Supporting Information For the Manuscript:

Evaluation of Several Two-Step Scoring Functions Based on Linear Interaction Energy, Effective Ligand Size, and Empirical Pair Potentials for Prediction of Protein-Ligand Binding Geometry and Free Energy.

Obaidur Rahaman,[1] Trilce P. Estrada[2], Douglas J. Doren,[1] Michela Taufer,[2] Charles L. Brooks III[3] and Roger S. Armen[4*]

[1]Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716 [2]Department of Computer and Information Sciences, University of Delaware, Newark, Delaware 19716 [3]Department of Chemistry, 930 N. University Ave, University of Michigan, Ann Arbor, MI 48109 [4]Department of Pharmaceutical Sciences, School of Pharmacy, Thomas Jefferson University, 130 S. 9th St, Ste 1510, Philadelphia, PA 19107.

Cross-validation groups for the dataset LPDB259 and LPDB160.

LPDB259:

(group1): 1a9m, 1abe, 1adb, 1adc, 1aeq, 1aet, 1avn, 1b6k, 1b6m, 1ba8, 1bai, 1bid, 1cil, 1cnx, 1did, 1dif, 1dwc, 1etr, 1fkf, 1fkg, 1flr, 1g2k, 1hhg, 1hhi, 1hpo, 1hsg, 1hsh, 1htf, 1hvh, 1hwr, 1ivp, 1mcf, 1mch, 1nnb, 1okl, 1ppc, 1tnh, 1tni, 1vac, 1xig, 2dbl, 2er7, 2web, 2wec, 2xis, 3ptb, 4er2, 4er4, 5cpp, 5hvp, 5tmn, 8abp.

(group2): 1a30, 1a4k, 1a50, 1aaq, 1ac4, 1ae8, 1aed, 1aek, 1aeo, 1afl, 1atl, 1bwa, 1bwb, 1bxo, 1bzm, 1c83, 1c87, 1c8k, 1cbs, 1cbx, 1cny, 1cvu, 1d3h, 1dbb, 1dbk, 1dbm, 1die, 1dwd, 1ela, 1ets, 1g35, 1hii, 1hiv, 1hps, 1hpv, 1hpx, 1hvj, 1hvk, 1k6p, 1ldm, 1lyb, 1pph, 1rne, 1sre, 2cpp, 2ifb, 3tmn, 4tln, 6abp, 6tmn, 7abp, 7hvp.

(group3): 1aeb, 1aee, 1aes, 1aeu, 1aev, 1anf, 1apu, 1apv, 1b6j, 1b6l, 1bim, 1bmn, 1bto, 1bv9, 1byg, 1cin, 1cpi, 1ctr, 1d3p, 1d4l, 1d4p, 1dbj, 1ent, 1fax, 1fkn, 1gno, 1hbv, 1hge, 1hhk, 1hih, 1hld, 1hxb, 1hxw, 1mcj, 1mcs, 1nsc, 1rbp, 2tsc, 2upj, 2wea, 3er3, 3er5, 4dfr, 4er1, 4hmg, 4tmn, 5abp, 5cna, 5tln, 9aat, 9abp, 9hvp.

(group4): 1a4w, 1abf, 1aeg, 1aeh, 1aen, 1afk, 1ajv, 1aoe, 1apw, 1b38, 1bap, 1bdr, 1bmm, 1c84, 1c88, 1cnw, 1ctt, 1dog, 1dwb, 1ebg, 1eed, 1ejn, 1ett, 1hdt, 1hew, 1hgj, 1hhj, 1hos, 1hrn, 1htg, 1ida, 1k6c, 1ldy, 1lgr, 1mcb, 1nsd, 1ppl, 1pso, 1sbp, 1tmn, 1tng, 1tnj, 2ctc, 2gbp, 2phh, 2vaa, 2vab, 3bto, 4hvp, 4ts1, 6cpa, 8cpp.

(group5): 1a07, 1ac0, 1ac8, 1add, 1adf, 1aef, 1aem, 1ajx, 1apb, 1apt, 1b39, 1bil, 1bra, 1bv7, 1c2t, 1cim, 1d3d, 1d4y, 1dih, 1dmp, 1elc, 1epo, 1epp, 1hgh, 1hsl, 1hte, 1hvi, 1hvl, 1hvr, 1lde, 1mnc, 1pgp, 1ppk, 1ppm, 1stp, 1tnk, 1tnl, 1tpp, 1ulb, 1uvs, 2acs, 2er6, 2ypi, 3cpa, 4phv, 5er2, 5tim, 6cpp, 7cpa, 7cpp, 7upj.

LPDB160:

(group1): 1aaq, 1adc, 1aee, 1aek, 1ajv, 1apu, 1avn, 1elc, 1ets, 1ets, 1hhg, 1hhj, 1hih, 1hvi, 1hvl, 1nsc, 1pgp, 1ppl, 1rbp, 1rne, 2cpp, 2ctc, 2er6, 2phh, 4er2, 4er4, 5cna, 5er2, 5hvp, 7hvp, 7upj, 9aat, 9hvp.

(group2): 1abe, 1abf, 1aev, 1bap, 1bra, 1ctt, 1dbb, 1dif, 1dih, 1dog, 1dwc, 1eed, 1epo, 1etr, 1ett, 1hpv, 1mcf, 1mch, 1mcj, 1ppc, 1pso, 1sre, 1tnk, 2acs, 2er7, 2tsc, 2upj, 2ypi, 3tmn, 4dfr, 4hmg, 6abp.

(group3): 1a4w, 1ac8, 1adf, 1aed, 1aeg, 1aen, 1aes, 1ajx, 1apb, 1apt, 1apv, 1dwd, 1ebg, 1hdt, 1hhk, 1hos, 1hps, 1hvk, 1ldm, 1lgr, 1nsd, 1ppk, 1stp, 1tnh, 2ifb, 3ptb, 4phv, 5abp, 5cpp, 5tim, 7cpp, 8abp.

(group4): 1add, 1aeb, 1aeh, 1aeq, 1aet, 1bid, 1bmn, 1dbm, 1dwb, 1ela, 1ent, 1hhi, 1hsg, 1hsl, 1htf, 1hvj, 1hvr, 1lyb, 1mcs, 1nnb, 1pph, 1ppm, 1tni, 1tnj, 1ulb, 1uvs, 1xig, 2xis, 4er1, 4hvp, 6cpp, 7abp.

(group5): 1ac4, 1ae8, 1aef, 1aem, 1aeo, 1aeu, 1apw, 1bmm, 1dbj, 1dbk, 1epp, 1fkf, 1gno, 1hbv, 1hpx, 1hte, 1htg, 1mcb, 1tng, 1tnl, 1tpp, 1vac, 2dbl, 2gbp, 2wea, 2web, 2wec, 3er3, 3er5, 4ts1, 8cpp, 9abp.

Table S1. Data look-up tables for distance dependent pair potentials $W_{pair}(r)$.

Distance	OH	CH	HH	NH	CC	ON	OC	NC	OO	other
1.00										
1.25										
1.50										
1.75	-2.25		0.61							
2.00	-2.27		0.87	2.00						2.00
2.25	-1.40		-0.80	1.56						-0.37
2.50	-0.92	2.00	-1.01	1.60						0.94
2.75	-1.06	0.99	-0.70	0.42		-0.69			-0.17	0.48
3.00	-0.57	0.11	-0.34	0.06		-1.49	2.00		-1.36	0.36
3.25	-0.25	-0.20	-0.15	0.10	2.00	-0.82	0.31	1.93	-0.68	0.07
3.50	-0.13	-0.16	-0.08	0.05	1.80	-0.07	-0.37	1.08	-0.44	0.12
3.75	0.06	-0.06	-0.03	0.03	0.51	0.18	-0.44	0.28	-0.16	0.06
4.00	0.16	0.00	-0.02	0.02	0.07	0.26	-0.33	0.05	0.08	-0.03
4.25	0.11	-0.03	0.02	-0.03	-0.01	0.14	-0.10	0.07	0.18	-0.07
4.50	0.10	-0.04	0.03	-0.01	-0.08	0.09	0.03	0.13	0.21	-0.01
4.75	0.10	-0.01	0.02	0.01	-0.12	0.00	0.05	0.06	0.22	-0.15
5.00	0.11	0.02	0.01	0.06	-0.17	-0.03	0.07	-0.10	-0.16	-0.05
5.25	0.09	0.02	0.02	0.00	-0.14	-0.12	0.08	-0.16	-0.04	-0.03
5.50	0.02	0.00	0.06	-0.04	-0.10	0.10	0.04	-0.18	0.09	0.10
5.75	0.03	-0.01	0.07	-0.05	-0.10	0.19	-0.01	-0.15	0.18	0.02
6.00	0.03	0.00	0.06	-0.05	-0.11	0.22	0.01	-0.14	0.19	-0.02

Table S2. Weights λ_{pair} for the scoring functinon S3w fit to LPDB259.

OH	CH	HH	NH	CC	ON	OC	NC	OO
0.380	3.994	0.000	7.727	0.00	1.428	0.00	0.00	0.456

Table S3. Enrichement Factors and ROC curves.

all poses Step1	re-rank top5 Step1	LPDB DP Step1	Rank Step2	all MW 180-700 <EF>	True EF 320-450 <EF>	Low MW 180-375 <EF>	Low MW 320-375 <EF>	High MW 375-450 <EF>	all MW 180-700 AUC	True AUC 320-450 AUC	Low MW 180-375 AUC	Low MW 320-375 AUC	High MW 375-450 AUC
LIE(Rdie)	LIE(Rdie)	-1.53	LIE(GBMV)	7.1	1.0	0.6	0.9	1.1	0.53	0.54	0.51	0.58	0.53
LIE(Rdie)	LIE(GBMV)	-1.62	LIE(GBMV)	7.3	1.4	1.1	1.6	1.2	0.54	0.55	0.49	0.57	0.60
Rdie	Rdie	-1.21	LIE(GBMV)	5.9	1.2	1.2	1.5	1.0	0.60	0.60	0.60	0.64	0.57
Rdie	LIE(GBMV)	-1.62	LIE(GBMV)	7.2	1.8	1.4	2.1	1.3	0.59	0.60	0.55	0.62	0.62
LIE(Rdie)	LIE(Rdie)	-1.53	S1	11.6	7.9	1.4	2.1	17.0	0.50	0.57	0.44	0.52	0.75
LIE(Rdie)	LIE(GBMV)	-1.62	S1	11.9	8.0	1.9	2.8	16.7	0.52	0.59	0.45	0.53	0.80
Rdie	Rdie	-1.21	S1	10.7	4.6	1.5	1.6	9.4	0.60	0.63	0.52	0.60	0.78
Rdie	LIE(GBMV)	-1.62	S1	11.2	6.8	1.7	1.8	14.4	0.55	0.62	0.49	0.57	0.82
LIE(Rdie)	LIE(Rdie)	-1.53	S2	4.6	5.9	6.2	8.6	3.2	0.66	0.72	0.69	0.78	0.64
LIE(Rdie)	LIE(GBMV)	-1.62	S2	5.1	6.6	6.3	8.8	4.1	0.68	0.75	0.69	0.77	0.73
Rdie	Rdie	-1.21	S2	3.8	3.4	4.8	6.1	0.7	0.64	0.67	0.61	0.66	0.72
Rdie	LIE(GBMV)	-1.62	S2	3.8	4.5	5.4	7.5	1.4	0.66	0.71	0.65	0.72	0.71
LIE(Rdie)	LIE(Rdie)	-1.53	S2h	2.1	1.0	1.4	1.5	0.6	0.62	0.61	0.63	0.65	0.58
LIE(Rdie)	LIE(GBMV)	-1.62	S2h	2.6	1.6	1.8	2.2	1.1	0.64	0.65	0.63	0.66	0.64
Rdie	Rdie	-1.21	S2h	1.4	0.8	0.7	0.9	0.7	0.60	0.59	0.59	0.60	0.60
Rdie	LIE(GBMV)	-1.62	S2h	2.6	1.6	1.8	2.2	1.1	0.64	0.65	0.63	0.66	0.64
LIE(Rdie)	LIE(Rdie)	-1.53	S3	1.2	0.5	0.0	0.0	1.3	0.44	0.43	0.36	0.36	0.59
LIE(Rdie)	LIE(GBMV)	-1.62	S3	1.8	1.0	0.3	0.4	1.9	0.46	0.43	0.39	0.39	0.54
Rdie	Rdie	-1.21	S3	1.0	0.8	0.2	0.3	1.8	0.41	0.40	0.36	0.36	0.48
Rdie	LIE(GBMV)	-1.62	S3	1.8	1.0	0.3	0.4	1.9	0.46	0.43	0.39	0.39	0.54
LIE(Rdie)	LIE(Rdie)	-1.53	S4	1.2	0.8	1.5	1.2	0.2	0.60	0.61	0.64	0.66	0.54
LIE(Rdie)	LIE(GBMV)	-1.62	S4	1.1	0.7	1.3	1.1	0.0	0.56	0.57	0.60	0.62	0.50
Rdie	Rdie	-1.21	S4	0.4	0.2	0.4	0.3	0.3	0.49	0.50	0.52	0.51	0.48
Rdie	LIE(GBMV)	-1.62	S4	1.1	0.7	1.3	1.1	0.0	0.56	0.57	0.60	0.62	0.50

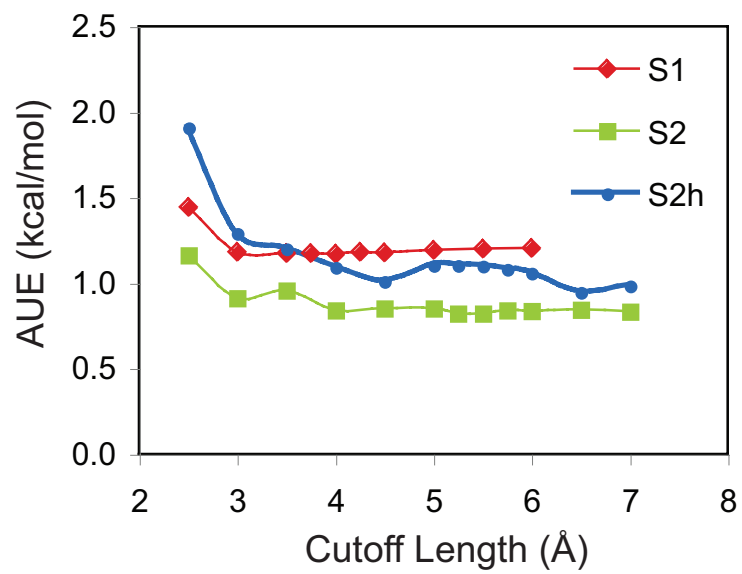Figure S1. Performance of S1, S2, S2h as a function of cutoff distance.

Figure S2. distance dependent pair potentials $W_{pair}(r)$.