Supporting Information

Combined Linear Interaction Energy and Alchemical Solvation Free Energy Approach for Protein–binding Affinity Computation

Eko Aditya Rifai,[1] Valerio Ferrario,[2] Jürgen Pleiss,[2] and Daan P. Geerke*,[1]

[1] *AIMMS Division of Molecular and Computational Toxicology, Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, the Netherlands*

[2] *Institute of Biochemistry and Technical Biochemistry, Universität Stuttgart, Allmandring 31, 70569 Stuttgart, Germany*

E–mail: d.p.geerke@vu.nl

Table S1: Weights W_i of individual simulations (starting from binding pose 1, 2 or 3) as used in the ALIE1 model for the ligands considered (listed by ID).

ID	Pose 1	Pose 2	Pose 3	**ID**	Pose 1	Pose 2	Pose 3
1	0.002	0.998	-	**15**	0.990	0.010	0.000
2	0.998	0.002	-	**16**	0.740	0.260	-
3	0.113	0.887	-	**17**	0.974	0.026	-
4	0.003	0.988	0.010	**18**	0.309	0.691	-
5	0.836	0.164	-	**19**	0.063	0.933	0.004
6	0.791	0.209	-	**20**	0.264	0.736	-
7	0.971	0.029	-	**21**	0.016	0.984	-
8	0.005	0.995	-	**22**	0.996	0.004	-
9	0.953	0.047	-	**23**	0.582	0.059	0.359
10	0.963	0.037	-	**24**	0.157	0.832	0.011
11	0.177	0.823	-	**25**	0.594	0.113	0.293
12	0.859	0.051	0.090	**26**	0.757	0.243	-
13	0.093	0.205	0.702	**27**	0.952	0.048	-
14	0.971	0.029	-	**28**	0.131	0.869	-

Table S2: Solvation free energies ΔG_{solv} and associated error estimates as calculated for the considered molecules (listed in Table 1).

No.	ΔG_{solv} (kJ mol^{-1})	Error ($\pm$)	**No.**	ΔG_{solv} (kJ mol^{-1})	Error ($\pm$)
1	–34.76	0.08	**15**	–34.81	0.08
2	–40.09	0.08	**16**	–6.14	0.07
3	–15.65	0.08	**17**	–28.15	0.08
4	–37.39	0.07	**18**	–45.02	0.08
5	–47.71	0.08	**19**	–32.16	0.09
6	–45.80	0.08	**20**	–35.17	0.08
7	–15.57	0.08	**21**	–16.65	0.08
8	–30.00	0.08	**22**	–44.71	0.09
9	–25.21	0.08	**23**	–50.13	0.11
10	–34.77	0.07	**24**	–44.33	0.09
11	–50.95	0.08	**25**	–29.32	0.08
12	–24.64	0.09	**26**	–34.60	0.08
13	–39.23	0.10	**27**	–18.09	0.07
14	–16.02	0.08	**28**	–16.88	0.07

Table S3: $\left\langle V^{ele}_{lig-surr}\right\rangle_{bound}$ and $\left\langle V^{ele}_{lig-surr}\right\rangle_{off}$ values (in kJ mol^{-1}) as calculated for the considered molecules (listed in Table 1) and as used in the ALIE2 model

No.	$\left\langle V^{ele}_{lig-surr}\right\rangle_{bound}$	$\left\langle V^{ele}_{lig-surr}\right\rangle_{off}$	**No.**	$\left\langle V^{ele}_{lig-surr}\right\rangle_{bound}$	$\left\langle V^{ele}_{lig-surr}\right\rangle_{off}$
1	–60.36	17.32	**15**	–59.41	–3.62
2	–57.41	–1.17	**16**	–6.93	3.56
3	–15.04	13.13	**17**	–36.02	1.48
4	–60.69	8.81	**18**	–42.90	–14.73
5	–51.78	1.54	**19**	–32.40	–12.03
6	–40.51	0.42	**20**	–23.13	13.81
7	–24.23	12.90	**21**	–6.09	15.71
8	–53.35	–8.86	**22**	–58.96	29.34
9	–42.68	0.52	**23**	–40.60	16.14
10	–38.71	1.77	**24**	–38.29	8.82
11	–52.66	10.33	**25**	–32.90	9.44
12	–30.91	–3.14	**26**	–38.53	–3.18
13	–30.40	9.81	**27**	–26.88	11.01
14	–5.71	9.10	**28**	–27.57	–8.54

Table S4: Model parameters for the ALIE1 models of CYP2A6 including some adaptations (i.e., calibrated either without compound **22**, without compound **16**, or with an offset γ–parameter, in kJ mol^{-1}), and for two adapted models with a correction for the ligand's SASA and number of rotatable bonds (with $T\Delta S$ in Equation 13 set to 20 kJ mol^{-1} or 6 kJ mol^{-1}, respectively, as indicated by the subscript for SASA), together with their respective root–mean–square errors (RMSE, with respect to experiment and in kJ mol^{-1}), standard deviation errors in the prediction from leave–one–out cross–validation (SDEP, in kJ mol^{-1}) and correlation metrics for the set of compounds considered in this work.

	ALIE1	w/o **22**	w/o **16**	γ	$SASA_{20}$/rot	$SASA_6$/rot
No. of comp.	28	27	27	28	28	28
α	0.34	0.30	0.34	0.15	0.26	0.29
β	0.66	0.79	0.65	0.63	0.66	0.66
γ	–	–	–	–23.42	–	–
RMSE	8.39	7.71	8.53	8.03	8.38	8.62
SDEP	9.29	8.31	9.53	9.45	9.26	9.54
r^2	0.59	0.64	0.50	0.61	0.56	0.54
q^2	0.48	0.58	0.35	0.46	0.44	0.41
Pearson's r	0.77	0.80	0.71	0.78	0.75	0.74
Spearman's ρ	0.76	0.76	0.73	0.78	0.74	0.74

Table S5: Experimental estimates for the CYP2E1 binding free energy ($\Delta G_{bind,exp}$ in kJ mol^{-1}) of the 28 compounds considered in the current work.

No.	$\Delta G_{bind,exp}$	**No.**	$\Delta G_{bind,exp}$
1	–27.04	**15**	–22.63
2	–23.40	**16**	–28.86
3	–31.41	**17**	–28.88
4	–28.66	**18**	–30.42
5	–25.39	**19**	–35.82
6	–32.81	**20**	–26.12
7	–21.50	**21**	–21.50
8	–35.10	**22**	–32.68
9	–42.59	**23**	–37.63
10	–24.54	**24**	–25.32
11	–28.17	**25**	–27.68
12	–35.10	**26**	–27.43
13	–38.56	**27**	–31.34
14	–31.71	**28**	–32.74

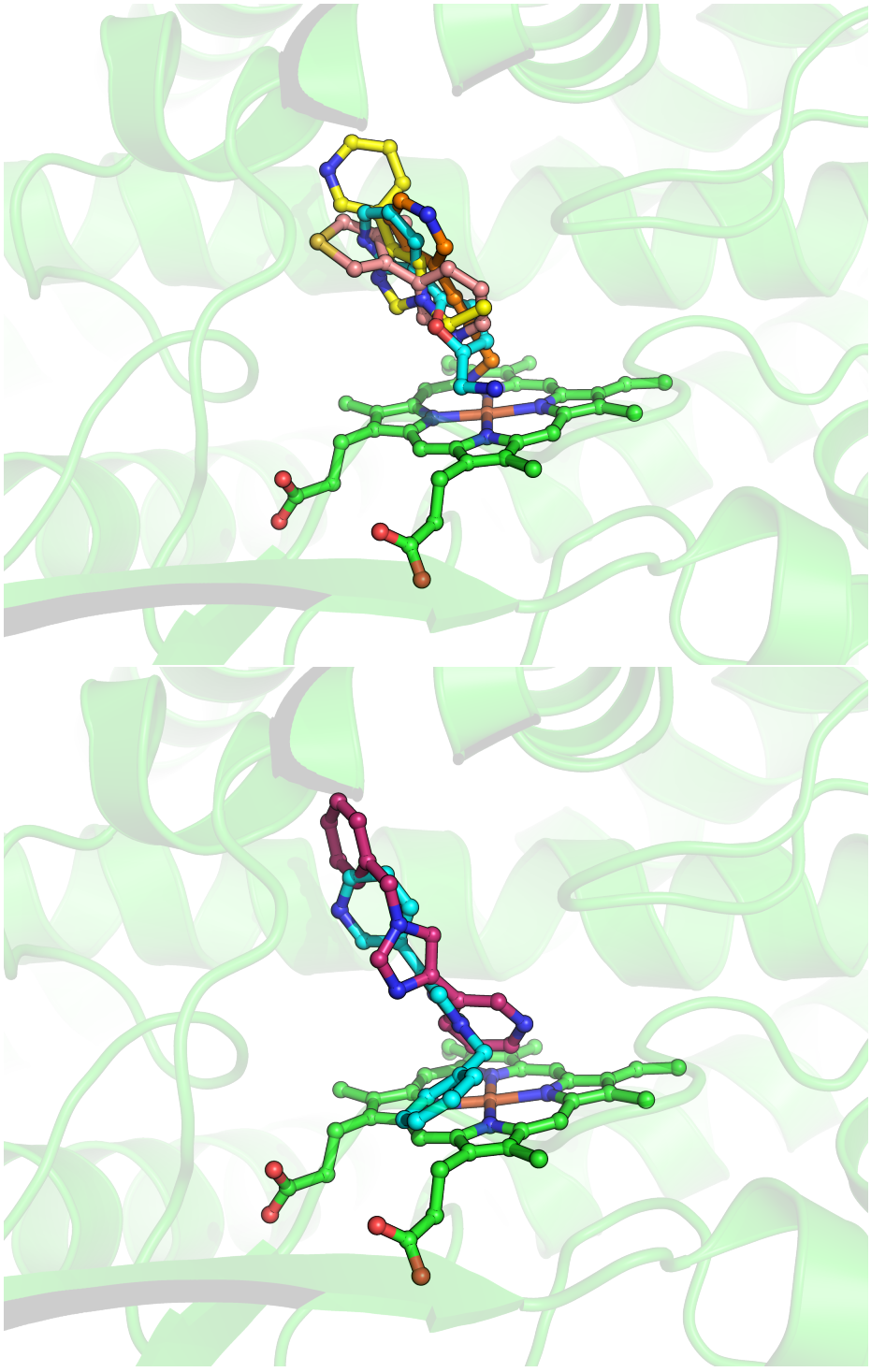

Figure S1: Upper panel: comparison of selected ligand-binding poses (for compounds **2** [cyan], **4** [orange], **14** [pink] and **24** [yellow]) that served as starting pose in MD simulations with $W_i > 0.8$ as used in the ALIE1 model. Lower panel: comparison of starting poses of individual simulations used in the ALIE1 model for compound **23**, i.e., poses 1 [cyan] and 3 [magenta] (with W_i of 0.58 and 0.36, respectively, Table S1).

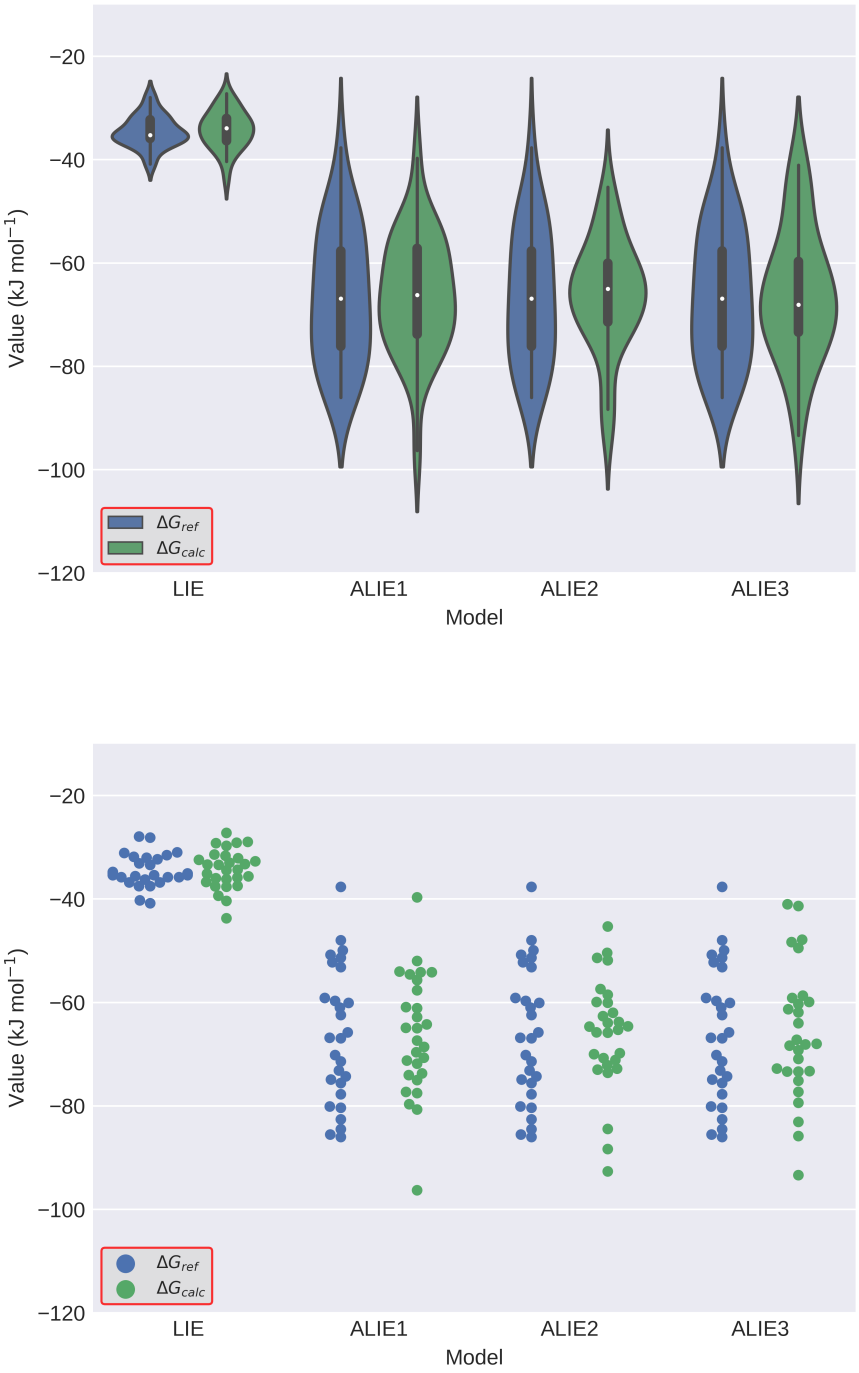

Figure S2: Comparison of distributions of the calculated (calc) and experimental (ref) ΔG_{bind} values for the CYP2A6 LIE model, and of the calculated (calc) and reference (ref) ΔG_{prot} values for the CYP2A6 ALIE models.

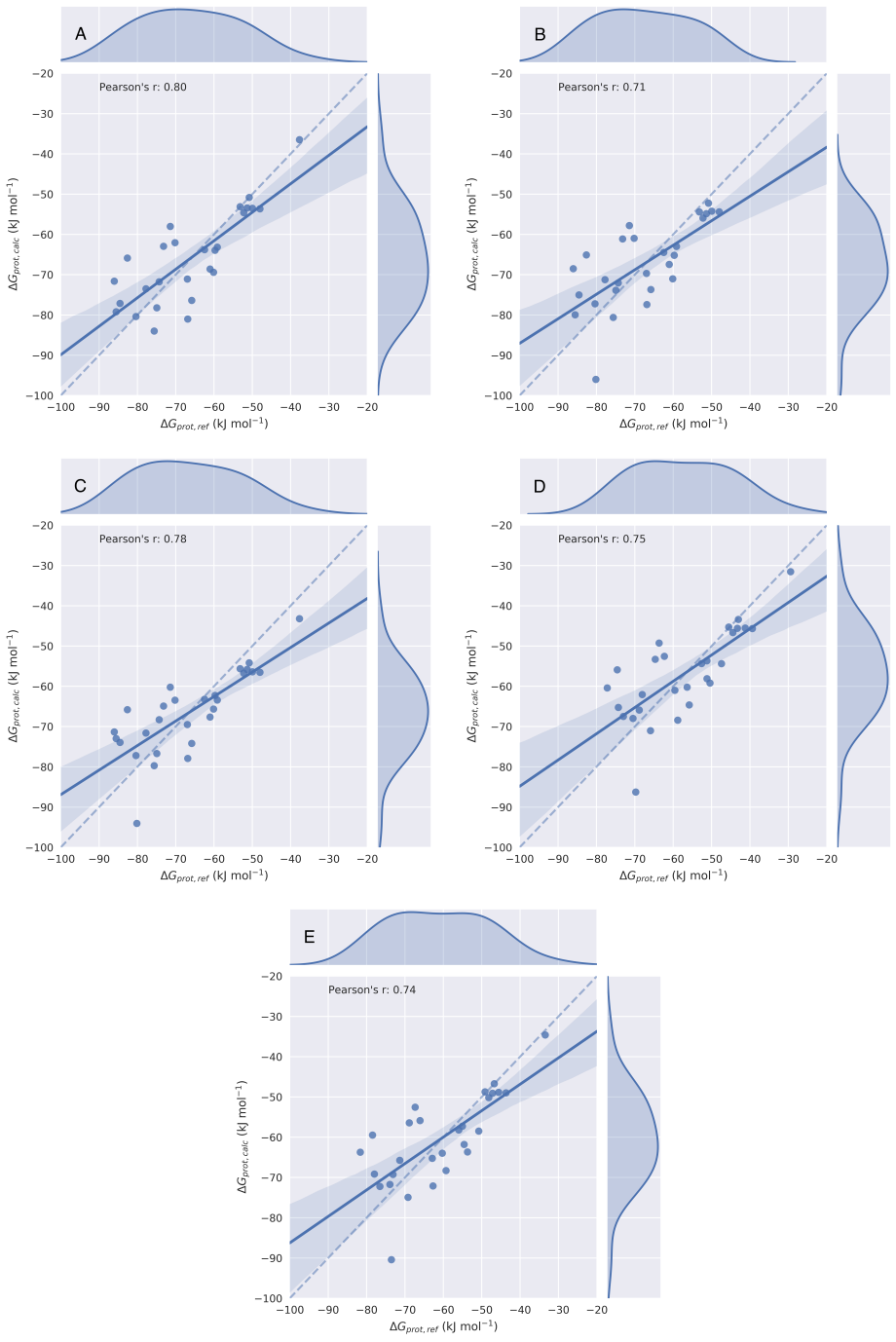

Figure S3: Scatter and kernel density plots of the ALIE1 models for CYP2A6 binding free energy calculation, calibrated either without compound **22** (A), without compound **16** (B), with all 28 compounds and an offset γ–parameter (C), and with a correction for the ligand's SASA and number of rotatable bonds (with $T\Delta S$ in Equation 13 set to 20 (D) or 6 (E) kJ mol^{-1}), illustrating obtained correlations between calculated free energies $\Delta G_{prot,calc}$ and the respective reference values. Dashed lines indicate ideal correlation, solid lines indicate actually obtained correlations, and shaded areas indicate 95% confidence intervals.

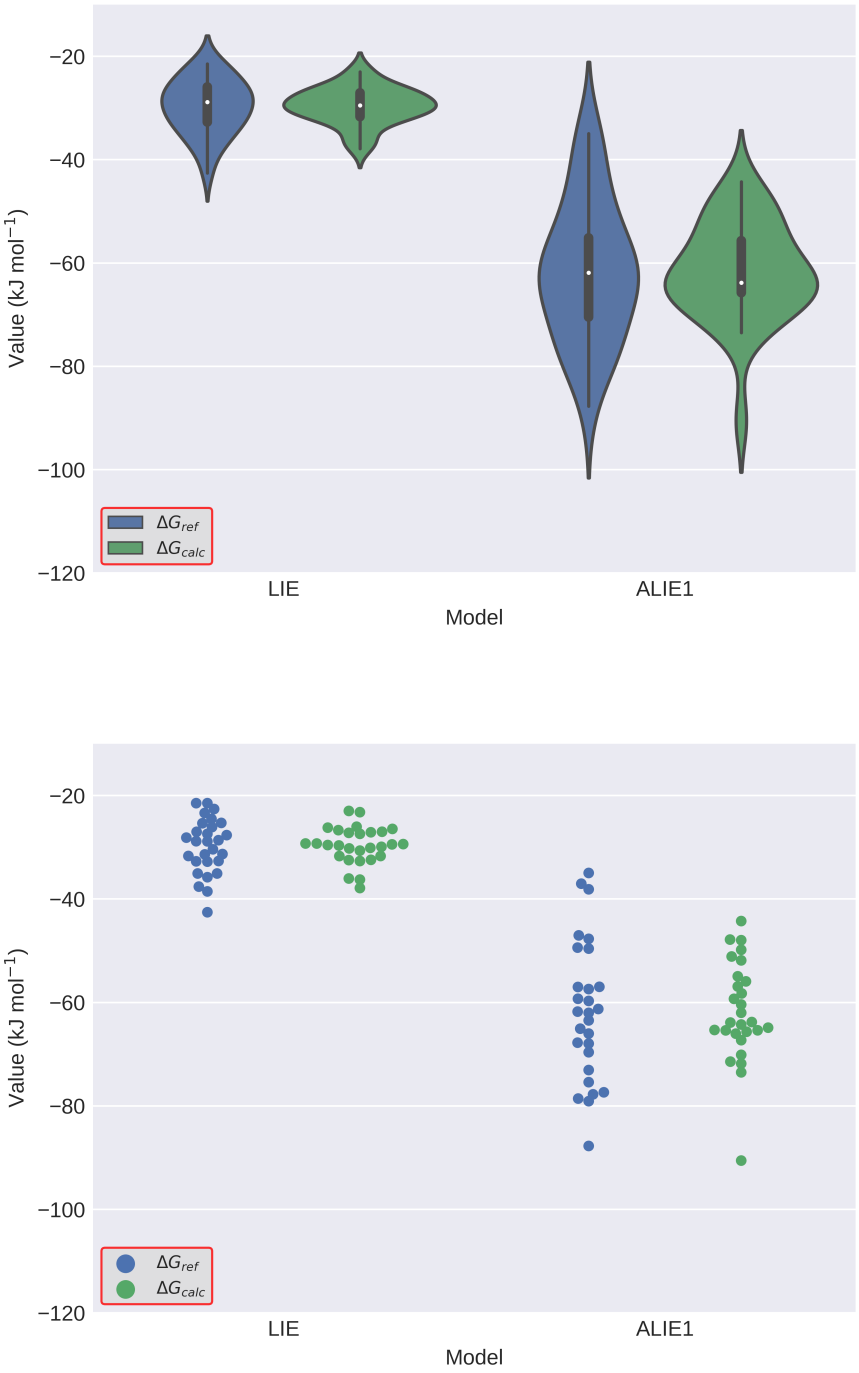

Figure S4: Comparison of distributions of the calculated (calc) and experimental (ref) ΔG_{bind} values for the CYP2E1 LIE model, and of the calculated (calc) and reference (ref) ΔG_{prot} values for the CYP2E1 ALIE1 model.