid-binding protein 6.30
2qbr Tyrosine-protein phosphatase, non-receptor type 1 6.33
2qbs Tyrosine-protein phosphatase, non-receptor type 1 6.68
2qdt Metallo-beta-lactamase L1 5.19
2qhy HIV-1 Protease (WT) 7.48
2qhz HIV-1 Protease (WT) 7.28
2qi1 HIV-1 Protease (WT) 7.30
2qrl Saccharopine dehydrogenase [NAD+, L-lysine-forming 4.00
2r5p Protease [HIV, subtype C] 8.48

2rde Uncharacterized protein VCA0042 7.01
2reg PUTATIVE GLYCINE BETAINE-BINDING ABC TRANSPORTER PROTEIN 5.57
2uy4 ENDOCHITINASE 4.68
2v7t 5'-FLUORO-5'-DEOXY ADENOSINE SYNTHETASE 7.48
2v7v 5'-FLUORO-5'-DEOXYADENOSINE SYNTHASE 6.66
2v8y EUKARYOTIC TRANSLATION INITIATION FACTOR 4E 5.71
2vkm BETA-SECRETASE 1 8.74
3bex Type III pantothenate kinase 5.57
3bgz Proto-oncogene serine/threonine-protein kinase Pim-1 6.26
3bxe Central glycolytic gene regulator 5.48
3c7i Growth factor receptor-bound protein 2 5.64
3ccz 3-hydroxy-3-methylglutaryl-coenzyme A reductase 7.29
3cd0 3-hydroxy-3-methylglutaryl-coenzyme A reductase 7.57
3cd7 3-hydroxy-3-methylglutaryl-coenzyme A reductase 7.84
3cda 3-hydroxy-3-methylglutaryl-coenzyme A reductase 7.58
3ddg alpha-mannosidase 2 6.00
3ekr Heat shock protein HSP 90-alpha 5.55
3ene Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma isoform 6.24
3f48 Transporter (LeuT) 5.48
3f8c Transcriptional regulator, PadR-like family 7.68

Listing of the OVER, UNDER, and GOOD Complexes Determined from Med |Err| Metrics

36 OVERs from ≥7 of 9 methods estimating absolute affinities and a +Med |Err| cutoff
(1fcx,1w3k, 1x9d, 2c92, 2cem, 2jfz, 2qmj, 2qnq, and 3b3c are not included in the list of 34 OVER complexes determined by regression and σ) Values at the end of each line are the pK_d for the complex.

1f74 N-ACETYL-NEURAMINATE LYASE 3.05
1fcx RETINOIC ACID RECEPTOR GAMMA-1 7.12
1gww N-ACETYLLACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE 4.28
1gx0 N-ACETYLLACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE 4.28
1h46 EXOGLUCANASE I 3.57
1ps3 Alpha-mannosidase II 2.28
1q6e beta-amylase 2.62
1q6g beta-amylase 2.06
1tog Aspartate aminotransferase 3.22
1w3k ENDOGLUCANASE 5A 4.30
1x8d Hypothetical protein yiiL 2.21
1x9d Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase 3.96
1y1m Glutamate [NMDA] receptor subunit zeta 1 1.82
1y1z Glutamate [NMDA] receptor subunit zeta 1 3.08
2aac ARAC 2.22
2add fructan 1-exohydrolase IIa 2.46
2azr Tyrosine-protein phosphatase, non-receptor type 1 3.64

2b1q hypothetical protein slr0953 1.59
2c92 6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE 5.68
2cem HIV-1 Protease (L63P) 7.92
2d2v hypothetical protein slr0953 1.00
2fw6 N5-carboxyaminoimidazole ribonucleotide mutase 2.92
2hdq Beta-lactamase 1.40
2i2c Probable inorganic polyphosphate/ATP-NAD kinase 1 4.70
2ihj Alpha-2,3/2,6-sialyltransferase/sialidase 4.59
2ihk Alpha-2,3/2,6-sialyltransferase/sialidase 4.16
2jff UDP-N-ACETYLMURAMOYLALANINE--D-GLUTAMATE LIGASE 3.62
2jfz GLUTAMATE RACEMASE 5.24
2oiq Proto-oncogene tyrosine-protein kinase Src 4.94
2pjo Ricin (EC 3.2.2.22) 0.05
2qmj Maltase-glucoamylase, intestinal 4.21
2qnq HIV-1 Protease (L63P) 6.11
2r3d Ricin (EC 3.2.2.22) -0.15
2v7u 5'-FLUORO-5'-DEOXY ADENOSINE SYNTHETASE 5.90
2vhw ALANINE DEHYDROGENASE 5.09
3b3c Bacterial leucyl aminopeptidase 3.98

28 UNDERs from ≥7 of 9 methods estimating absolute affinities and a -Med |Err| cutoff
(1txf, 2jdu, 3b50, 3eko, and 4ubp are <u>not</u> included in the list of 29 UNDER complexes determined by regression and σ) Values at the end of each line are the pK_d for the complex.

1duv ORNITHINE TRANSCARBAMOYLASE 11.80
1q0y Fab 9B1, Light chain 9.00
1qkt Estrogen receptor alpha 9.04
1sw1 osmoprotection protein (proX) 7.30
1swk CORE-STREPTAVIDIN 12.00
1txf Glutamate receptor, ionotropic kainate 1 7.24
1ui0 Uracil-DNA Glycosylase 7.06
2c1q BIOTIN BINDING PROTEIN A 13.00
2hzl TRAP-T family sorbitol/mannitol transporter, periplasmic binding protein, SmoM 6.57
2i0a HIV-1 Protease (WT) 11.40
2i0d HIV-1 Protease (WT) 12.10
2jbj GLUTAMATE CARBOXYPEPTIDASE 2 9.70
2jdu FUCOSE-BINDING LECTIN PA-IIL 6.72
2jj3 Estrogen receptor beta 9.55
2pog Estrogen receptor alpha 9.54
2qeh D7R4 Protein 8.62
2qi3 HIV-1 Protease (WT) 10.20
2qi4 HIV-1 Protease (WT) 10.44
2qi5 HIV-1 Protease (WT) 10.85
2qi6 HIV-1 Protease (WT) 10.57
2rca Glutamate [NMDA] receptor subunit 3B 7.79

3b50 Sialic acid-binding periplasmic protein siaP 7.55
3b92 ADAM 17 8.00
3brn Lipocalin 8.70
3bu1 Lipocalin 8.15
3czv Carbonic anhydrase 13 7.80
3eko Heat shock protein HSP 90-alpha 6.70
4ubp PROTEIN (UREASE (CHAIN A)) 5.59

34 GOODs from ≥7 of 9 methods estimating absolute affinities and a 1.1 logK cutoff
(Nine complexes are in italics below; they are <u>not</u> included in the list of 123 GOOD complexes determined by regression and σ) Values at the end of each line are the pK_d for the complex.

1bcj MANNOSE-BINDING PROTEIN-A 3.70
1enu tRNA guanine transglycosylase 5.08
1fh9 BETA-1,4-XYLANASE 6.43
1ha2 SERUM ALBUMIN 5.54
1i7z CHIMERA OF IG KAPPA 6.40
1lhw SEX HORMONE-BINDING GLOBULIN 8.16
1p1n Glutamate receptor 2 6.80
1s50 Glutamate Receptor 6 5.85
1s7y Glutamate Receptor 6 5.85
1s9t Glutamate Receptor 6 6.60
1toi Aspartate aminotransferase 4.05
1uwu BETA-GALACTOSIDASE 5.98
1xw6 Glutathione S-transferase Mu 1 5.62
1y20 Glutamate [NMDA] receptor subunit zeta 1 5.32
1yc1 Heat shock protein HSP 90-alpha 6.17
1yxd dihydrodipicolinate synthase 4.57
2b3f glucose-binding protein 6.03
2d3z polyprotein 6.64
2fai Estrogen receptor 6.24
2fqt S-ribosylhomocysteine lyase 6.14
2fqw Membrane lipoprotein tmpC 6.68
2ou0 Lysozyme 3.23
2p4y Peroxisome proliferator-activated receptor gamma 9.00
2p7g Estrogen-related receptor gamma 6.53
2p98 Methionine aminopeptidase 6.19
2pwd Sucrose isomerase 4.40
2pzv Steroid Delta-isomerase 3.87
2qbq Tyrosine-protein phosphatase, non-receptor type 1 7.44
2uy4 ENDOCHITINASE 4.68
2vmd DISCOIDIN-2 3.02
2zmm Tyrosine-protein phosphatase, non-receptor type 1 7.60
3ccz 3-hydroxy-3-methylglutaryl-coenzyme A reductase 7.29
3cd0 3-hydroxy-3-methylglutaryl-coenzyme A reductase 7.57

3ddg alpha-mannosidase 2 6.00