Supporting Information for "Simple Entropy Terms for End-Point Binding Free Energy Calculations"

William M. Menzer,[†] Chen Li,[‡] Wenji Sun,[†] Bing Xie,[‡] and David D. L. Minh[*,‡]

Department of Biological Sciences, Illinois Institute of Technology, Chicago, IL 60616, USA, and Department of Chemistry, Illinois Institute of Technology, Chicago, IL 60616, USA

E-mail: dminh@iit.edu

[*]To whom correspondence should be addressed
[†]Illinois Tech Biological Sciences
[‡]Illinois Tech Chemistry

Table S1: **Details of each of the 6 datasets**

Dataset Name	# Ligands	Structural Data Source	Binding Data Source
bromodomains	8	PDB identifiers 3MXF, 3SVG, 3U5J, 3U5L, 4J0R, 4MR3, 4MR4, and 4OGJ	Isothermal titration calorimetry reported by Aldeghi et al.[1]
FXR	31	Roche crystal structures donated to D3R (`http://dx.doi.org/10.15782/D6RP4P`) Identifiers: 1btoj, 1dqzc, 1fggu, 1fymf, 1hoia, 1hqmf, 1hvih, 1ifba, 1ivlk, 1ixdk, 1jski, 1kjyp, 1kmrz, 1lkvm, 1lrpw, 1nqqw, 1pdbc, 1rfrr, 1rtaz, 1rwui, 1rzch, 1skpl, 1uxjy, 1wgph, 1wles, 1xivt, 1xxar, 1yfjn, 1ymrc, 1ytut, and 1zhbd	IC_{50} measured with a competitive binding assay, the scintillation proximity assay
hPNMT	7	PDB identifiers 1HNN, 1N7J, 2G70, 2G71, 2G72, 2G8N, and 2ONY	k_i reported by Gee et al.[2]
lysozyme	9	PDB identifiers 181L, 182L, 183L, 184L, 185L, 186L, 187L, 188L, and 1NHB	Isothermal titration calorimetry reported by Morton and Matthews[3]
MAP4K4	18	Subset of Genentech crystal structures donated to D3R (`http://dx.doi.org/10.15782/D6WC7Z`) Identifiers: MAP01, MAP02, MAP03, MAP04, MAP05, MAP06, MAP07, MAP08, MAP09, MAP11, MAP12, MAP13, MAP14, MAP15, MAP16, MAP17, MAP18, and MAP19	k_i measured with Z-LYTE biochemical assay
Interaction Entropy	14	PDB identifiers 2BRB, 2IWX, 2VW5, 2WBG, 2X00, 2XDL, 2YGE, 2ZJW, 3AO4, 3K5V, 3KGP, 3OWJ, 4DES, and 4DEW.	Compiled by Duan et al.[4]

Table S2: **Comparison between PROPKA and AmberTools default protonation states, showing distances to the closest ligand atom in the crystal structure.** No differences were observed in the bromodomain dataset.

Dataset	ID	Residue	PROPKA	AmberTools	Distance (Å)
FXR	1nqqw	205	HID	HIE	2.300
		52	HID	HIE	3.739
		204	HID	HIE	4.290
		203	HID	HIE	7.158
		102	HID	HIE	9.483
		71	HIP	HIE	17.371
		12	HID	HIE	24.410
		180	HID	HIE	27.777
	1fggu	52	HID	HIE	2.315
		204	HID	HIE	7.859
		203	HIP	HIE	10.443
		12	HID	HIE	17.336
		180	HID	HIE	20.035
		71	HID	HIE	22.760
	1btoj	52	HID	HIE	2.620
		102	HID	HIE	7.423
		204	HID	HIE	9.034
		203	HID	HIE	9.973
		71	HID	HIE	21.140
		12	HID	HIE	22.236
		180	HID	HIE	24.164
	1ytut	205	HID	HIE	2.886
		52	HID	HIE	3.676
		204	HID	HIE	7.229
		102	HID	HIE	9.779
		203	HID	HIE	10.503
		231	HID	HIE	18.377
		12	HID	HIE	20.159
		180	HID	HIE	22.836
		71	HID	HIE	23.213
	1ivlk	205	HID	HIE	3.045
		52	HID	HIE	4.034
		210	HID	HIE	6.723
		204	HID	HIE	8.206
		102	HID	HIE	8.630
		203	HIP	HIE	10.394
		224	HID	HIE	16.506
		12	HID	HIE	21.100
		71	HIP	HIE	22.707
		180	HID	HIE	24.537

FXR	1hvih	205	HID	HIE	3.098
		52	HID	HIE	3.548
		102	HID	HIE	5.856
		213	HID	HIE	6.969
		204	HID	HIE	8.249
		203	HIP	HIE	10.433
		226	HID	HIE	17.274
		12	HID	HIE	21.343
		71	HIP	HIE	22.822
		180	HID	HIE	25.494
	1ixdk	52	HID	HIE	3.212
		204	HIP	HIE	7.606
		102	HID	HIE	9.095
		203	HID	HIE	11.280
		12	HID	HIE	16.604
		232	HID	HIE	18.924
		180	HID	HIE	19.983
		71	HID	HIE	22.592
	1ymrc	200	HID	HIE	3.244
		47	HID	HIE	3.564
		97	HID	HIE	6.331
		199	HID	HIE	7.374
		212	HID	HIE	9.970
		198	HID	HIE	10.198
		12	HID	HIE	21.406
		175	HID	HIE	23.016
		66	HID	HIE	23.338
	1hoia	52	HID	HIE	3.296
		205	HID	HIE	5.349
		204	HIP	HIE	7.192
		102	HID	HIE	8.983
		217	HID	HIE	10.556
		203	HID	HIE	10.756
		12	HID	HIE	17.907
		232	HID	HIE	18.429
		180	HID	HIE	20.598
		71	HID	HIE	23.403
	1fymf	52	HID	HIE	3.330
		204	HID	HIE	7.326
		102	HID	HIE	9.123
		217	HID	HIE	10.334
		203	HID	HIE	11.106
		12	HID	HIE	17.812
		231	HID	HIE	18.946

FXR		180	HID	HIE	20.231
		71	HID	HIE	22.949
	1kjyp	205	HID	HIE	3.331
		52	HID	HIE	3.730
		204	HID	HIE	7.070
		102	HID	HIE	7.668
		203	HID	HIE	9.809
		230	HID	HIE	16.257
		12	HID	HIE	18.001
		71	HIP	HIE	22.447
		180	HID	HIE	24.897
	1yfjn	52	HID	HIE	3.394
		204	HIP	HIE	7.231
		102	HID	HIE	9.301
		203	HID	HIE	11.114
		12	HID	HIE	17.502
		231	HID	HIE	18.377
		180	HID	HIE	21.367
		71	HID	HIE	23.106
	1rfrr	54	HID	HIE	3.426
		207	HID	HIE	3.808
		104	HID	HIE	4.336
		206	HID	HIE	6.999
		205	HID	HIE	9.765
		219	HID	HIE	10.219
		14	HID	HIE	20.403
		73	HID	HIE	22.919
		182	HID	HIE	23.981
		2	HID	HIE	34.745
	1hqmf	52	HID	HIE	3.429
		205	HID	HIE	3.590
		102	HID	HIE	7.848
		204	HID	HIE	7.908
		203	HID	HIE	10.585
		12	HID	HIE	15.918
		180	HID	HIE	20.444
		71	HID	HIE	22.353
	1wgph	52	HID	HIE	3.461
		205	HID	HIE	3.941
		204	HID	HIE	6.269
		102	HID	HIE	8.851
		203	HID	HIE	9.539
		217	HID	HIE	9.641
		232	HID	HIE	15.981
		12	HID	HIE	17.881

FXR		71	HID	HIE	23.038
		180	HID	HIE	23.986
	1xxar	205	HID	HIE	3.461
		52	HID	HIE	3.932
		102	HID	HIE	7.271
		204	HID	HIE	8.604
		203	HID	HIE	11.186
		226	HID	HIE	17.369
		12	HID	HIE	18.395
		180	HID	HIE	24.996
	1rtaz	52	HID	HIE	3.499
		204	HIP	HIE	7.151
		102	HID	HIE	8.993
		217	HID	HIE	10.465
		203	HID	HIE	11.001
		232	HID	HIE	18.082
		12	HID	HIE	18.226
		180	HID	HIE	22.379
		71	HID	HIE	23.191
	1sjpr	206	HID	HIE	3.570
		53	HID	HIE	3.994
		103	HID	HIE	7.377
		205	HID	HIE	8.606
		213	HID	HIE	9.809
		204	HIP	HIE	11.115
		13	HID	HIE	18.270
		72	HID	HIE	22.662
		181	HID	HIE	24.980
	1dqzc	52	HID	HIE	3.582
		205	HID	HIE	4.337
		204	HIP	HIE	6.426
		203	HID	HIE	9.766
		102	HID	HIE	9.808
		232	HID	HIE	16.543
		12	HID	HIE	19.926
		180	HID	HIE	20.799
		71	HID	HIE	23.265
	1rwui	52	HID	HIE	3.623
		205	HID	HIE	5.970
		102	HID	HIE	9.030
		204	HIP	HIE	10.414
		217	HID	HIE	10.691
		203	HID	HIE	12.912
		232	HID	HIE	18.875
		180	HID	HIE	19.707

FXR		12	HID	HIE	20.525
		71	HID	HIE	23.156
	1zhbd	52	HID	HIE	3.623
		204	HID	HIE	4.368
		101	HID	HIE	7.566
		203	HID	HIE	9.308
		202	HIP	HIE	11.759
		229	HID	HIE	16.833
		12	HID	HIE	21.667
		71	HIP	HIE	23.103
		179	HID	HIE	25.614
	1skpl	52	HID	HIE	3.632
		205	HID	HIE	6.288
		102	HID	HIE	8.966
		204	HIP	HIE	10.220
		217	HID	HIE	10.778
		203	HID	HIE	13.114
		232	HID	HIE	19.113
		12	HID	HIE	19.132
		180	HID	HIE	21.688
		71	HID	HIE	23.208
	1rzch	52	HID	HIE	3.661
		205	HID	HIE	3.754
		204	HIP	HIE	6.068
		102	HID	HIE	9.550
		203	HID	HIE	9.767
		232	HID	HIE	17.646
		180	HID	HIE	20.238
		12	HID	HIE	20.506
		71	HID	HIE	23.244
	1jski	52	HID	HIE	3.682
		205	HID	HIE	4.349
		204	HID	HIE	7.697
		102	HID	HIE	9.095
		203	HID	HIE	11.281
		12	HID	HIE	20.297
		180	HID	HIE	22.262
		71	HID	HIE	23.324
	1xivt	52	HID	HIE	3.683
		205	HID	HIE	3.971
		204	HIP	HIE	7.838
		102	HID	HIE	8.924
		203	HID	HIE	11.249
		233	HID	HIE	18.791
		180	HID	HIE	19.766

FXR		12	HID	HIE	20.470
		71	HID	HIE	23.189
	1lkvm	52	HID	HIE	3.696
		205	HID	HIE	4.318
		204	HID	HIE	6.344
		102	HID	HIE	8.978
		203	HID	HIE	9.948
		232	HID	HIE	17.612
		12	HID	HIE	20.227
		180	HID	HIE	21.263
		71	HID	HIE	23.258
	1txca	52	HID	HIE	3.697
		205	HID	HIE	5.413
		204	HIP	HIE	7.425
		217	HID	HIE	9.907
		102	HID	HIE	9.956
		203	HID	HIE	11.037
		232	HID	HIE	18.715
		12	HID	HIE	20.626
		180	HID	HIE	20.728
		71	HID	HIE	23.396
	1ifba	205	HID	HIE	3.721
		52	HID	HIE	3.752
		204	HIP	HIE	7.393
		102	HID	HIE	9.099
		217	HID	HIE	10.300
		203	HID	HIE	10.899
		232	HID	HIE	18.397
		12	HID	HIE	20.090
		180	HID	HIE	21.280
		71	HID	HIE	23.391
	1wles	51	HID	HIE	3.809
		204	HID	HIE	5.259
		203	HIP	HIE	7.283
		101	HID	HIE	9.296
		202	HID	HIE	10.830
		11	HID	HIE	19.033
		179	HID	HIE	20.069
		70	HID	HIE	23.511
	1uxjy	52	HID	HIE	3.831
		204	HIP	HIE	7.449
		102	HID	HIE	9.872
		203	HID	HIE	10.845
		233	HID	HIE	18.467
		12	HID	HIE	20.719

FXR		180	HID	HIE	21.063
		71	HID	HIE	23.371
	1pdbc	52	HID	HIE	3.888
		204	HID	HIE	3.973
		203	HID	HIE	7.508
		102	HID	HIE	8.360
		211	HID	HIE	9.704
		202	HID	HIE	9.829
		223	HID	HIE	16.684
		71	HID	HIE	21.427
		12	HID	HIE	24.621
		180	HID	HIE	25.355
	1kmrz	200	HID	HIE	3.911
		52	HID	HIE	4.466
		210	HID	HIE	8.099
		199	HID	HIE	8.190
		198	HIP	HIE	10.580
		225	HID	HIE	15.939
		12	HID	HIE	17.398
		71	HIP	HIE	22.738
		175	HID	HIE	22.812
hPNMT	2g8n	197	GLH	GLU	2.361
		163	GLH	GLU	5.150
		107	HID	HIE	7.623
		239	HID	HIE	9.185
		71	HID	HIE	16.386
		178	HID	HIE	18.417
	2ony	196	GLH	GLU	2.644
		162	GLH	GLU	5.672
		106	HID	HIE	7.686
		238	HID	HIE	9.467
		54	ASH	ASP	11.933
		137	HID	HIE	14.696
		177	HID	HIE	19.711
	2g71	199	GLH	GLU	2.734
		37	LYN	LYS	2.781
		165	GLH	GLU	5.064
		109	HID	HIE	8.683
		241	HID	HIE	9.553
		140	HID	HIE	12.195
		57	ASH	ASP	12.202
		180	HID	HIE	17.556
		73	HID	HIE	19.013
	2g72	203	GLH	GLU	2.761
		169	GLH	GLU	5.216

hPNMT		245	HID	HIE	9.465
		113	HID	HIE	9.510
		144	HID	HIE	12.853
		77	HID	HIE	18.284
		184	HID	HIE	18.494
	1hnn	198	GLH	GLU	2.819
		164	GLH	GLU	5.750
		240	HID	HIE	9.208
		56	ASH	ASP	12.039
		108	HID	HIE	12.239
		139	HID	HIE	14.906
		179	HID	HIE	20.033
		72	HID	HIE	21.930
	2g70	200	GLH	GLU	2.856
		38	LYN	LYS	3.371
		166	GLH	GLU	4.126
		242	HID	HIE	9.468
		110	HID	HIE	12.111
		58	ASH	ASP	12.538
		181	HID	HIE	18.437
		74	HID	HIE	22.113
	1n7j	198	GLH	GLU	3.048
		36	LYN	LYS	4.138
		164	GLH	GLU	5.764
		240	HID	HIE	10.265
		56	ASH	ASP	12.315
		139	HID	HIE	14.553
		108	HID	HIE	14.576
		179	HID	HIE	19.569
		72	HID	HIE	22.251
lysozyme	188l	31	HID	HIE	17.698
	181l	31	HID	HIE	18.531
	182l	31	HID	HIE	18.421
	183l	31	HID	HIE	18.135
	184l	31	HID	HIE	18.675
	185l	31	HID	HIE	18.133
	186l	31	HID	HIE	18.620
	187l	31	HID	HIE	18.692
	1nhb	31	HID	HIE	18.755
MAP4K4	MAP01	133	HIP	HIE	8.731
		287	HID	HIE	11.770
		29	HID	HIE	12.664
		63	HID	HIE	13.310
		125	HID	HIE	16.099

MAP4K4		130	HID	HIE	18.643
		129	HID	HIE	19.419
		219	HID	HIE	23.615
	MAP02	64	HID	HIE	10.682
		286	HID	HIE	12.223
		30	HID	HIE	13.185
		127	HID	HIE	13.790
		131	HID	HIE	17.616
		132	HID	HIE	18.116
		218	HID	HIE	22.019
		81	HID	HIE	22.609
	MAP03	121	HID	HIE	8.500
		24	HID	HIE	12.231
		48	HID	HIE	12.987
		113	HID	HIE	15.473
		268	HID	HIE	16.622
		118	HID	HIE	18.840
		117	HID	HIE	18.980
		67	HID	HIE	23.436
		249	HID	HIE	28.032
	MAP04	66	HID	HIE	13.047
		32	HID	HIE	13.795
		279	HID	HIE	15.961
		131	HID	HIE	17.263
		132	HID	HIE	17.331
	MAP05	278	HID	HIE	12.809
		127	HID	HIE	14.994
		131	HID	HIE	16.973
		132	HID	HIE	17.251
		81	HID	HIE	22.322
		210	HID	HIE	22.736
	MAP06	58	GLH	GLU	4.527
		33	HID	HIE	12.524
		67	HID	HIE	13.850
		281	HID	HIE	15.620
		133	HID	HIE	18.153
		134	HID	HIE	18.459
		213	HID	HIE	18.785
		83	HIP	HIE	21.880
	MAP07	136	HID	HIE	8.001
		283	HID	HIE	12.751
		63	HID	HIE	12.949
		133	HID	HIE	18.113
		132	HID	HIE	18.449
		215	HID	HIE	23.306

MAP4K4	MAP08	134	HID	HIE	8.245
		27	HID	HIE	12.964
		276	HID	HIE	12.988
		61	HID	HIE	13.814
		126	HID	HIE	16.223
		131	HID	HIE	18.516
		130	HID	HIE	19.075
		80	HIP	HIE	19.256
	MAP09	135	HIP	HIE	7.971
		28	HID	HIE	12.328
		289	HID	HIE	12.796
		127	HID	HIE	13.596
		131	HID	HIE	17.245
		132	HID	HIE	18.697
		221	HID	HIE	23.736
	MAP11	282	HID	HIE	10.585
		52	GLH	GLU	11.700
		65	HID	HIE	13.436
		130	HID	HIE	15.249
		135	HID	HIE	18.443
		134	HID	HIE	18.449
		84	HID	HIE	22.100
	MAP12	138	HID	HIE	8.449
		65	HID	HIE	10.889
		31	HID	HIE	12.484
		286	HID	HIE	12.993
		134	HID	HIE	17.004
		135	HID	HIE	17.277
		84	HID	HIE	21.194
	MAP13	138	HID	HIE	8.549
		65	HID	HIE	11.087
		31	HID	HIE	12.468
		284	HID	HIE	13.182
		130	HID	HIE	14.238
		134	HID	HIE	16.948
		135	HID	HIE	17.238
		84	HID	HIE	21.220
	MAP14	131	HID	HIE	8.243
		62	HID	HIE	12.597
		28	HID	HIE	12.654
		123	HID	HIE	15.510
		285	HID	HIE	16.705
		128	HID	HIE	18.259
		127	HID	HIE	18.660

MAP4K4	MAP15	127	HID	HIE	8.596
		29	HID	HIE	12.271
		58	HID	HIE	12.997
		119	HID	HIE	15.207
		270	HID	HIE	16.807
		124	HID	HIE	17.178
		123	HID	HIE	17.318
		202	HID	HIE	22.755
	MAP16	125	HID	HIE	7.370
		52	HID	HIE	12.263
		29	HID	HIE	12.623
		177	ASH	ASP	13.754
		117	HID	HIE	14.124
		119	HID	HIE	15.719
		267	HID	HIE	16.510
		121	HID	HIE	16.793
		122	HID	HIE	16.896
		71	HID	HIE	21.643
		199	HID	HIE	23.730
	MAP17	28	HID	HIE	13.175
		63	HID	HIE	13.844
		62	HID	HIE	14.691
		127	HID	HIE	17.070
		131	HID	HIE	20.561
		132	HID	HIE	20.894
		81	HID	HIE	23.533
	MAP18	278	HID	HIE	10.908
		62	HID	HIE	12.897
		127	HID	HIE	15.879
		129	HID	HIE	16.899
		131	HID	HIE	17.621
		132	HID	HIE	17.796
		81	HID	HIE	21.379
		210	HID	HIE	22.405
	MAP19	130	HID	HIE	8.213
		58	HID	HIE	11.251
		29	HID	HIE	12.532
		57	HID	HIE	12.658
		122	HID	HIE	14.698
		272	HID	HIE	15.811
		126	HID	HIE	16.641
		127	HID	HIE	16.767
		76	HID	HIE	20.990
		204	HID	HIE	22.736

IE	2wbg	164	GLH	GLU	2.394
		403	GLH	GLU	2.628
		296	HID	HIE	3.549
		178	HID	HIE	6.418
		376	HID	HIE	10.739
		173	HID	HIE	13.657
		34	HID	HIE	13.956
		55	HID	HIE	13.982
		199	HID	HIE	15.290
		38	HIP	HIE	17.019
		365	HIP	HIE	19.602
	3ao4	255	HID	HIE	2.945
		253	HID	HIE	6.659
		117	HID	HIE	16.894
		58	HID	HIE	17.161
		197	HID	HIE	20.278
		150	HIP	HIE	22.237
		103	HID	HIE	26.669
		11	HIP	HIE	26.821
	3owj	158	HID	HIE	3.441
		113	HID	HIE	3.970
		166	HIP	HIE	12.907
		27	HID	HIE	18.148
		232	HID	HIE	19.538
		234	HID	HIE	21.185
		181	HID	HIE	24.029
		307	HID	HIE	26.193
		284	HID	HIE	28.653
		16	HID	HIE	29.528
	2x00	198	CYX	CYS	3.590
		199	CYX	CYS	4.573
		195	HID	HIE	7.236
		135	CYX	CYS	10.876
		148	CYX	CYS	11.567
		119	HID	HIE	30.552
		9	HID	HIE	43.907

IE	3kgp	188	CYX	CYS	3.692
		216	CYX	CYS	4.249
		42	HID	HIE	5.089
		88	HID	HIE	7.202
		27	CYX	CYS	7.310
		43	CYX	CYS	7.784
		177	CYX	CYS	9.045
		161	CYX	CYS	10.527
		89	HID	HIE	12.384
		129	CYX	CYS	12.928
		165	HID	HIE	13.947
		198	CYX	CYS	14.688
		158	HID	HIE	15.891
		80	HID	HIE	17.107
		22	HID	HIE	17.295
		230	HID	HIE	18.056
		238	HID	HIE	22.922
		35	CYX	CYS	23.769
		104	CYX	CYS	27.676
	2zjw	160	HID	HIE	5.030
		115	HID	HIE	5.912
		168	HIP	HIE	12.976
		29	HID	HIE	16.592
		234	HID	HIE	20.058
		183	HID	HIE	20.781
		236	HID	HIE	22.029
		309	HID	HIE	25.190
		18	HID	HIE	27.659
		286	HID	HIE	28.672
		276	HID	HIE	35.096
	2xdl	61	HID	HIE	9.487
		173	HID	HIE	12.887
	4des	47	HIP	HIE	9.537
		79	HID	HIE	13.516
		81	HIP	HIE	15.093
		22	HID	HIE	17.001
	3k5v	266	HIP	HIE	10.095
		137	HIP	HIE	17.904
		172	HID	HIE	27.423
		23	HID	HIE	31.680
	2vw5	197	HID	HIE	14.401
	2iwx	197	HID	HIE	14.420
	2brb	181	HIP	HIE	23.721
		239	HID	HIE	29.679

Table S3: **experimental and estimated free energies for all receptor ligand systems using a second order expansion with a 3Å RMSD cutoff and without the inclusion of ligand external or site entropy. All values are reported in kcal/mol.**

Set	Ligand	Experimental	1	2	3	4	EXP	1 + NM
bromodomains	3MXF	-9.6	-39.64	-29.50	-25.43	-15.81	-28.20	-17.20
	4OGJ	-8.9	-71.59	-27.25	166.98	1107.51	-44.34	-52.79
	3SVG	-7.3	-26.02	-17.52	-17.71	-19.20	-19.61	-4.63
	3U5J	-7.4	-24.24	-6.12	-6.78	-45.91	-17.28	-6.37
	4MR3	-9.0	-37.87	-19.62	-43.26	-24.09	-27.64	-21.79
	4MR4	-7.8	-33.15	-15.38	-36.08	30.36	-22.22	-14.06
	3U5L	-8.2	-35.63	-16.92	-17.55	-60.40	-21.82	-15.48
	4J0R	-8.8	-29.27	-17.61	-18.89	-9.75	-20.54	-8.30
FXR	1ymrc	-11.56	-57.16	-47.12	-45.25	-44.54	-47.30	-34.40
	1kjyp	-6.54	-89.76	-41.50	93.03	1437.06	-52.00	-56.71
	1yfjn	-6.67	-79.29	-46.45	-20.34	21.31	-53.81	-48.38
	1rtaz	-5.93	-75.40	-57.41	-32.30	68.74	-53.43	-45.51
	1rzch	-5.67	-66.96	-55.50	-52.24	-50.79	-55.41	-41.68
	1wles	-5.39	-72.70	-54.29	-45.74	-25.36	-30.48	-41.75
	1skpl	-6.63	-64.65	-49.63	-47.10	-46.00	-52.10	-36.15
	1nqqw	-9.33	-66.64	-56.55	-50.46	-45.06	-55.43	-43.56
	1wgph	-6.43	-79.16	-58.33	-41.20	7.88	-56.27	-48.57
	1uxjy	-5.47	-68.02	-51.95	-46.59	-46.98	-55.75	-42.24
	1zhbd	-7.90	-56.97	-45.56	-43.62	-48.18	-47.55	-33.22
	1dqzc	-7.38	-70.49	-51.46	-66.19	-40.90	-57.04	-43.72
	1rwui	-5.42	-65.80	-52.79	-48.96	-49.13	-54.69	-39.36
	1rfrr	-6.81	-66.61	-55.14	-52.22	-54.75	-55.50	-42.13
	1ivlk	-10.24	-69.57	-43.87	-72.04	-22.73	-51.73	-42.64
	1jski	-5.92	-64.20	-55.91	-54.23	-53.73	-56.31	-40.65
	1kmrz	-6.39	-74.76	-30.18	4.28	-320.89	-54.35	-41.44
	1ifba	-6.58	-68.23	-57.89	-54.38	-48.40	-41.96	-41.51
	1hvih	-10.20	-58.25	-47.25	-48.59	-41.40	-48.13	-31.68
	1xivt	-8.07	-65.78	-56.41	-53.40	-52.08	-53.91	-39.39
	1ixdk	-7.20	-112.75	-82.46	-65.28	-8.41	-83.11	-79.00
	1fggu	-5.35	-61.11	22.40	-87.03	-1791.79	-43.98	-36.68
	1btoj	-10.64	-75.80	-29.52	11.67	-456.49	-54.34	-45.06
	1hqmf	-8.09	-79.91	-33.48	174.47	741.65	-54.00	-47.91
	1lrpw	-8.44	-51.80	-42.93	-40.65	-41.22	-42.01	-28.43
	1pdbc	-9.41	-55.56	-49.11	-47.84	-46.93	-48.08	-34.22
	1hoia	-6.30	-70.86	-43.68	-14.63	21.84	-52.48	-44.39
	1ytut	-9.33	-66.55	-55.00	-53.25	-51.68	-47.53	-38.35
	1lkvm	-8.61	-65.82	-54.61	-51.24	-50.88	-54.53	-37.05
	1xxar	-6.97	-82.49	-63.26	-45.18	-2.69	-61.91	-51.39
	1fymf	-5.41	-83.38	-60.74	-31.74	89.40	-57.11	-56.68

Set	Ligand	Experimental	1	2	3	4	EXP	1 + NM
hPNMT	2G70	-10.6	-47.18	-37.94	-33.18	-22.18	-32.26	-25.13
	2G71	-9.1	-66.27	-51.54	-43.96	-33.48	-52.60	-42.26
	2G72	-10.2	-74.88	-46.59	-54.84	-39.81	-57.33	-51.65
	1HNN	-9.4	-50.38	-38.19	-34.13	-34.90	-36.40	-26.91
	2ONY	-10.6	-72.00	-59.51	-59.42	-41.47	-60.54	-46.48
	1N7J	-10.1	-44.87	-38.08	-37.66	-36.68	-37.86	-27.35
	2G8N	-10.1	-68.76	-51.27	-48.35	-49.18	-55.39	-45.09
lysozyme	181L	-5.19	-15.12	-13.49	-12.96	-12.70	-12.58	-2.81
	182L	-5.42	-19.56	-16.90	-16.04	-15.93	-16.29	-7.04
	183L	-5.09	-20.19	-17.26	-16.96	-16.88	-16.56	-6.86
	184L	-6.48	-25.66	-22.40	-21.67	-21.20	-19.99	-10.58
	185L	-4.89	-19.44	-16.78	-16.34	-16.33	-16.55	-7.84
	186L	-6.70	-27.00	-24.38	-23.52	-23.02	-22.84	-13.17
	187L	-4.67	-20.06	-17.69	-17.29	-16.97	-16.06	-5.65
	188L	-4.60	-21.33	-17.88	-18.36	-17.95	-170.00	-9.20
	1NHB	-5.76	-20.90	-18.73	-18.12	-17.82	-16.72	-7.11
MAP4k4	MAP01	-11.56	-85.40	-51.56	109.42	2406.04	-44.90	-71.30
	MAP02	-11.30	-54.88	-30.80	-45.34	49.49	-42.12	-30.80
	MAP03	-7.75	-29.25	-23.93	-22.92	-22.29	-22.20	-14.59
	MAP04	-8.24	-29.21	-15.95	-8.55	-35.53	-20.50	-13.95
	MAP05	-10.97	-38.92	-30.06	-26.99	-26.94	-30.80	-19.24
	MAP06	-9.77	-38.48	-28.40	-28.57	-25.37	-27.81	-18.14
	MAP07	-11.29	-43.21	-32.48	-28.96	-32.50	-34.41	-23.61
	MAP08	-9.60	-46.46	-36.00	-33.60	-32.78	-36.23	-29.82
	MAP09	-10.53	-51.72	-41.70	-37.42	-37.20	-40.08	-32.65
	MAP11	-10.50	-42.02	-22.10	2.07	-6.16	-27.29	-31.97
	MAP12	-11.5	-57.59	-43.16	-46.35	-52.96	-47.33	-31.55
	MAP13	-7.91	-49.09	-31.36	-44.32	-42.05	-37.00	-26.20
	MAP14	-9.68	-66.58	-22.86	-20.94	-396.60	-48.73	-39.62
	MAP15	-10.87	-46.38	-36.87	-33.91	-32.83	-36.40	-22.77
	MAP16	-9.86	-53.82	-43.66	-39.75	-35.79	-43.36	-29.84
	MAP17	-6.86	-38.16	-27.48	-25.09	-21.28	-27.09	-18.63
	MAP18	-9.41	-48.98	-38.49	-36.11	-32.42	-39.40	-38.70
	MAP19	-11.68	-43.79	-33.61	-32.18	-32.35	-33.23	-25.17
Interaction Entropy	2WBG	-6.15	-57.41	-30.23	14.83	225.16	-37.36	-25.73
	2IWX	-9.23	-39.43	-30.83	-28.25	-23.47	-29.44	-19.57
	3KGP	-3.55	-50.80	-13.20	3.06	-67.41	-32.29	-30.12
	4DEW	-9.67	-14.94	-6.94	0.27	15.36	-7.87	6.09
	3K5V	-8.71	-37.54	-22.84	-21.32	-27.42	-27.02	-19.34
	2BRB	-6.72	-38.97	-25.60	-23.98	-29.61	-26.89	-21.58
	2XDL	-4.29	-34.05	-26.73	-26.01	-23.07	-26.25	-13.30
	4DES	-8.09	-14.51	-8.97	-10.36	-19.61	-12.12	-0.38
	3OWJ	-8.38	-34.36	-22.49	-10.15	-28.74	-28.01	-26.96
	2VW5	-11.77	-52.44	-36.93	-33.67	-31.66	-39.06	-27.66
	2ZJW	-10.64	-23.56	-17.73	-14.60	-14.34	-18.11	-5.43
	2X00	-15.65	-32.39	-18.37	-16.55	-27.75	-24.10	-16.39
	3AO4	-2.86	-24.44	-14.87	-5.94	-23.77	-18.94	-14.42
	2YGE	-7.00	-50.98	-34.31	-31.15	-26.68	-36.41	-22.57

Table S4: **Correlation metrics for calculations from Duan et al.[4].** In this paper, MD simulations were run either with restraints for 2 ns or unconstrained for 6 ns. Free energies were calculated by normal modes (NM) or interaction entropy (IE), the latter of which is numerically equivalent to the exponential average.

		R	ρ	τ
2 ns	NM	-0.11 (0.25)	-0.08 (0.28)	-0.03 (0.22)
	IE	0.22 (0.25)	0.35 (0.25)	0.2 (0.19)
6 ns	NM	0.31 (0.32)	0.13 (0.28)	0.1 (0.21)
	IE	0.08 (0.29)	-0.07 (0.31)	-0.1 (0.24)

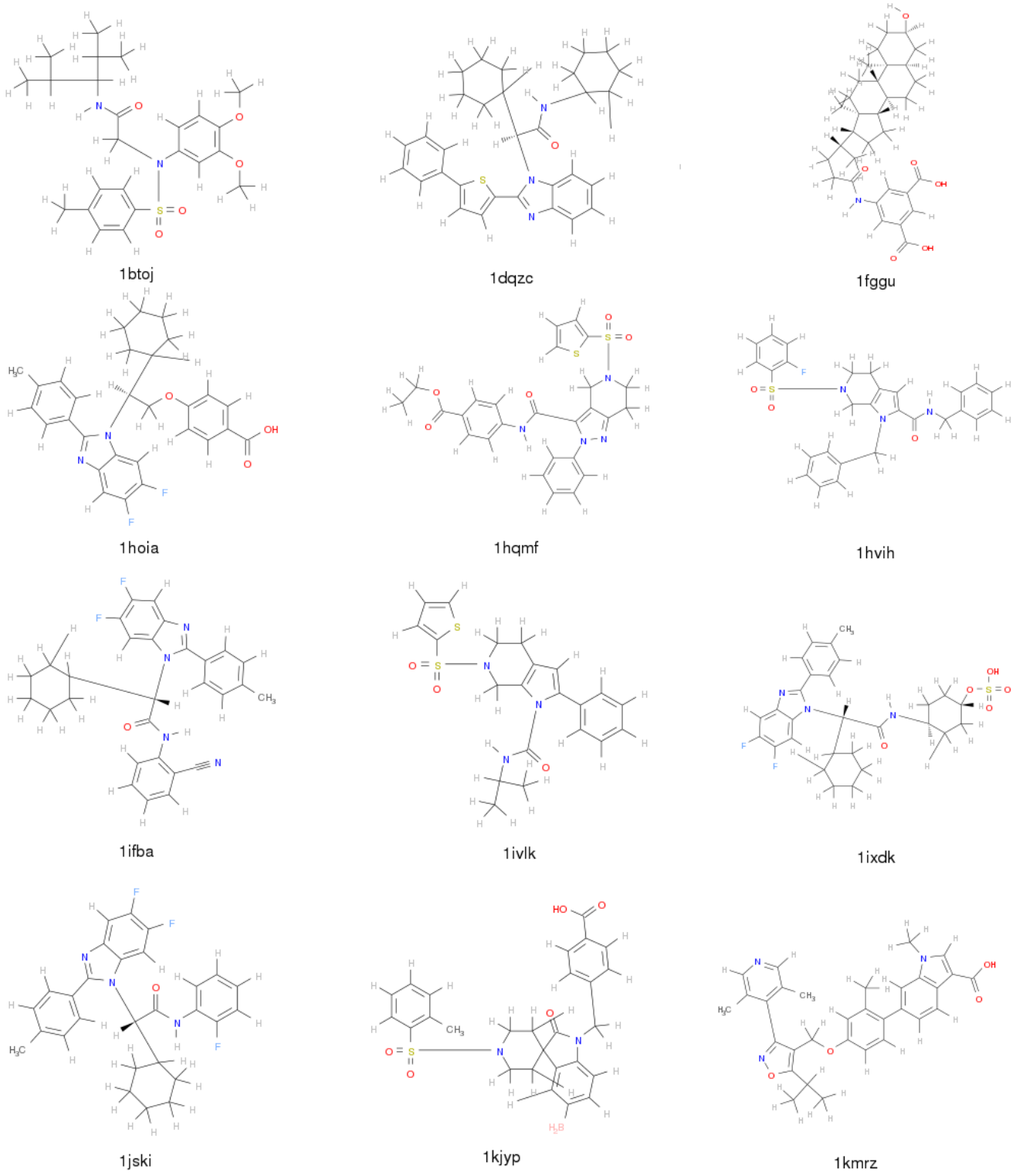

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

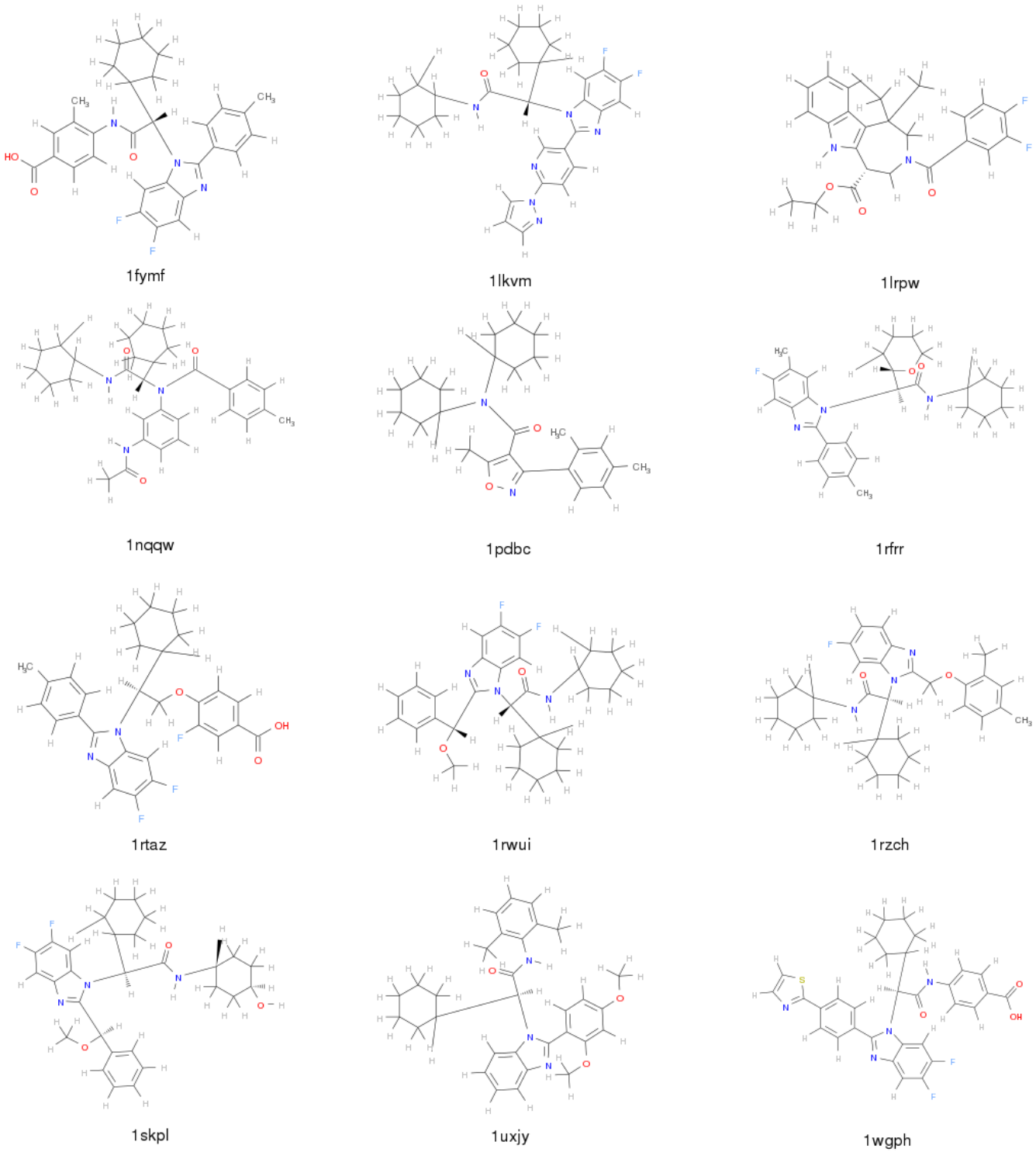

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

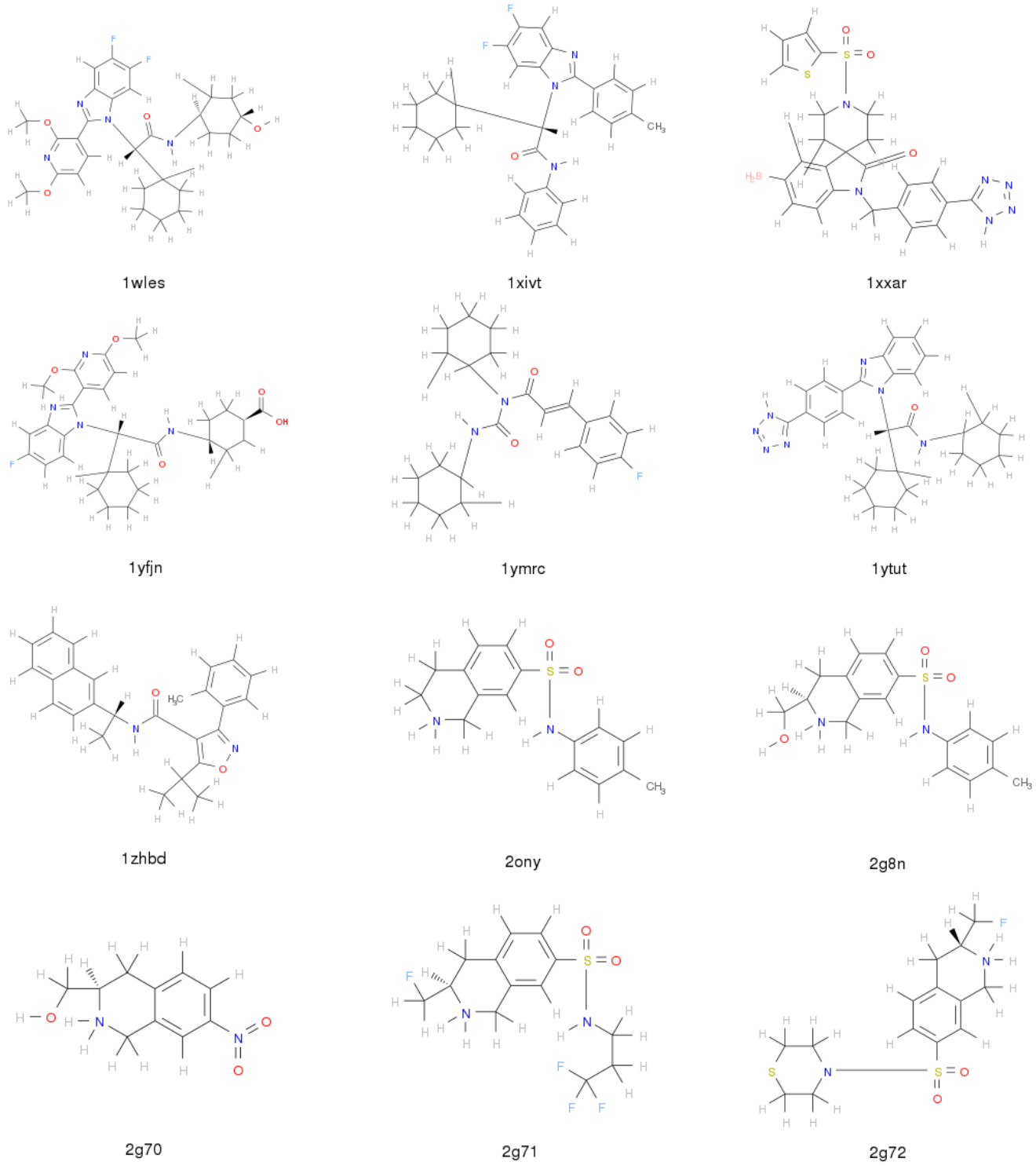

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

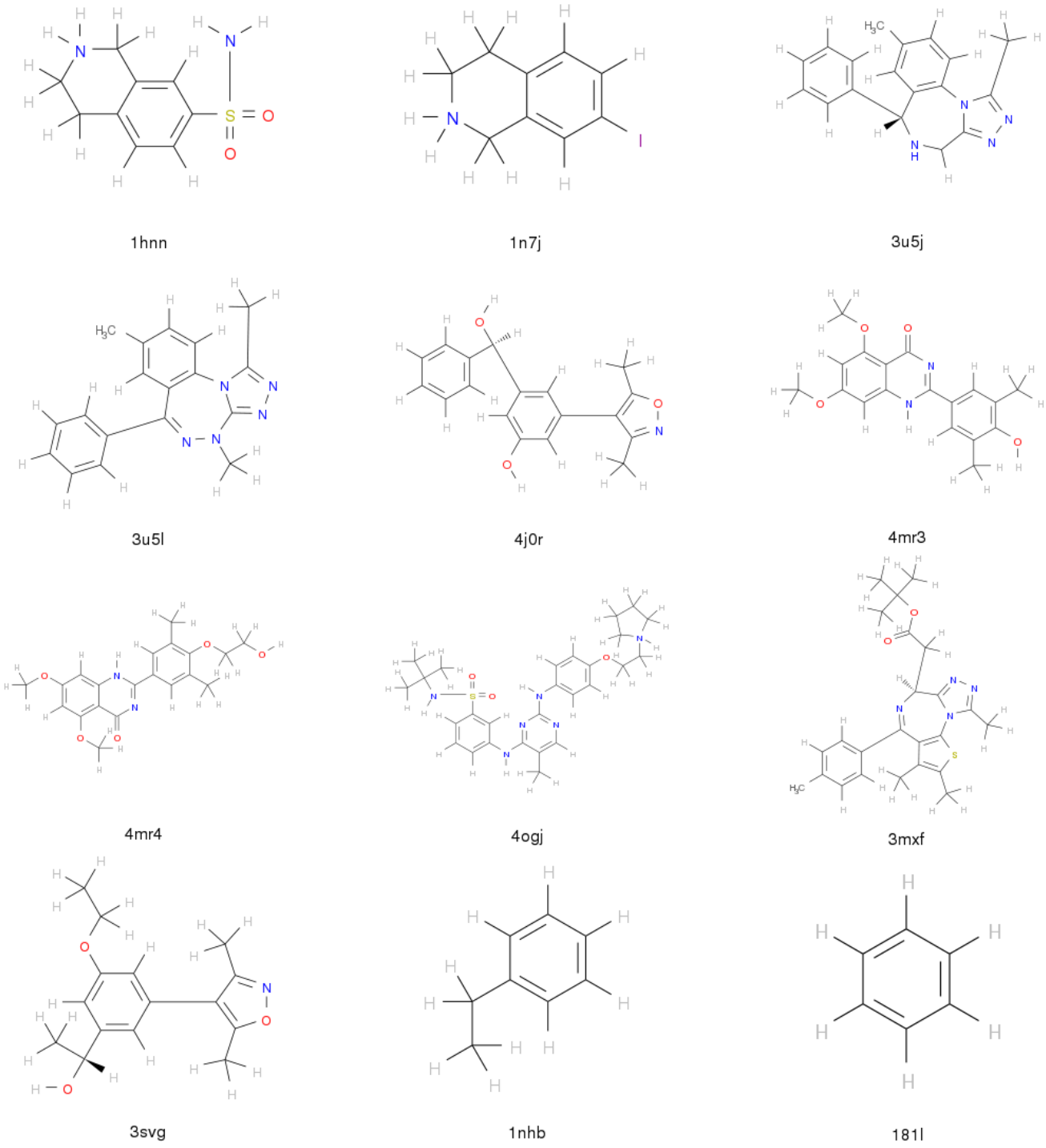

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

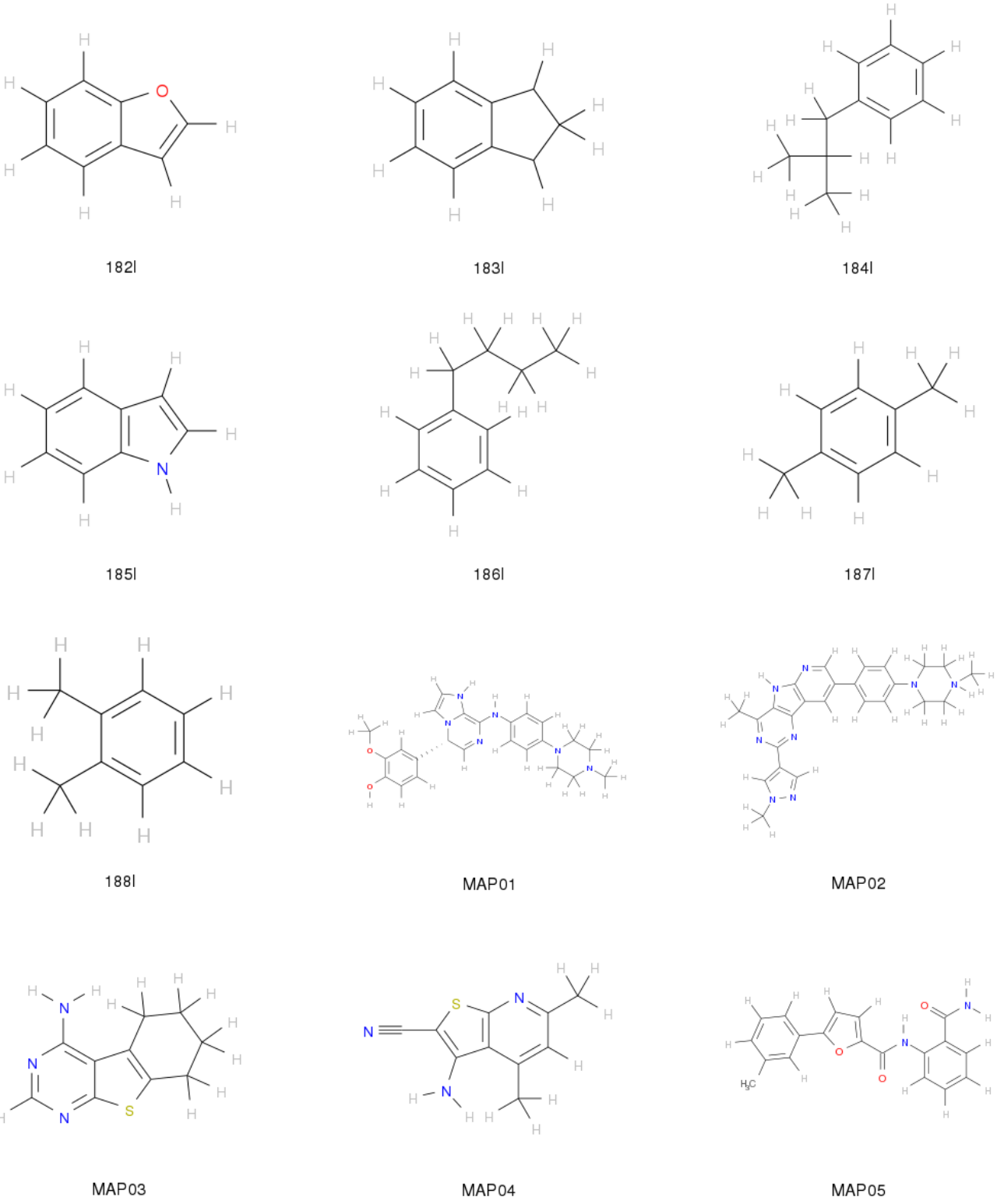

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

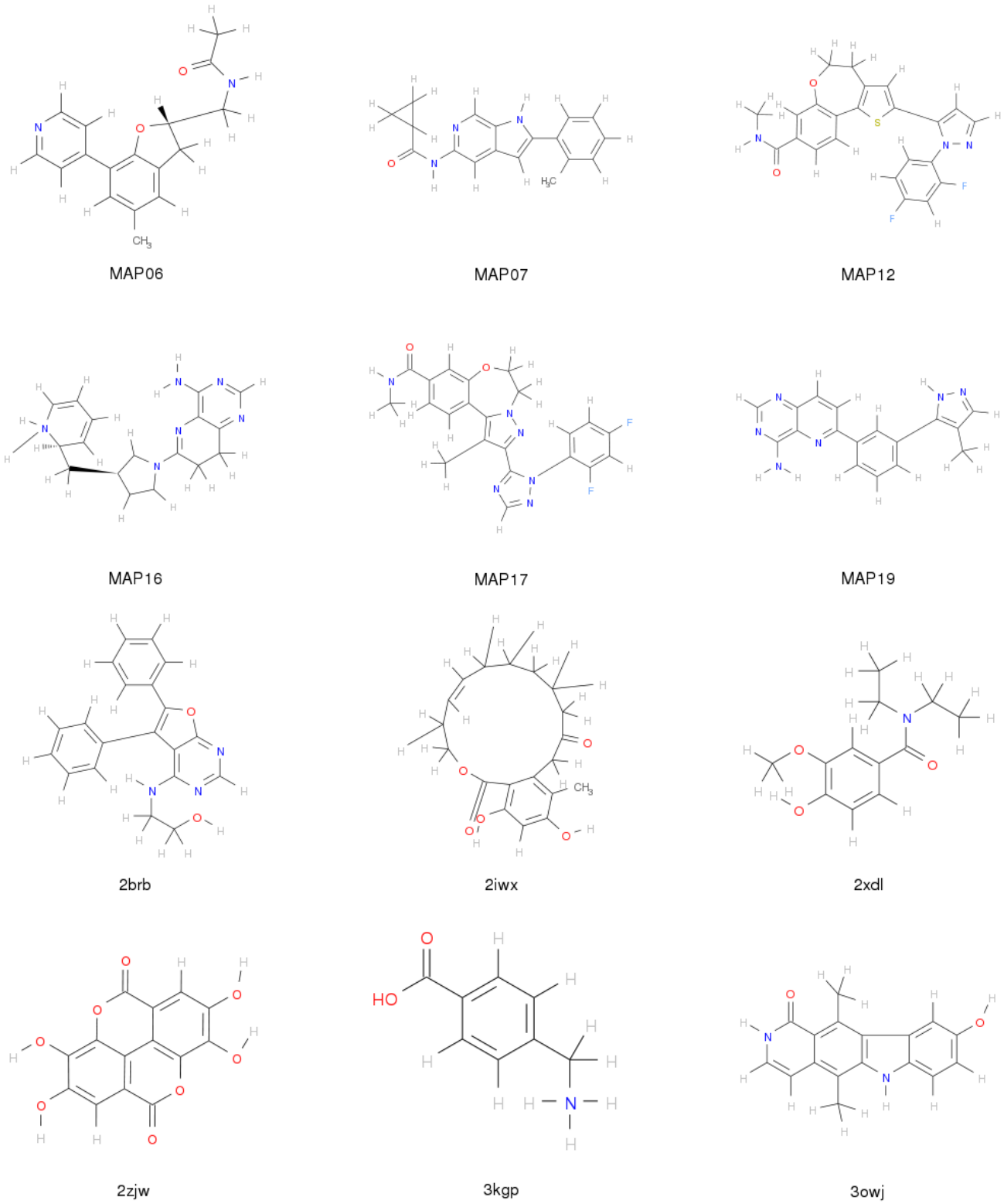

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

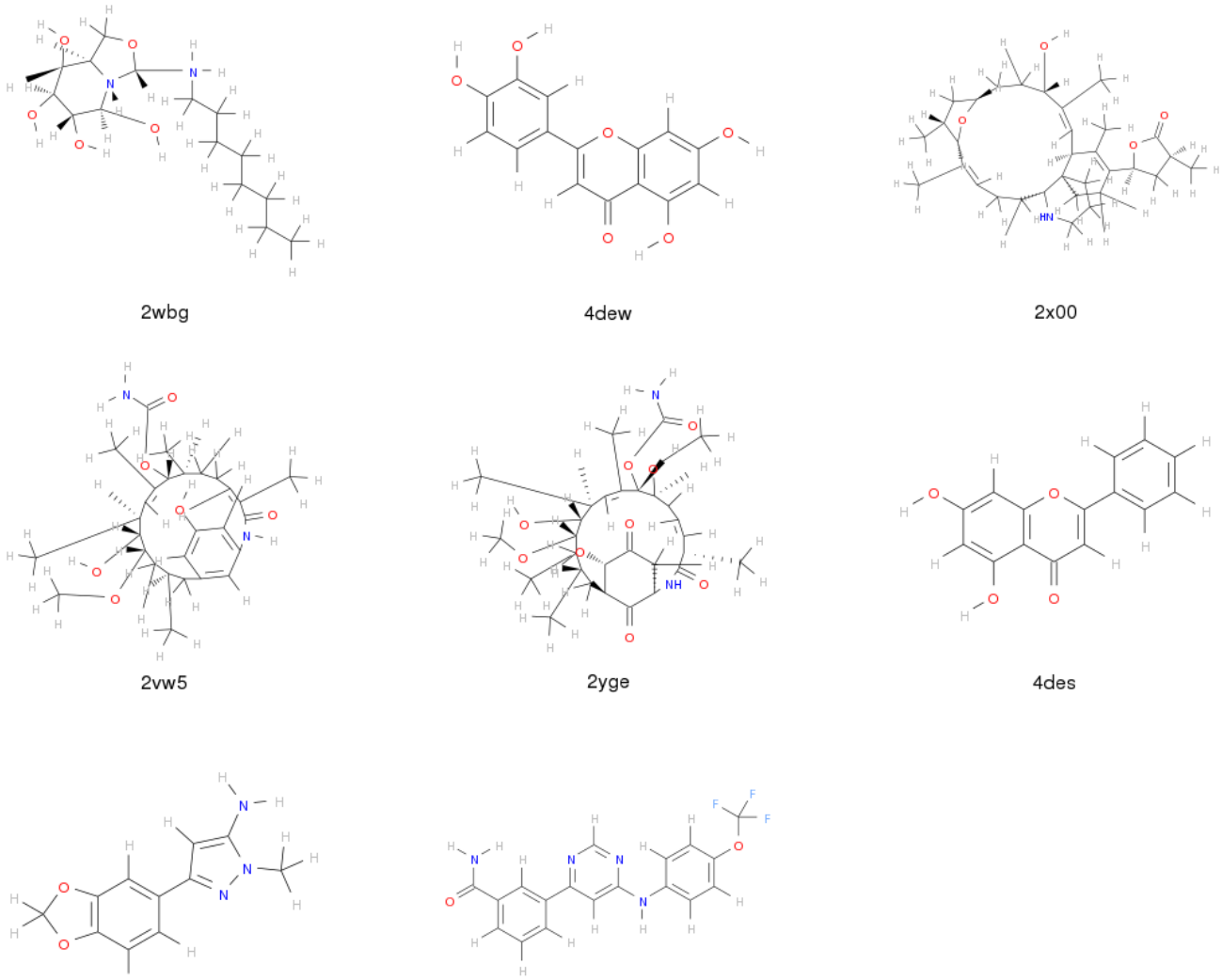

Figure S1: **Chemical structures of all ligands showing their protonation state, labeled by an identifier for the complex.** Diagrams were generated using Open Babel 2.3.2[5]

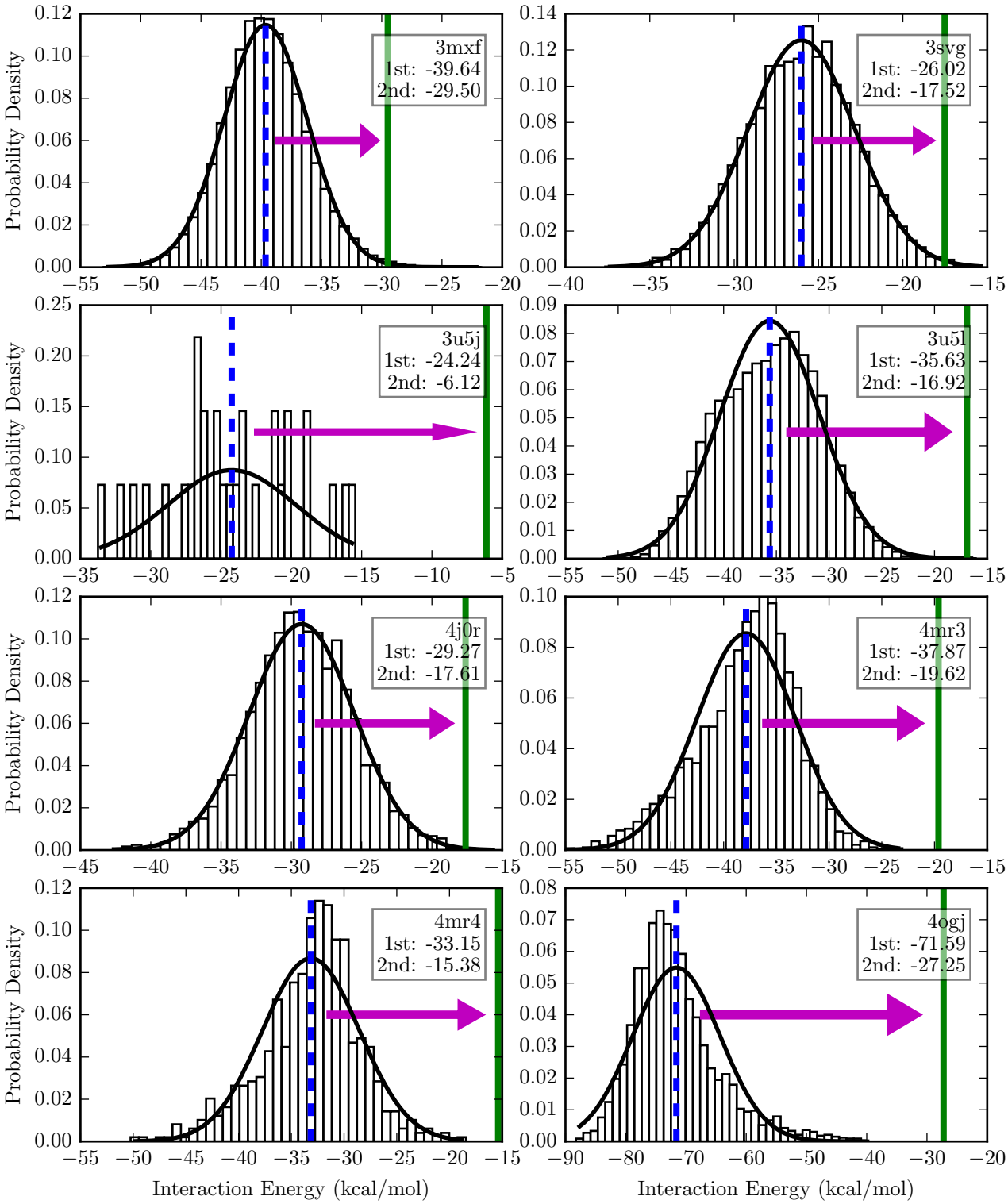

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 3mxf, 3svg, 3u5j, 3u5l, 4j0r, 4mr3, 4mr4, 4ogj from the bromodomains dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

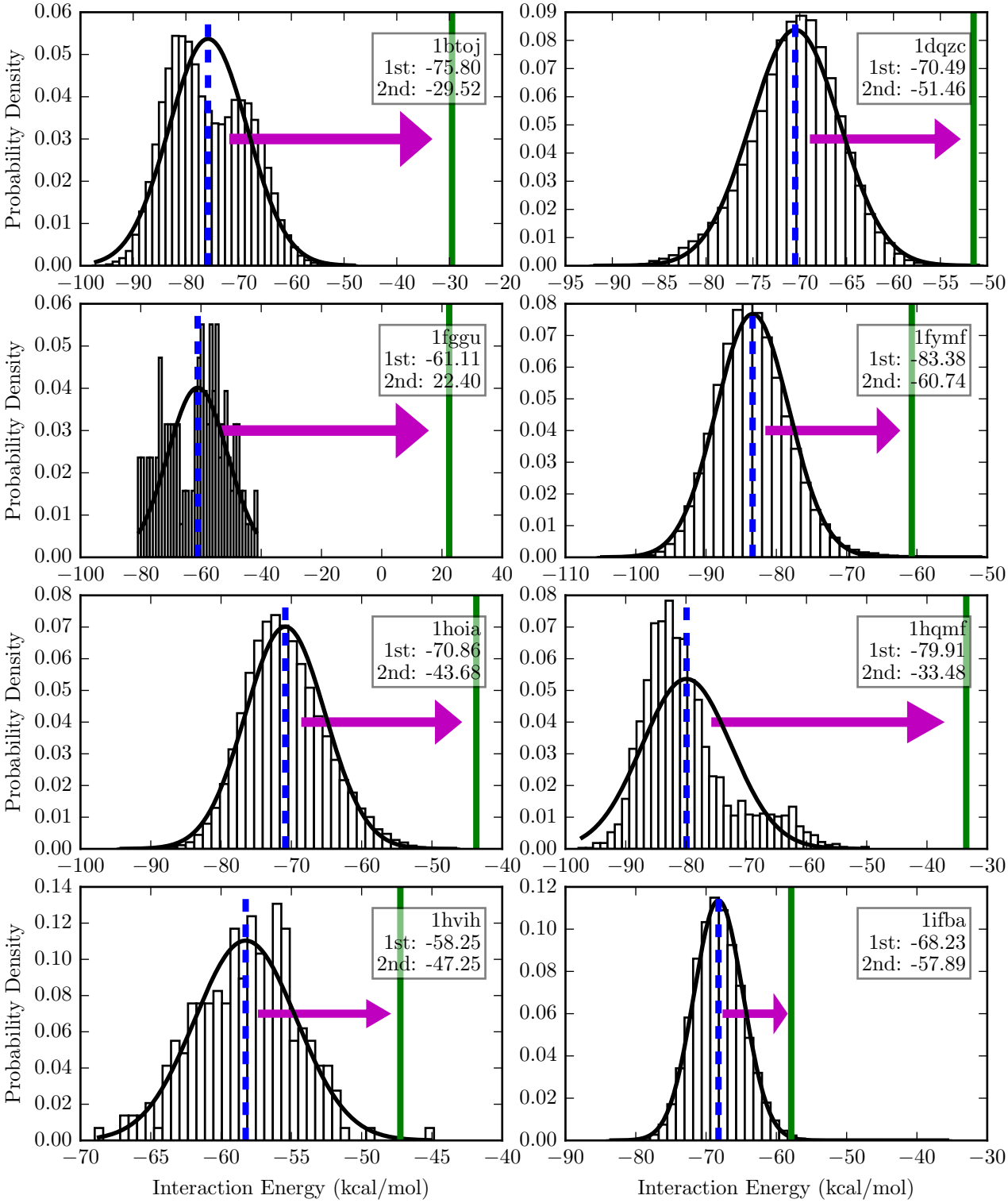

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 1btoj, 1dqzc, 1fggu, 1fymf, 1hoia, 1hqmf, 1hvih, 1ifba from the FXR dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

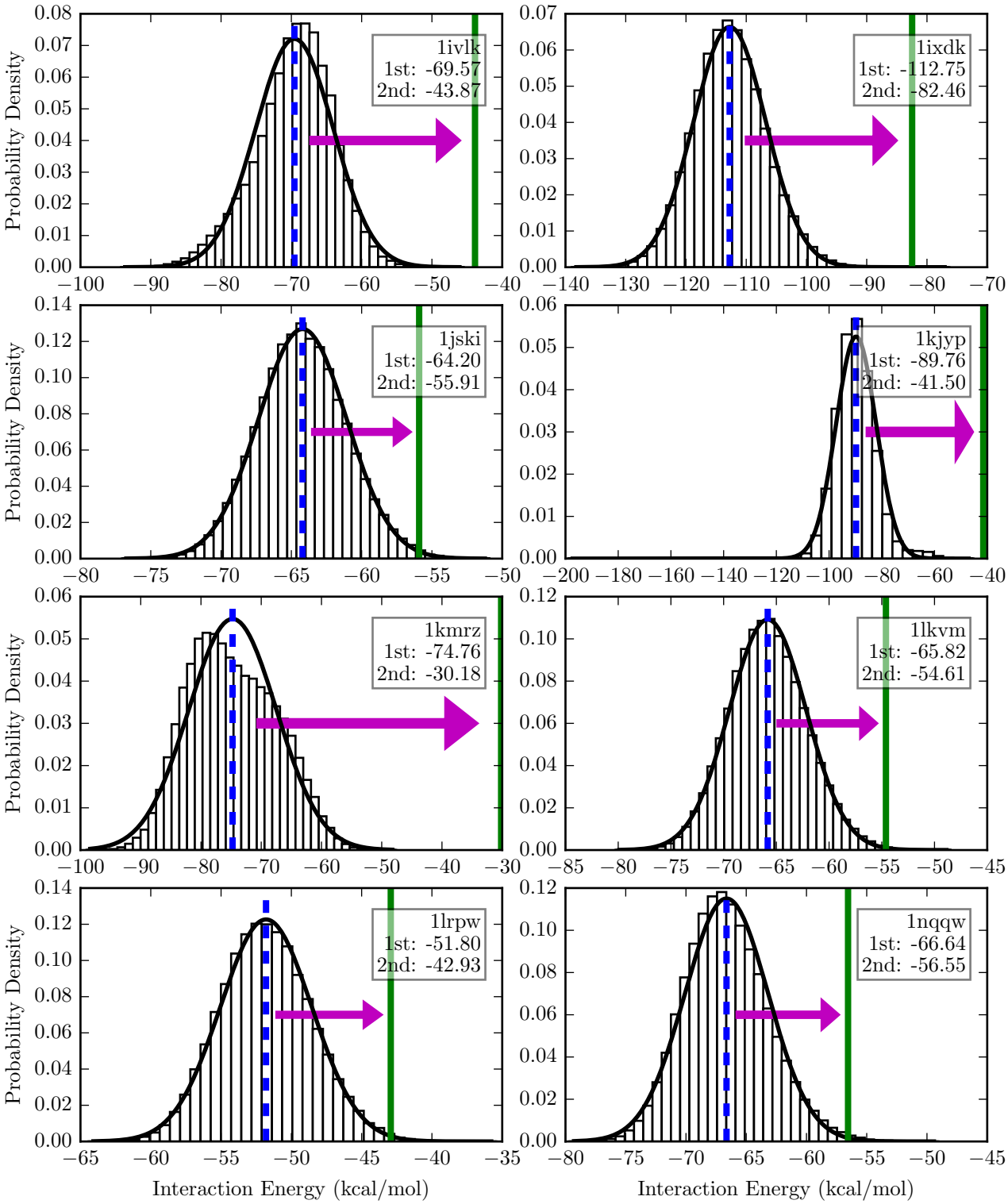

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 1ivlk, 1ixdk, 1jski, 1kjyp, 1kmrz, 1lkvm, 1lrpw, 1nqqw from the FXR dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

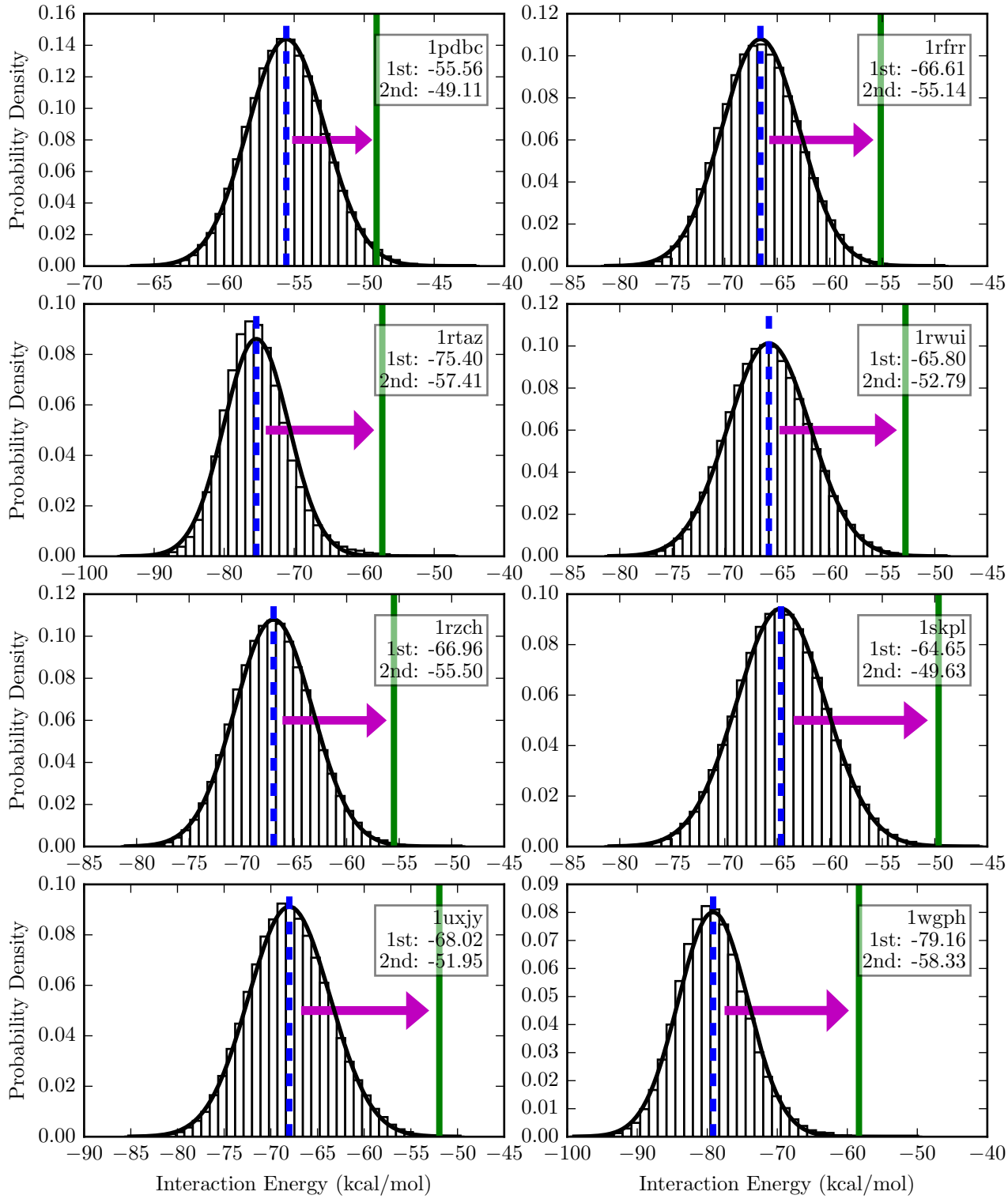

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 1pdbc, 1rfrr, 1rtaz, 1rwui, 1rzch, 1skpl, 1uxjy, 1wgph from the FXR dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

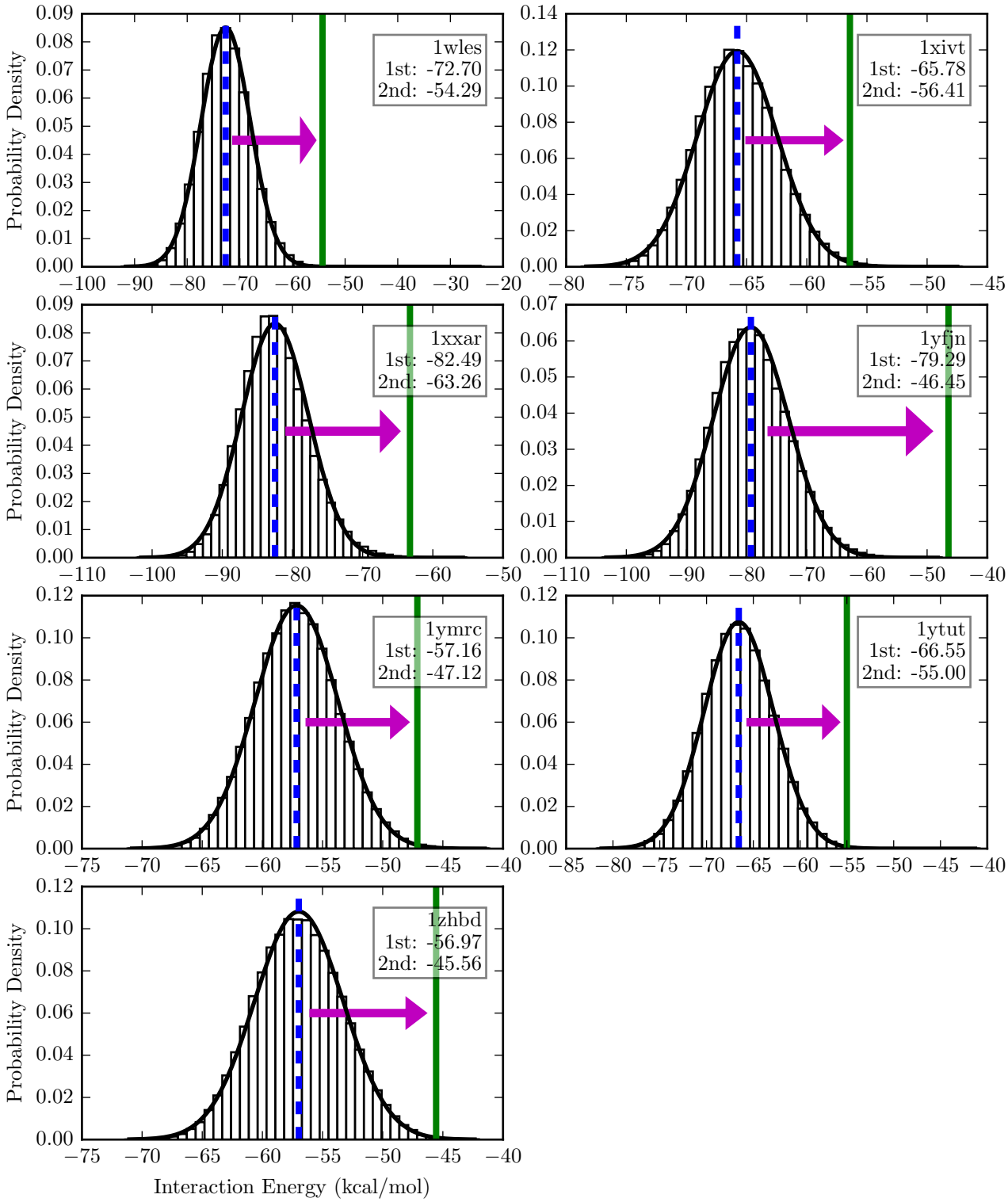

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 1wles, 1xivt, 1xxar, 1yfjn, 1ymrc, 1ytut, 1zhbd from the FXR dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

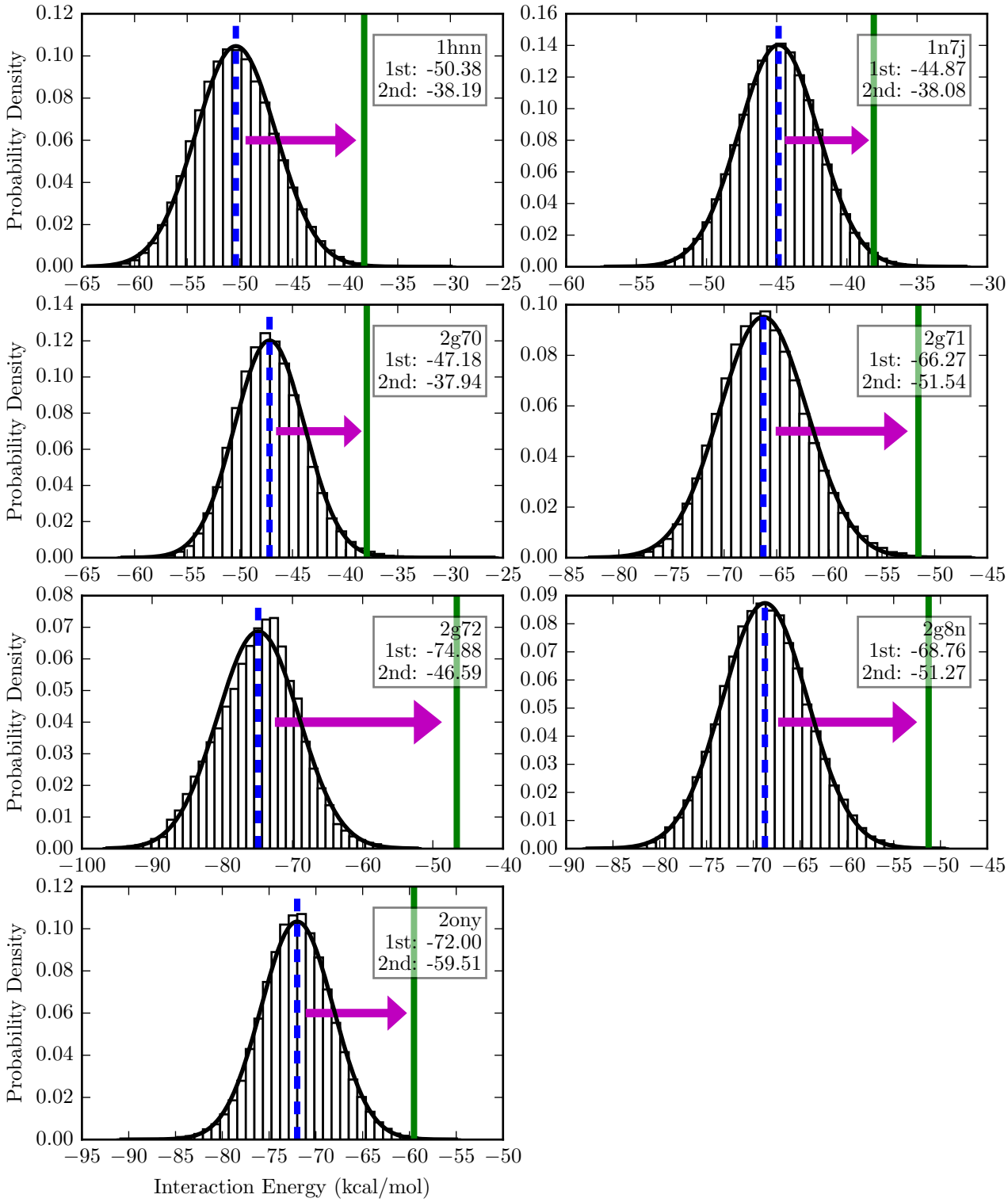

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 1hnn, 1n7j, 2g70, 2g71, 2g72, 2g8n, 2ony from the hPNMT dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

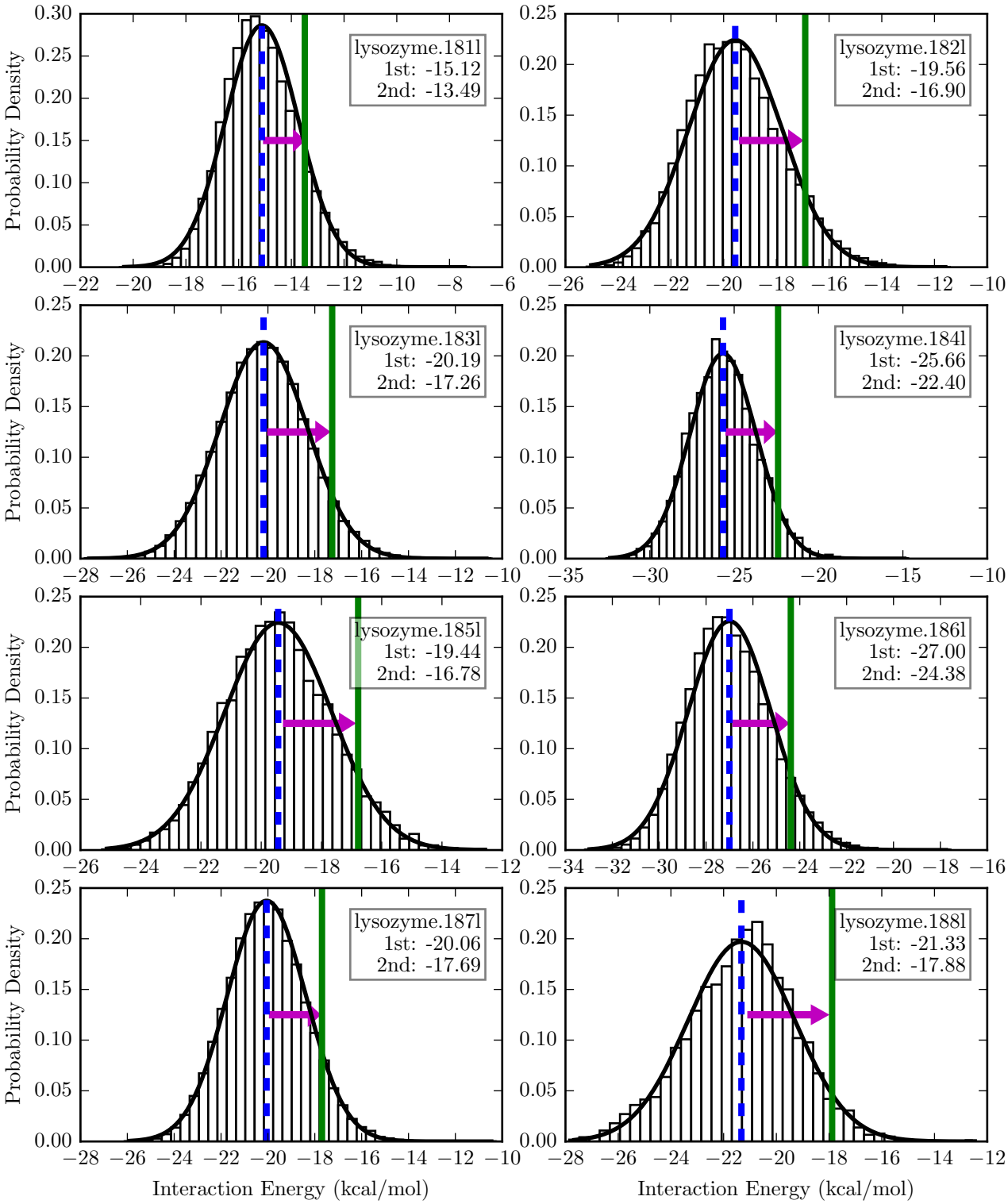

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 181l, 182l, 183l, 184l, 185l, 186l, 187l, 188l from the lysozyme dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

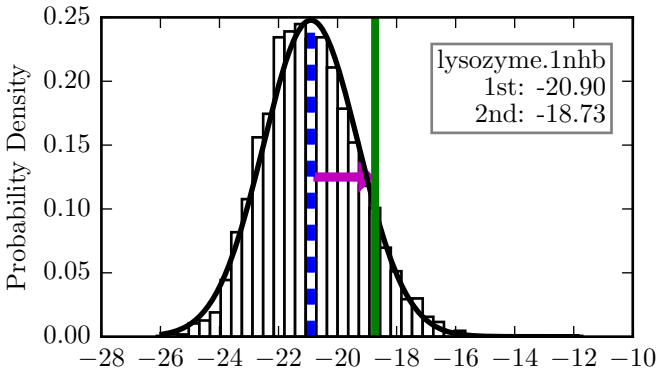

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 1nhb from the lysozyme dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

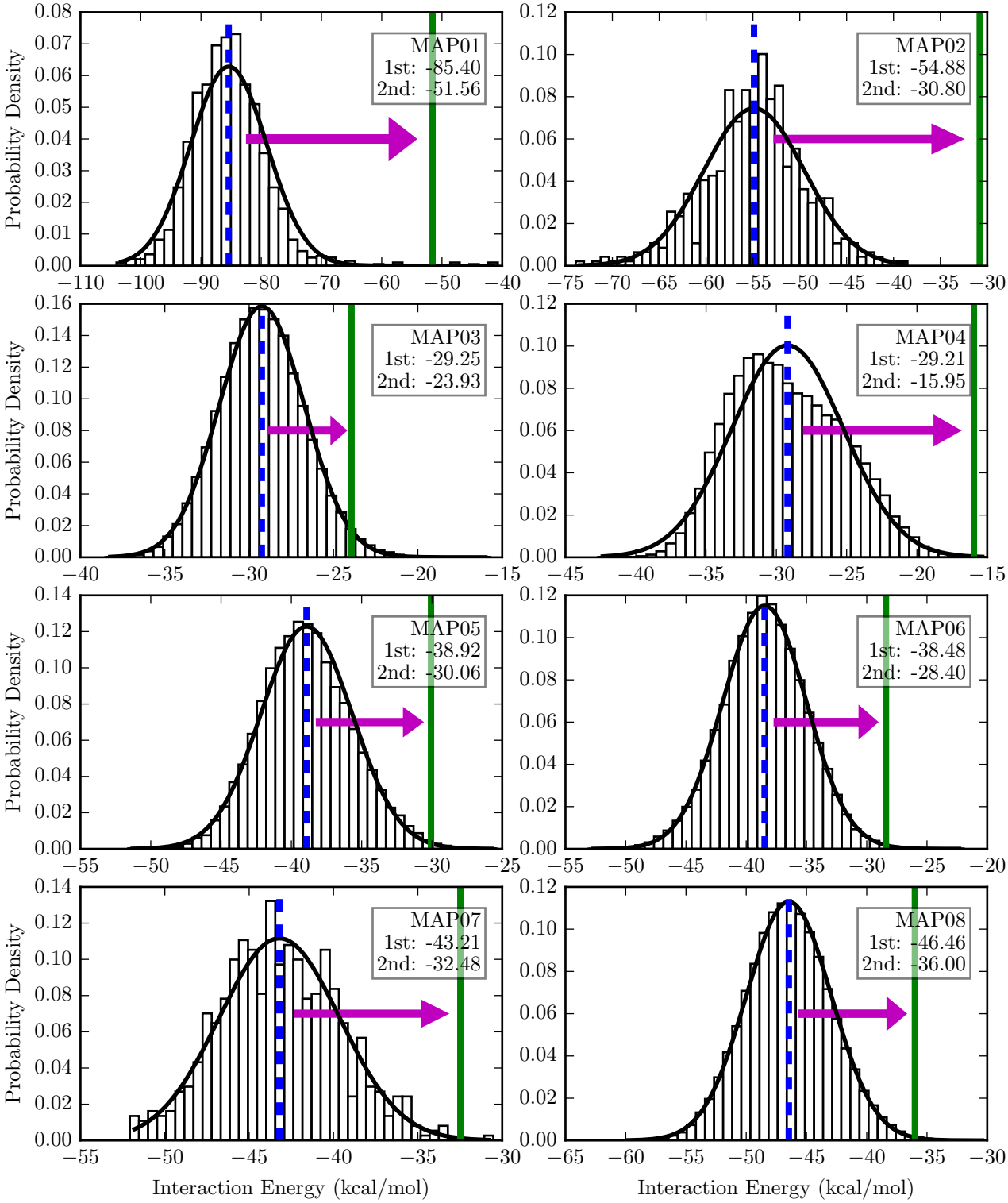

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): MAP01, MAP02, MAP03, MAP04, MAP05, MAP06, MAP07, MAP08 from the MAPK4K4 dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

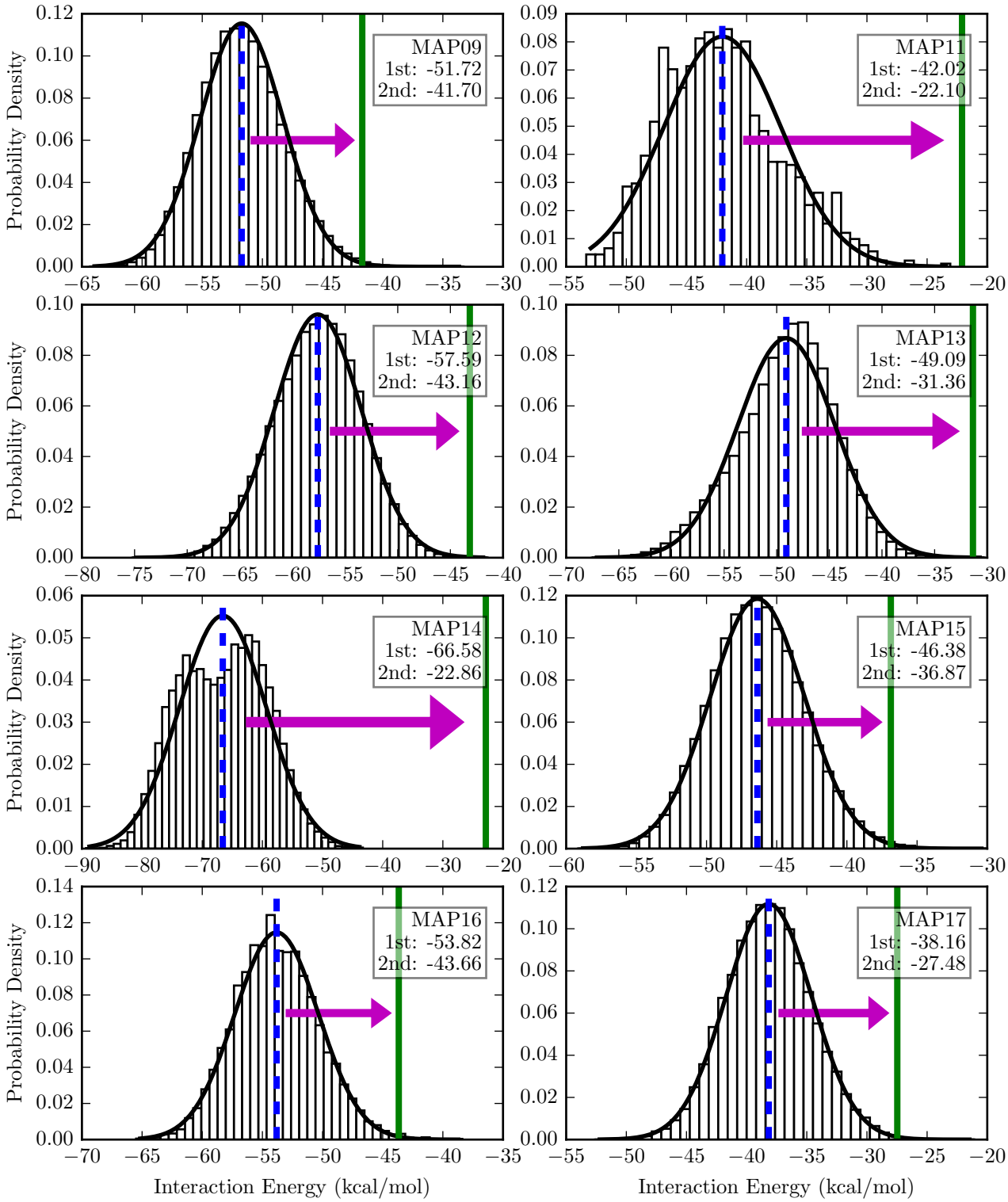

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): MAP09, MAP11, MAP12, MAP13, MAP14, MAP15, MAP16, MAP17 from the MAPK4K4 dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

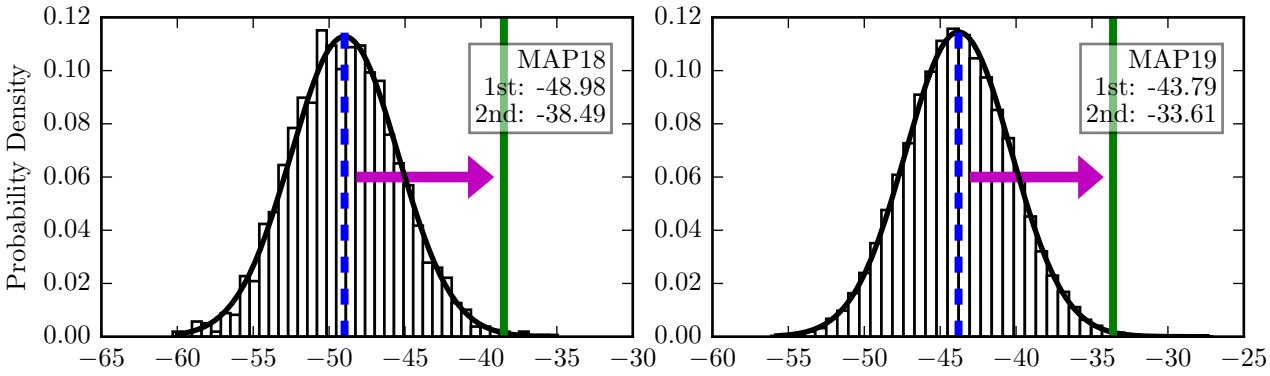

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): MAP18, MAP19 from the MAPK4K4 dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

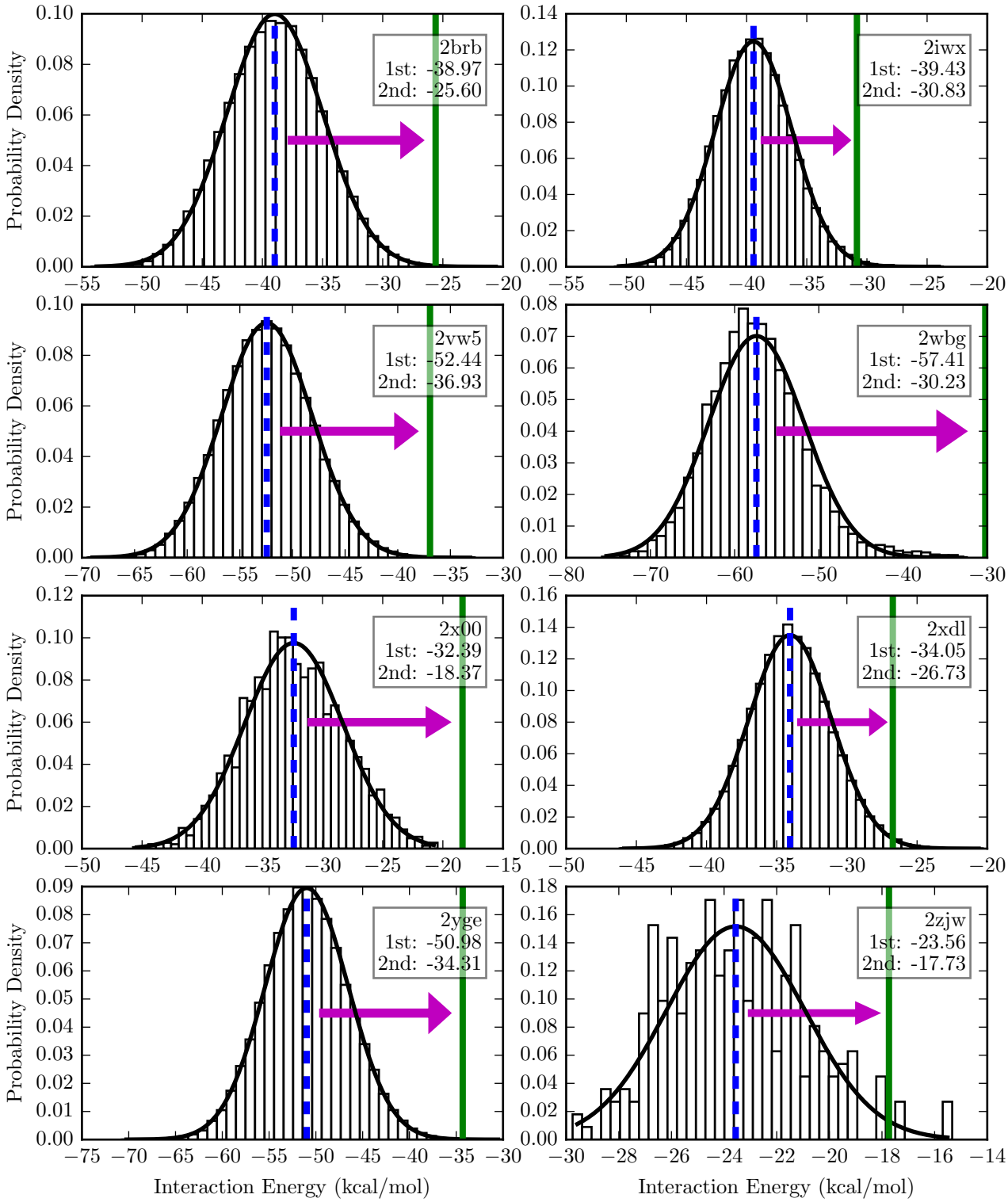

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 2brb, 2iwx, 2vw5, 2wbg, 2x00, 2xdl, 2yge, 2zjw from the IE dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

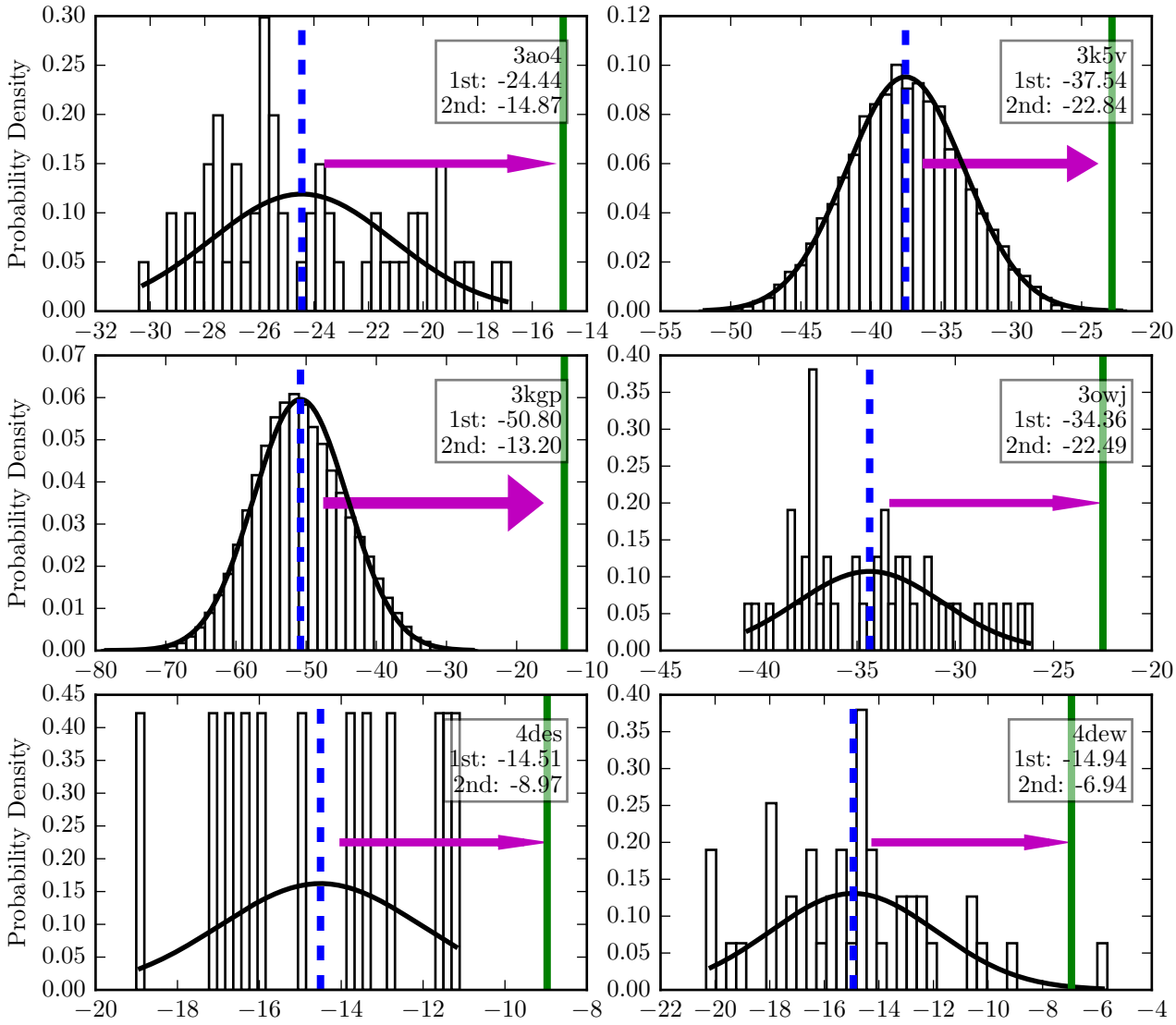

Figure S2: The bars are a normalized histogram of interaction energies observed in simulations of protein-ligand complexes (after ligand RMSD filtering): 3ao4, 3k5v, 3kgp, 3owj, 4des, 4dew from the IE dataset. The mean interaction energy is shown with a dashed blue vertical line. The second-order truncation of the cumulant expansion is shown with a solid green line. A Gaussian distribution based on the sample mean and standard deviation is shown as a solid black line.

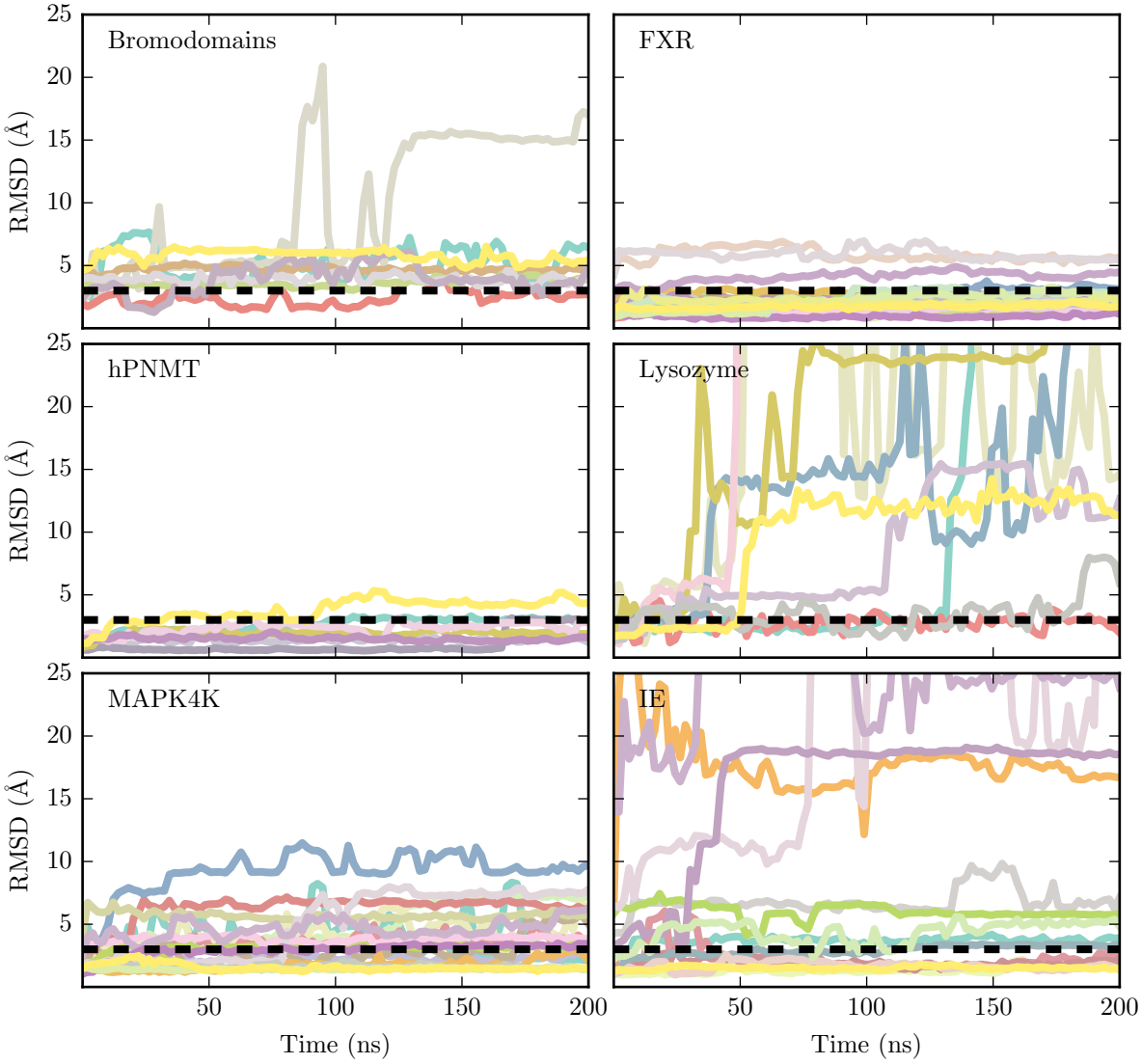

Figure S3: **Ligand RMSD as a function of time.** The running average of the ligand RMSD compared to the crystal structure over 2 ns intervals is shown. The y axis is restricted to between 0 and 25 Å. The dashed line is at 3 Å, which we used as the native pose cutoff.

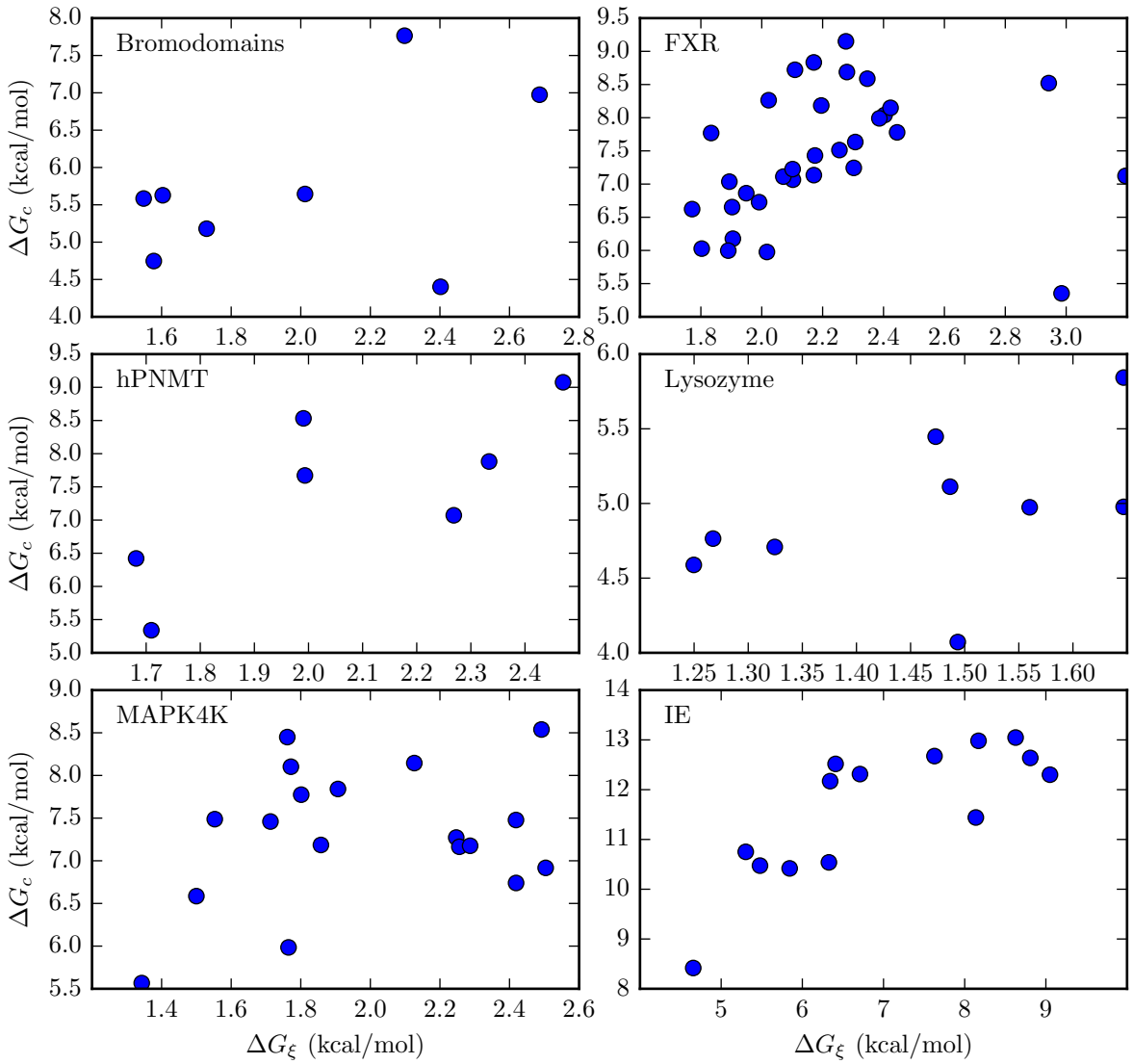

Figure S4: **Free energy corrections based on ligand external entropy:** restricting the ligand to the binding site (ΔG_ξ) and of imposing a confining potential on translational and rotational degrees of freedom (ΔG_c). The latter is based on $\Delta G_c = \Delta G_{c,L} - \Delta G_{c,RL}$ and does not include importance sampling effects through ΔG^c_{RL}.

```
#! /usr/bin/env python

import os, sys, glob, pickle, gzip
import numpy as np
from numpy import arange, dtype
import scipy.stats
from scipy.stats import spearmanr, kendalltau, gaussian_kde
import math
import itertools
import pymbar

BETA = 1./(8.3144621E-3/4.184*300) # mol/kcal
rmsd_filters = [3.0, np.inf]

if not 'system_dir' in locals().keys():
  system_dir = '/home/cli78/MM_PBSA/bromodomains/'

#########################
# Parse input arguments #
#########################

import argparse
parser = argparse.ArgumentParser()
parser.add_argument('--reference_data', \
  default=system_dir + 'experimental.dat', \
  help='File with reference free energies')

parser.add_argument('--sampler_dir', \
  default=system_dir, \
  help='Directory with sampler energies')
parser.add_argument('--sampler_total_E_FN', \
  default='trajectory.log', \
  help='File name for sampler complex energies')

parser.add_argument('--nm_dir', \
  default=system_dir, \
  help='Directory with normal modes analysis')
parser.add_argument('--nm_RL_FN', \
  default='complex_nm.out', \
  help='File name for normal modes of the complex')
parser.add_argument('--nm_R_FN', \
  default='receptor_nm.out', \
  help='File name for normal modes of the receptor')
parser.add_argument('--nm_L_FN', \
  default='ligand_nm.out', \
```

```
  help='File name for normal modes of the ligand')

parser.add_argument('--GBSA_dir', \
  default=system_dir, \
  help='Directory with GBSA energies')
parser.add_argument('--GBSA_RL_FN', \
  default='trajectory.log', \
  help='File name for GBSA complex energies')
parser.add_argument('--GBSA_R_FN', \
  default='protein.dat', \
  help='File name for GBSA receptor energies')
parser.add_argument('--GBSA_L_FN', \
  default='ligand.dat', \
  help='File name for GBSA ligand energies')

parser.add_argument('--ligand_dir',
  default=system_dir, \
  help='Directory with ligand RMSDs, prmtop, and dcd files')
parser.add_argument('--ligand_RMSD_FN', \
  default='ligand_rmsd.dat', help='File name for ligand rmsds')
parser.add_argument('--ligand_prmtop_FN', \
  default='vacuum.prmtop', help='File name for ligand prmtop')
parser.add_argument('--ligand_dcd_FN', \
  default='ligand_aligned.dcd', \
  help='Trajectory of the ligand from the simulation of the complex')

parser.add_argument('--skip_analysis', \
  action='store_true', \
  help='Does not do analysis')
parser.add_argument('-f', help='Dummy argument for ipython')

args = parser.parse_args()
del argparse

########################
# Read reference data #
########################

F = open(args.reference_data)
dat = F.readlines()
F.close()

ref_dG = {}
for line in dat:
  (id, dG) = line.split()
```

```
  ref_dG[id] = float(dG)
ids = ref_dG.keys()
del dat, F

#############
# Load data #
#############

def load_potential_energy(FN):
  """
  Loads a file with potential energies, returning them in kcal/mol
  """
  if os.path.isfile(FN):
    F = open(FN, 'r')
    dat = F.read().strip()
    F.close()
    if FN.endswith('.log'):
      lines = dat.split('\n')
      return np.array([float(line.split('\t')[3]) for line in lines[1:]])/4.184
    elif FN.endswith('.dat'):
      lines = dat.split('\n')
      return np.array([float(line) for line in lines])/4.184
    elif FN.endswith('.0'):
      frames = dat.split('Processing frame')[1:]
      energies = np.array(\
        [f.replace('1-4 ','1-4_')[f.find('BOND'):].split()[2:35:3] \
          for f in frames],dtype=float)
      # return np.sum(energies,1)/4.184
      return np.sum(energies[:,[3,4,5,9,10]],1)/4.184
    elif FN.endswith('.out'):
      # The normal modes output is in cal/mol/K
      # The conversion is to kcal/mol at 298.15 K
      if dat=='':
        print 'Empty file '+FN
        return np.array(np.nan)
      entropies = np.array([d.split()[-1] for d in dat.split('\n')], \
        dtype=float)*298.15/1000.
      return entropies
      pass
    else:
      raise Exception('Extension not recognized')
  else:
    print 'Missing file '+FN
    return np.array(np.nan)
```

```
def load_RMSD(FN, nvals):
  """
  Loads a file with ligand RMSDs, in Angstroms
  """
  if os.path.isfile(FN):
    F = open(FN, 'r')
    dat = F.read().strip()
    F.close()
    return np.array([np.float(d.split()[1]) for d in dat.split('\n')[1:]])
  else:
    print 'Missing file ' + FN
    return np.zeros(nvals)

def get_external_dof(traj):
  '''
  Calculate external degrees of freedom:
  (1) the center of mass and
  (2) the euler angle based on the principle axes of rotation
  '''
  import mdtraj
  com = mdtraj.compute_center_of_mass(traj)
  inertial_tensors = mdtraj.geometry.compute_inertia_tensor(traj)

  euler_angles = []
  # Principal axes are eigenvectors of the interial tensors
  # This makes the first principal axis Z, second Y, and third X
  for inertial_tensor in inertial_tensors:
    [w,v] = np.linalg.eig(inertial_tensor)
    (X,Y,Z) = [v[:,i] for i in np.argsort(w)]
    # Euler angles are determined from principal axes
    # https://en.wikipedia.org/wiki/Euler_angles
    euler_angles.append([np.arctan2(Z[0],Z[1]), \
                         np.arccos(Z[2]), \
                         np.arctan2(X[2],Y[2])])
  euler_angles = np.array(euler_angles)
  return (com, euler_angles)

Es = dict([(id,{}) for id in ids])

for id in ids:
  Es[id]['sampler'] = load_potential_energy(\
    os.path.join(args.sampler_dir,id,args.sampler_total_E_FN))
  for solvent in ['GBSA','nm']:
    dir = os.path.join(getattr(args,solvent+'_dir'),id)
    for moeity in ['L','R','RL']:
```

```
      Es[id]['%s_%s'%(solvent,moeity)] = load_potential_energy(\
        os.path.join(dir, getattr(args,'%s_%s_FN'%(solvent,moeity))))
  Es[id]['RMSD'] = load_RMSD(\
    os.path.join(args.ligand_dir,id,args.ligand_RMSD_FN), \
    len(Es[id]['sampler']))
  if np.any([np.any(np.isnan(Es[id][key])) for key in Es[id].keys()]):
    print 'Removing '+id
    del Es[id]
    ids = [item for item in ids if not item==id]

import mdtraj
from netCDF4 import Dataset
import sys
sys.stdout.write('Getting external degrees of freedom for')

if not os.path.isdir('external_dof'):
  os.mkdir('external_dof')
for id in ids:
  sys.stdout.write(' ' + id)

  loaded = False
  FN = os.path.join('external_dof',id + '.nc')
  if os.path.isfile(FN):
    try:
      nc = Dataset(FN,'r')
      Es[id]['com'] = nc.variables['com'][:]
      Es[id]['euler_angles'] = nc.variables['euler_angles'][:]
      nc.close()
      loaded = True
    except:
      loaded = False
  if not loaded:
    traj_FN = os.path.join(args.ligand_dir,id,args.ligand_dcd_FN)
    if not os.path.isfile(traj_FN):
      print 'Ligand trajectory missing in '+traj_FN
      (Es[id]['com'], Es[id]['euler_angles']) = (np.zeros(3), np.zeros(3))
      continue
    prmtop_FN = os.path.join(args.ligand_dir,id,args.ligand_prmtop_FN)
    traj = mdtraj.load(traj_FN, top=prmtop_FN)
    (Es[id]['com'], Es[id]['euler_angles']) = get_external_dof(traj)
    nc = Dataset(FN,'w',format='NETCDF4')
    nc.createDimension('n_frames', Es[id]['com'].shape[0])
    nc.createDimension('n_coords', 3)
    nc.createVariable('com','f8',('n_frames','n_coords'))
    nc.createVariable('euler_angles','f8',('n_frames','n_coords'))
```

```
    nc.variables['com'][:] = Es[id]['com']
    nc.variables['euler_angles'][:] = Es[id]['euler_angles']
    nc.close()
sys.stdout.write('\n\n')

print 'Number of energies:'
print '\tSampler\t|\tGBSA\t\t\t|\tnm\t\t\t|\tCOM\t|\tRMSD'
for id in ids:
  count_str = id + '\t%d'%(len(Es[id]['sampler'])) + '\t|\t'
  count_str += '\t|\t'.join(['\t'.join([\
      'nan' if np.isnan(Es[id]['%s_%s'%(solvent,moeity)]).any() \
      else '%d'%len(Es[id]['%s_%s'%(solvent,moeity)]) \
        for moeity in ['L','R','RL']]) for solvent in ['GBSA','nm']])
  count_str += '\t|\t' + '%d'%len(Es[id]['com'])
  count_str += '\t|\t' + '%d'%len(Es[id]['RMSD'])
  print count_str
print

#################################
# Binding free energy estimates #
#################################

def estimate_dG(Psi, nm_TdS, DeltaU=None):
  """
  Returns an array of free energy estimates:
  based on cumulants 1-4, EXP, and then NM.
  Psi are interaction energies.
  nm_TdS are entropy terms from normal modes analysis.
  DeltaU are reweighting factors for individual configurations.
  """
  if DeltaU is None:
    cumulants = [np.nan] + [scipy.stats.kstat(Psi,n) for n in range(1,5)]
    dG = np.cumsum([BETA**(n-1)*cumulants[n]/math.factorial(n) \
      for n in range(1,5)])
    exp = (np.log(np.mean(np.exp(BETA*Psi-max(BETA*Psi))))+max(BETA*Psi))/BETA
  else:
    num = Psi - DeltaU
    den = -DeltaU
    cumulants_num = [np.nan] + [scipy.stats.kstat(num,n) for n in range(1,5)]
    cumulants_den = [np.nan] + [scipy.stats.kstat(den,n) for n in range(1,5)]
    dG = np.cumsum([BETA**(n-1)*(cumulants_num[n] - cumulants_den[n])/\
      math.factorial(n) for n in range(1,5)])
    exp = (np.log(np.average(np.exp(BETA*num-max(BETA*num)))) + max(BETA*num) \
      - np.log(np.average(np.exp(BETA*den-max(BETA*den)))) - max(BETA*den))/BETA
  return np.array(list(dG) + [exp, dG[0] - nm_TdS])
```

```
def calc_confinement_fe(com, euler, nbins=100):
  """
  Obtains the free energy of
  (1) putting the unbound ligand into the binding site, dG_xi
  (2) adding constraints to the unbound ligand and
      removing them from the bound complex, dG_c
  """
  # Scott's rule for Gaussian kernel density estimate bandwidth
  npts = com.shape[0]
  scotts_factor = np.power(npts,(-1./5))

  dG_xi = []
  dG_c = []
  U_c = []

  # Translational degrees of freedom

  # The standard state volume for a single molecule
  # in a box of size 1 L is 1.66053928 nanometers**3
  L_o = 1.66053928**(1/3.)

  def calc_translational(x):
    tail = (max(x)-min(x))*0.1
    edges = np.linspace(min(x)-tail, max(x)+tail, nbins)
    delta = (edges[1]-edges[0])
    centers = edges + delta/2.

    # Obtain confinement potential
    kernel = gaussian_kde(x)
    rho = kernel(centers)
    U = -np.log(rho)/BETA

    # Obtain confinement energies for bound state
    inds = np.array(np.floor(\
      (x-edges[0])/delta),dtype=np.int)
    inds[inds==nbins] = nbins-1
    DeltaU = U[inds]

    # Estimate confinement free energies
    # For dG_cL, use numerical integral for expectation
    dG_cL = -1/BETA*np.log(np.mean(np.exp(-BETA*U)))
    # For dG_cRL, use sample mean estimator
    dG_cRL = -1/BETA*np.log(np.mean(\
      np.exp(-BETA*DeltaU+min(BETA*DeltaU)))) + min(DeltaU)
```

```
    # Standard state correction
    L = edges[-1]-edges[0]+delta
    dG_xi = -1/BETA * np.log(L / L_o)

    return dG_xi, dG_cL - dG_cRL, DeltaU

  # Performs principal components analysis
  com_c = com - np.mean(com, 0)
  [w,v] = np.linalg.eig(np.dot(com_c.T,com_c)/com_c.shape[0])
  com_pca = np.dot(com_c,v)

  for dim in range(3):
    dG_xi_d, dG_c_d, U_c_d = calc_translational(com_pca[:,dim])
    dG_xi.append(dG_xi_d)
    dG_c.extend([dG_xi_d + dG_c_d])
    U_c.append(U_c_d)

  # Rotational degrees of freedom

  edges_2pi = np.linspace(-np.pi,np.pi,nbins)
  edges_pi = np.linspace(0,np.pi,nbins)

  def calc_rotational(x, edges, sine=False):
    delta = edges[1]-edges[0]
    centers = edges[:-1] + delta/2.

    # Obtain confinement potential
# See https://stackoverflow.com/questions/18921419/\
# implementing-a-2d-fft-based-kernel-density-estimator-in-python-and-comparing-i
    H = np.histogram(x,edges,density=True)[0]
    sigma = scotts_factor*10.*delta
    ker = np.exp(-(centers-centers[int(len(centers)/2)])**2/\
                 2/sigma**2)/sigma/np.sqrt(2*np.pi)
    rho = np.abs(np.fft.fftshift(np.fft.ifft(\
      np.fft.fft(H)*np.fft.fft(ker))).real)
    rho /= np.sum(rho)*delta
    U = -np.log(np.array(rho))/BETA

    # Obtain confinement energies for bound state
    inds = np.array(np.floor(\
      (x-edges[0])/delta),dtype=np.int)
    inds[inds==nbins] = nbins-1
    DeltaU = U[inds]

    # Estimate confinement free energies
```

```
    # For dG_cL, use numerical integral for expectation
    if sine:
      dG_cL = -1/BETA*np.log(np.sum(np.sin(centers)*np.exp(-BETA*U))/\
                             np.sum(np.sin(centers)))
    else:
      dG_cL = -1/BETA*np.log(np.mean(np.exp(-BETA*U)))
    # For dG_cRL, use sample mean estimator
    dG_cRL = -1/BETA*np.log(np.mean(\
      np.exp(-BETA*DeltaU+min(BETA*DeltaU)))) + min(DeltaU)

    return dG_cL - dG_cRL, DeltaU

  dG_c_d, U_c_d = calc_rotational(euler[:,0], edges_2pi)
  dG_c.append(dG_c_d)
  U_c.append(U_c_d)
  dG_c_d, U_c_d = calc_rotational(euler[:,1], edges_pi, sine=True)
  dG_c.append(dG_c_d)
  U_c.append(U_c_d)
  dG_c_d, U_c_d = calc_rotational(euler[:,2], edges_2pi)
  dG_c.append(dG_c_d)
  U_c.append(U_c_d)

  return np.sum(dG_xi), np.sum(dG_c), np.sum(U_c,0)

dG = dict([(id,{}) for id in ids])

# Actually does free energy estimates
for id in ids:
  for solvent in ['GBSA']:
    for rmsd_filter in rmsd_filters:
      stride = len(Es[id]['RMSD'])/len(Es[id][solvent+'_RL'])
      ok_rmsd = Es[id]['RMSD'][::stride]<rmsd_filter
      Psi = (Es[id][solvent+'_RL'] - Es[id][solvent+'_R'] \
             - Es[id][solvent+'_L'])[ok_rmsd]

      # Entropy from normal modes analysis
      if not np.isnan(Es[id]['nm_RL']).any():
        nm = Es[id]['nm_RL'] - Es[id]['nm_R'] - Es[id]['nm_L']
        stride = len(ok_rmsd)/len(Es[id]['nm_RL'])
        nm_TdS = np.mean(nm[ok_rmsd[::stride]])
      else:
        nm_TdS = np.nan

      # Free energy estimate without binding site correction
      # or ligand confinement free energy
```

```
      # and without reweighting
      dG[id]['%s_no_rw_%.2f_o'%(solvent,rmsd_filter)] = \
        estimate_dG(Psi, nm_TdS)
      # or with reweighting
      stride = len(Es[id]['sampler'])/len(Es[id][solvent+'_RL'])
      DeltaU = (Es[id][solvent+'_RL'] - Es[id]['sampler'][::stride])[ok_rmsd]
      dG[id]['%s_rw_%.2f_o'%(solvent,rmsd_filter)] = \
        estimate_dG(Psi, nm_TdS, DeltaU)

      # External dof corrections
      (dG_xi, dG_c, U_c) = calc_confinement_fe(\
        Es[id]['com'][ok_rmsd[:Es[id]['com'].shape[0]]], \
        Es[id]['euler_angles'][ok_rmsd[:Es[id]['euler_angles'].shape[0]]])

      # With binding site correction
      dG[id]['%s_no_rw_%.2f_xi'%(solvent, rmsd_filter)] = \
        estimate_dG(Psi, nm_TdS) + dG_xi
      dG[id]['%s_rw_%.2f_xi'%(solvent, rmsd_filter)] = \
        estimate_dG(Psi, nm_TdS, DeltaU) + dG_xi
      # With ligand confinement free energy
      dG[id]['%s_no_rw_%.2f_c'%(solvent, rmsd_filter)] = \
        estimate_dG(Psi[(Psi.shape[0] - U_c.shape[0]):], nm_TdS, \
        U_c) + dG_c
      dG[id]['%s_rw_%.2f_c'%(solvent, rmsd_filter)] = \
        estimate_dG(Psi[(Psi.shape[0] - U_c.shape[0]):], nm_TdS, \
        DeltaU[(Psi.shape[0] - U_c.shape[0]):] + U_c) + dG_c

for confinement in ['o','xi','c']:
  key = 'GBSA_no_rw_%.2f_%s'%(rmsd_filters[0], confinement)
  print 'Free energies for '+key
  for id in ids:
    print id + '\t' + ' '.join(['%6.2f'%fe \
      for fe in dG[id][key]])
  print

########################
# Performance analysis #
########################

from scipy.stats import spearmanr, kendalltau

def PearsonR(x, y):
  return np.corrcoef(x,y)[0,1]

def SpearmanR(x, y):
```

```
  return spearmanr(x,y)[0]

def KendallTau(x, y):
  return kendalltau(x,y)[0]

def RMSE(x,y):
  return np.sqrt(np.mean((x-y)**2))

def aRMSE(x,y):
  mean_diff = np.mean(x)-np.mean(y)
  return np.sqrt(np.mean((x-y-mean_diff)**2))

def bootstrap_std(x,y,func):
  vals = []
  rep = 0
  while rep<1000:
    inds = np.random.randint(len(x), size=len(x))
    val = func(x[inds],y[inds])
    if not np.isnan(val):
      vals.append(func(x[inds],y[inds]))
      rep += 1
  return np.std(vals)
metrics = ['PearsonR', 'SpearmanR', 'KendallTau', 'RMSE', 'aRMSE']

if not args.skip_analysis:
  dG_r = np.array([ref_dG[id] for id in ids])
  print 'Performance analysis:'
  print
  print '&&%17s'%'' + ''.join(['&%13s'%t \
    for t in ['1','2','3','4','EXP','1+nm']]) + '\\\\'

  solvent = 'GBSA' # also could be 'PBSA'
  rw = 'no_rw' # also could be 'rw'
  for rmsd_filter in rmsd_filters:
    for confinement in ['_o','_xi','_c']:
      prefix = repr(rmsd_filter) + '&' + \
        {'_o':' No', '_xi':'Site', '_c':'Yes'}[confinement] + '&'
      tables = {}
      for metric in metrics:
        tables[metric] = []
      rows = {}
      for metric in metrics:
        rows[metric] = [prefix + '%11s'%metric]
      key = '%s_%s_%.2f%s'%(solvent,rw,rmsd_filter,confinement)
      for order_ind in range(6):
```

```
  dG_t = np.array([dG[id][key][order_ind] for id in ids])
  rows['PearsonR'].append('&%6.2f (%.2f)'%(\
    PearsonR(dG_r,dG_t),bootstrap_std(dG_r,dG_t,locals()['PearsonR'])))
  rows['SpearmanR'].append('&%6.2f (%.2f)'%(\
    SpearmanR(dG_r,dG_t),bootstrap_std(dG_r,dG_t,locals()['SpearmanR'])))
  rows['KendallTau'].append('&%6.2f (%.2f)'%(\
    KendallTau(dG_r,dG_t),bootstrap_std(dG_r,dG_t,locals()['KendallTau'])))
  rows['RMSE'].append('&%6.2f (%.2f)'%(\
    RMSE(dG_r,dG_t),bootstrap_std(dG_r,dG_t,locals()['RMSE'])))
  rows['aRMSE'].append('&%6.2f (%.2f)'%(\
    aRMSE(dG_r,dG_t),bootstrap_std(dG_r,dG_t,locals()['aRMSE'])))
for metric in metrics:
  tables[metric].append(rows[metric][0] + \
    ''.join(['%14s'%r for r in rows[metric][1:]]) + '\\\\')

# Print out a table
for metric in metrics:
  print '\n'.join(tables[metric])
print
```

References

(1) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, *7*, 207–218.

(2) Gee, C. L.; Drinkwater, N.; Tyndall, J. D. A.; Grunewald, G. L.; Wu, Q.; McLeish, M. J.; Martin, J. L. Enzyme Adaptation to Inhibitor Binding: A Cryptic Binding Site in Phenylethanolamine N-Methyltransferase. *J. Med. Chem.* **2007**, *50*, 4845–4853.

(3) Morton, A.; Matthews, B. W. Specificity of Ligand Binding in a Buried Nonpolar Cavity of T4 Lysozyme: Linkage of Dynamics and Structural Plasticity. *Biochemistry* **1995**, *34*, 8576–8588.

(4) Duan, L.; Liu, X.; Zhang, J. Z. H. Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein-Ligand Binding Free Energy. *J. Am. Chem. Soc.* **2016**, *138*, 5722–5728.

(5) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminformatics* **2011**, *3*, 33.