Supporting information for: Uncertainty Quantification in Alchemical Free Energy Methods

Agastya P. Bhati,[†] Shunzhou Wan,[†] Yuan Hu,[‡,¶] Brad Sherborne,[‡] and Peter V. Coveney[*,†]

[†]*Centre for Computational Science, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom*

[‡]*Modeling and Informatics, Merck & Co., Inc., 2000, Galloping Hill Road, Kenilworth, NJ 07033, USA*

[¶]*Current address: Modeling & Informatics, Discovery, Alkermes Inc., 852 Winter Street, Waltham, MA, 02451, USA*

E-mail: p.v.coveney@ucl.ac.uk

Phone: +44 (0)20 7679 4802

In the main text, we have described different approaches based on ensemble simulation to calculate absolute as well as relative free energies using "exact" free energy methods. It has been demonstrated that running a single replica may have several issues and certainly does not provide a handle with which to quantify the associated uncertainties in the predicted free energies; this is true irrespective of simulation length. Here, we provide additional details behind our results reported in the main text. Table S1 provides the free energy predictions for all complexes studied from each replica as well as their ensemble averages calculated

APB and SW contributed equally to the results

using the TIES analysis. Figure S1 shows the variation of $\Delta G^{protein}_{vdw+elec}$, the most fluctuating component of absolute free energy, for simulations up to 10 ns for all complexes studied. Table S2 shows the energy decomposition of absolute free energies for all complexes. Table S3 compares the computed values of $\Delta G^{protein}_{vdw+elec}$, the most fluctuating component of absolute free energy, with and without MBAR analysis which demonstrate that MBAR has no effect on the accuracy of the results.

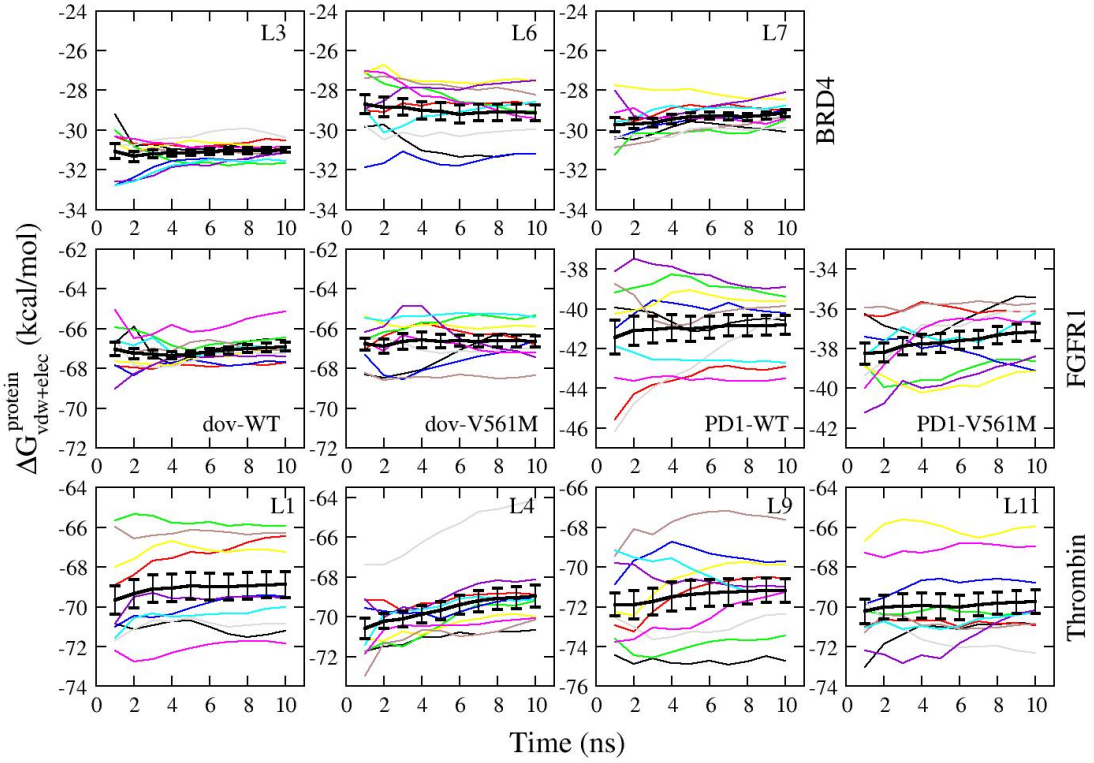

Figure S1: Convergence of the absolute binding free energy calculations for all molecular systems investigated in this study. $\Delta G^{protein}_{vdw+elec}$ is used here to display the convergence as it is the component with largest variance. The coloured lines represent individual replicas while the black lines denotes results from TIES analysis with associated error bars. While most of the TIES results have converged within 4 ns, some of the individual replicas have not within 10 ns.

Table S1: Free energy predictions for all complexes studied using 5 replicas for the four schemes (I to IV†). The largest values among all replicas are highlighted in **bold** and the smallest ones in *italics*. All values are in kcal/mol.

System	FE-type	Scheme	rep1	rep2	rep3	rep4	rep5	TIES-analysis
V561M mutant (forward) with PD173074	ΔG^{com}_{alch}	I	2.94	**3.12**	2.79	2.91	*2.41*	2.84(0.16)
		II	3.00	**3.16**	2.93	2.67	*2.20*	2.79(0.15)
		III	**2.40**	1.85	2.04	2.06	*1.78*	2.03(0.06)
		IV	**2.30**	1.87	2.00	2.11	*1.79*	2.01(0.05)
	ΔG^{pro}_{alch}	I	-0.75	**-0.45**	-0.80	-0.79	*-0.83*	-0.72(0.09)
		II	-0.71	**-0.60**	-0.83	*-0.92*	-0.72	-0.75(0.08)
		III	-1.25	*-1.30*	-1.26	-1.12	**-1.06**	-1.20(0.05)
		IV	*-1.31*	-1.26	-1.18	**-1.03**	-1.12	-1.18(0.04)
V561M mutant (reverse) with PD173074	ΔG^{com}_{alch}	I	1.86	**1.89**	1.69	*1.67*	1.86	1.79(0.10)
		II	1.77	1.85	*1.48*	1.58	**2.10**	1.76(0.09)
		III	*1.99*	2.39	2.09	**2.52**	2.32	2.26(0.07)
		IV	*2.00*	2.36	2.05	**2.51**	2.33	2.25(0.07)
	ΔG^{pro}_{alch}	I	**-0.53**	-0.89	-0.92	*-1.08*	-0.90	-0.86(0.08)
		II	**-0.66**	-0.79	*-1.07*	-1.04	-0.87	-0.89(0.08)
		III	-1.02	-1.32	**-0.92**	-1.18	*-1.35*	-1.16(0.07)
		IV	-1.06	-1.24	**-1.00**	-1.24	*-1.31*	-1.17(0.06)
V561M mutant (forward) with TKI258	ΔG^{com}_{alch}	III	*-1.73*	-1.52	**-1.05**	-1.22	-1.23	-1.35(0.07)
		IV	*-1.76*	-1.55	**-1.00**	-1.27	-1.26	-1.37(0.07)
L1-L9 with thrombin	ΔG^{com}_{alch}	III	**1.29**	*0.54*	0.86	1.06	1.04	0.96(0.09)
		IV	**1.22**	*0.54*	0.84	1.16	1.10	0.97(0.08)
	ΔG^{lig}_{alch}	III	1.61	**1.68**	1.60	*1.58*	1.66	1.63(0.04)
		IV	**1.68**	1.66	1.63	*1.60*	1.65	1.64(0.03)
L4-L11 with thrombin	ΔG^{com}_{alch}	III	-2.00	**-1.58**	-1.97	*-2.83*	-2.77	-2.23(0.13)
		IV	-1.95	**-1.59**	-2.00	*-2.86*	-2.80	-2.24(0.12)
	ΔG^{lig}_{alch}	III	*-1.29*	-1.05	-1.22	-1.26	**-1.01**	-1.17(0.04)
		IV	*-1.35*	-1.10	-1.21	-1.26	**-1.04**	-1.19(0.03)
L3-L6 with BRD4	ΔG^{com}_{alch}	III	*-4.73*	-4.17	**-4.14**	-4.34	-4.32	-4.34(0.08)
		IV	*-4.86*	-4.15	**-4.11**	-4.33	-4.30	-4.35(0.07)
	ΔG^{lig}_{alch}	III	**-5.18**	-5.63	-5.29	-5.58	*-5.74*	-5.48(0.06)
		IV	**-5.24**	-5.59	-5.34	-5.55	*-5.72*	-5.49(0.04)
L3-L7 with BRD4	ΔG^{com}_{alch}	III	*4.79*	4.89	5.00	**5.22**	4.90	4.96(0.08)
		IV	*4.72*	4.86	4.97	**5.21**	4.87	4.92(0.07)
	ΔG^{lig}_{alch}	III	5.27	5.19	*5.15*	**5.40**	5.17	5.23(0.06)
		IV	5.28	5.22	*5.03*	**5.39**	5.06	5.19(0.05)

† In scheme IV, the samples from states which are electrostatically fully decoupled from the state of interest are excluded from MBAR analysis. This is because the energies of such samples at the state of interest may approach infinitely high values due to overlapping atoms by virtue of the non-softcore electrostatic potential used in these simulations.

Table S2: Free energy contributions (kcal/mol) from the 5 steps (see Section 2.2.2 in the main text) in the thermodynamic cycle for the absolute binding free energy calculation. The errors shown in brackets are derived from the bootstrapped standard error.

Lig	Protein	$\Delta G^{lig}_{vdw+elec}$	ΔG^{lig}_{restr}	$\Delta G^{protein}_{vdw+elec}$ (A)	$\Delta G^{protein}_{restr}$ (B)	A + B	Finite size correction	ΔG_{alch}	ΔG_{exp}
dov	wt	-38.81(0.02)	-13.73	-66.91(0.23)	-3.39(0.01)	-70.30(0.23)	10.35	-7.41(0.23)	-9.0(0.68)
dov	v561m	-38.81(0.02)	-13.73	-66.63(0.28)	-3.42(0.01)	-70.06(0.28)	10.35	-7.17(0.28)	-9.6(0.46)
pd1	wt	-16.46(0.06)	-13.39	-40.79(0.67)	-2.73(0.01)	-43.52(0.66)	-	-13.67(0.66)	-11.4(0.09)
pd1	v561m	-16.46(0.06)	-13.39	-37.16(0.44)	-2.72(0.01)	-39.88(0.44)	-	-10.03(0.44)	-8.7(0.10)
l3	brd4	-10.01(0.02)	-12.83	-31.01(0.12)	-3.31(0.01)	-34.32(0.12)	-	-11.48(0.12)	-9.3(0.05)
l6	brd4	-11.46(0.02)	-12.85	-29.14(0.39)	-4.76(0.23)	-33.90(0.23)	-	-9.59(0.23)	-7.7(0.01)
l7	brd4	-12.48(0.02)	-12.37	-29.15(0.19)	-3.69(0.10)	-32.85(0.16)	-	-8.00(0.16)	-8.0(0.10)
l1	thrombin	-40.61(0.17)	-12.69	-68.88(0.67)	-2.32(0.02)	-71.21(0.67)	8.18	-9.73(0.69)	-8.46
l4	thrombin	-41.42(0.16)	-12.68	-68.94(0.56)	-2.36(0.02)	-71.30(0.54)	7.30	-9.90(0.56)	-7.48
l9	thrombin	-41.65(0.10)	-12.67	-71.17(0.60)	-2.17(0.01)	-73.35(0.60)	7.10	-11.93(0.61)	-8.89
l11	thrombin	-40.19(0.07)	-12.68	-69.74(0.59)	-2.27(0.01)	-72.01(0.58)	6.58	-12.56(0.58)	-8.56

Table S3: Comparison of $\Delta G^{protein}_{vdw+elec}$ with and without MBAR analysis in the absolute binding free energy calculations. Employment of MBAR in TIES calculation has no effect on the calculated free energy changes and hence is not considered in further analyses of absolute binding free energies.

Lig	Protein	$\Delta G^{protein}_{vdw+elec}$	
		without MBAR	with MBAR
dov	wt	-66.91(0.23)	-66.77(0.23)
dov	v561m	-66.63 (0.28)	-66.48 (0.29)
pd1	wt	-40.79 (0.67)	-40.76 (0.70)
pd1	v561m	-37.16 (0.44)	-37.24 (0.42)
l3	brd4	-31.01 (0.12)	-30.78 (0.11)
l6	brd4	-29.14 (0.39)	-29.24 (0.37)
l7	brd4	-29.15 (0.19)	-29.16 (0.20)
l1	thrombin	-68.88 (0.67)	-68.97 (0.68)
l4	thrombin	-68.94 (0.56)	-68.95 (0.57)
l9	thrombin	-71.17 (0.60)	-71.23 (0.58)
l11	thrombin	-69.74 (0.59)	-69.79 (0.60)

It should be noted in Table S3 that the uncertainty in the MBAR results is slightly larger in some cases. This may be because MBAR requires evaluation of the potential energies and energy derivatives for all samples from the single precision trajectory files. The values thus computed are less precise than the double precision ones obtained directly from the simulation which are used for the calculation of ΔG values without MBAR.