Robust Scoring Functions for Protein-Ligand Interactions with Quantum Chemical Charge Models

Jui-Chih Wang,[a] *Jung-Hsin Lin,* [b,c,d,*]*Chung-Ming Chen,*[a] *Alex L. Perryman,*[e] *Arthur J. Olson*[e]

[a] Institute of Biomedical Engineering, National Taiwan University, Taipei, Taiwan. [b] Division of Mechanics, Research Center for Applied Sciences, Academia Sinica, Taiwan. [c] School of Pharmacy, National Taiwan University, Taipei, Taiwan. [d] Institute of Biomedical Sciences, Academia Sinica, Taiwan. [e] Department of Molecular Biology, The Scripps Research Institute, La Jolla, USA.

To whom correspondence should be addressed. J.-H. Lin, Tel: +8862-27823212 ext 886. Fax: +8862-27823060. E-mail: jlin@ntu.edu.tw

Supporting Information

Table S1. The outliers of the RGG model

PDB	$\Delta G_{exp.}$	$\Delta G_{pred.}$	Protein name
1adc	-6.42	-15.94	ALCOHOL DEHYDROGENASE
1apt	-12.82	-8.82	PENICILLOPEPSIN
1bzm	-8.23	-4.72	CARBONIC ANHYDRASE I
1cbx	-8.65	-7.03	CARBOXYPEPTIDASE A
1cil	-12.94	-6.44	CARBONIC ANHYDRASE II
1cim	-12.1	-5.70	CARBONIC ANHYDRASE II
1cin	-11.97	-6.13	CARBONIC ANHYDRASE II
1cnw	-10.6	-3.04	CARBONIC ANHYDRASE II
1cnx	-10.11	-4.05	CARBONIC ANHYDRASE II
1cny	-10.78	-2.41	CARBONIC ANHYDRASE II
1dih	-7.83	-12.49	DIHYDRODIPICOLINATE REDUCTASE
1ebg	-14.76	-9.51	ENOLASE
1hsl	-9.95	-6.34	HISTIDINE BINDING PROTEIN
1lde	-9.41	-5.41	ALCOHOL DEHYDROGENASE
1ldy	-11.06	-4.16	ALCOHOL DEHYDROGENASE
1lyb	-15.57	-9.82	CATHEPSIN D
1okl	-12.39	-5.73	CARBONIC ANHYDRASE II
1pgp	-7.77	-3.62	6-PHOSPHOGLUCONATE DEHYDROGENASE
1sbp	-9.44	-5.78	SULFATE-BINDING PROTEIN
1stp	-18.39	-10.02	STREPTAVIDIN
1tmn	-9.95	-8.37	THERMOLYSIN
2xis	-7.94	-3.48	XYLOSE ISOMERASE
3er3	-9.68	-4.11	ENDOTHIAPEPSIN
3er5	-12.35	-10.16	ENDOTHIAPEPSIN
4dfr	-13.22	-9.67	DIHYDROFOLATE REDUCTASE
5tmn	-10.96	-6.86	THERMOLYSIN
6cpa	-15.71	-8.41	CARBOXYPEPTIDASE A
7cpa	-19.03	-9.21	CARBOXYPEPTIDASE A
9aat	-11.21	-7.08	ASPARTATE AMINOTRANSFERASE

Table S2. The outliers of the RAP model

PDB	$\Delta G_{exp.}$	$\Delta G_{pred.}$	Protein name
1adb	-11.45	-16.60	ALCOHOL DEHYDROGENASE
1adc	-6.42	-16.08	ALCOHOL DEHYDROGENASE
1ajv	-10.59	-14.94	HIV-1 PROTEASE
1apt	-12.82	-9.24	PENICILLOPEPSIN
1bra	-2.49	-6.18	TRYPSIN
1cil	-12.94	-6.98	CARBONIC ANHYDRASE II
1cim	-12.1	-6.34	CARBONIC ANHYDRASE II
1cin	-11.97	-6.73	CARBONIC ANHYDRASE II
1cnw	-10.6	-3.78	CARBONIC ANHYDRASE II
1cnx	-10.11	-4.59	CARBONIC ANHYDRASE II
1cny	-10.78	-3.13	CARBONIC ANHYDRASE II
1dih	-7.83	-12.87	DIHYDRODIPICOLINATE REDUCTASE
1ldy	-11.06	-4.48	ALCOHOL DEHYDROGENASE
1lyb	-15.57	-9.88	CATHEPSIN D
1okl	-12.39	-6.35	CARBONIC ANHYDRASE II
1sre	-5.46	-8.87	STREPTAVIDIN
1stp	-18.39	-9.86	STREPTAVIDIN
1tnk	-2.03	-5.95	TRYPSIN
1tnl	-2.56	-6.79	TRYPSIN
3er3	-9.68	-5.43	ENDOTHIAPEPSIN
3er5	-12.35	-10.65	ENDOTHIAPEPSIN
6cpa	-15.71	-9.45	CARBOXYPEPTIDASE A
7cpa	-19.03	-10.38	CARBOXYPEPTIDASE A

Table S3. The outliers of the RRP model

PDB	$\Delta G_{exp.}$	$\Delta G_{pred.}$	Protein name
1adb	-11.45	-16.75	ALCOHOL DEHYDROGENASE
1adc	-6.42	-16.04	ALCOHOL DEHYDROGENASE
1ajv	-10.59	-14.51	HIV-1 PROTEASE
1bra	-2.49	-6.21	TRYPSIN
1cbx	-8.65	-9.09	CARBOXYPEPTIDASE A
1dih	-7.83	-12.85	DIHYDRODIPICOLINATE REDUCTASE
1ebg	-14.76	-13.20	ENOLASE
1hhj	-9	-13.18	HISTOCOMPATIBILITY ANTIGEN
1lde	-9.41	-7.04	ALCOHOL DEHYDROGENASE
1lyb	-15.57	-9.17	CATHEPSIN D
1stp	-18.39	-9.73	STREPTAVIDIN
1tmn	-9.95	-9.70	THERMOLYSIN
1tni	-2.32	-5.81	TRYPSIN
1tnj	-2.67	-6.21	TRYPSIN
1tnk	-2.03	-5.97	TRYPSIN
1tnl	-2.56	-6.79	TRYPSIN
4er4	-9.27	-14.01	ENDOTHIAPEPSIN
4hvp	-8.33	-12.65	HIV-1 PROTEASE
4tmn	-13.89	-13.43	THERMOLYSIN
6tmn	-6.88	-8.33	THERMOLYSIN

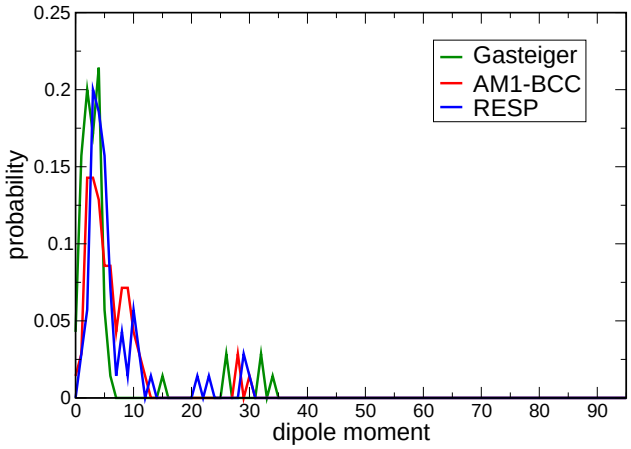

Figure S1. The distributions of dipole moments (in unit Debye) of 70 neutral ligands of 189 dataset (from LPDB). The values of dipole moment were calculated by three kinds of atomic charges.

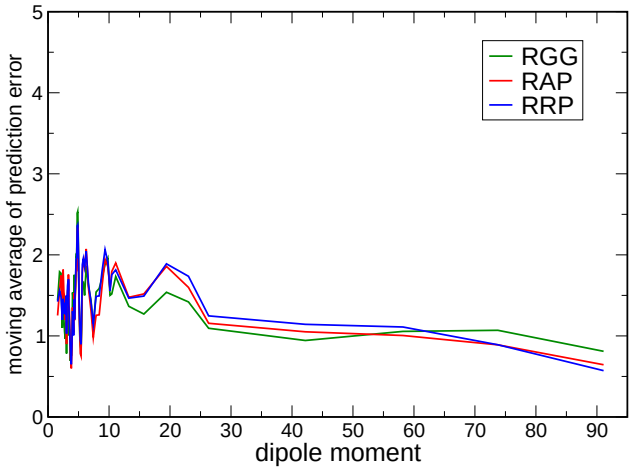

Figure S2. Moving average of prediction error versus dipole moment (Debye) of 70 neutral ligands of 189 dataset (from LPDB). The prediction error was deviation between estimated and experimental binding free energy. The values of dipole moment were calculated by RESP atomic charges.

Table S4. RMS prediction errors of AutoDock4 scoring functions for neutral ligands in the training set

scoring function	70 cases	10 cases[a]
AutoDock4^{RGG}	1.684	1.357
AutoDock4^{RAP}	1.689	1.367
AutoDock4^{RRP}	1.709	1.456
original AutoDock4^{GG}	2.293	1.726

[a]These 10 cases are large dipole moment ligands. (Debye > 12.5)

All values of RMS prediction errors are in kcal/mol.