Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations

Supporting Information

Matteo Aldeghi[1], Alexander Heifetz[2], Michael J. Bodkin[2], Stefan Knapp[3,4,5] and Philip C. Biggin[1]*

[1]Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK.

[2]Evotec (U.K.) Ltd., 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire OX14 4RZ, U.K.

[3]Structural Genomics Consortium, Nuffield Department of Clinical Medicine, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK.

[4]Target Discovery Institute, Nuffield Department of Clinical Medicine, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.

[5]Institute for Pharmaceutical Chemistry, Goethe University Frankfurt, 60438 Frankfurt, Germany.

*To whom correspondence should be addressed.

Email: philip.biggin@bioch.ox.ac.uk

Tel. +44 1865 613305

Fax. +44 1865 613238

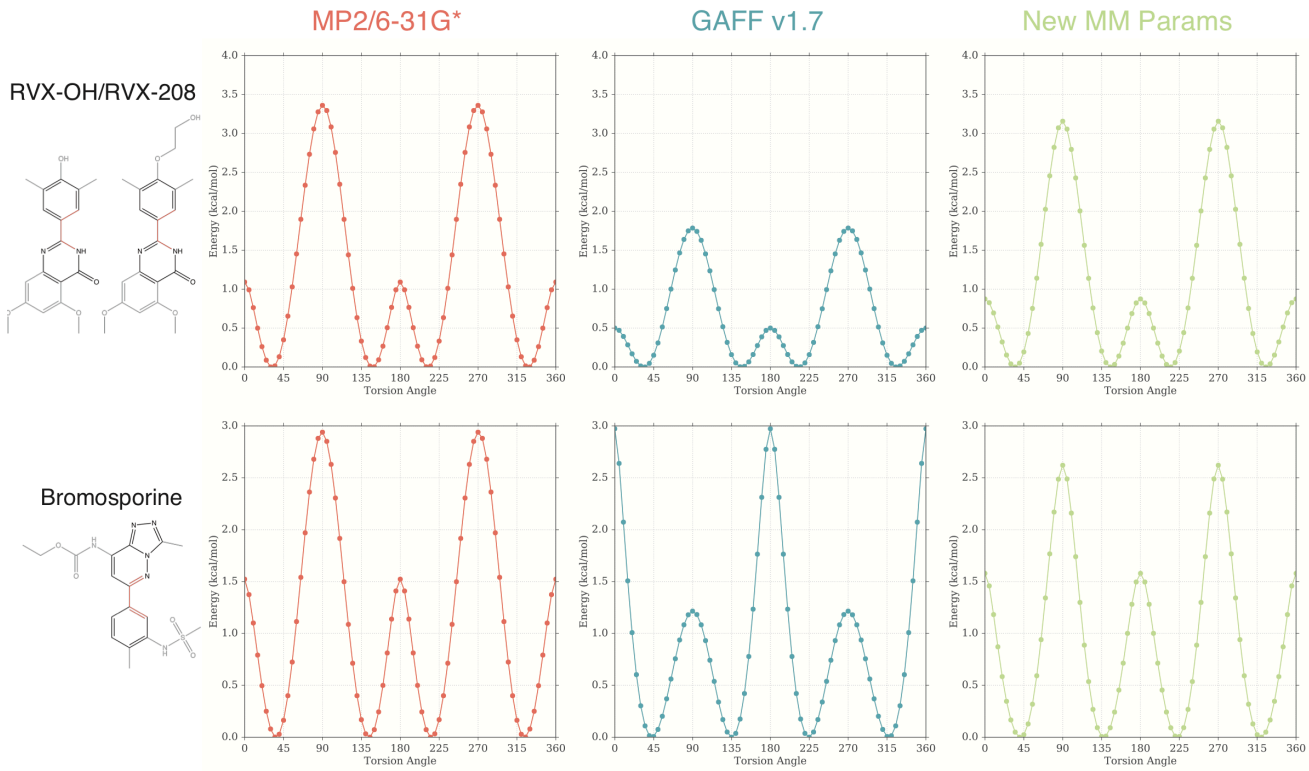
Table S1

Information on the protein models used in the calculations.

Protein	Family	PDB ID	Resolution (Å)	R-Value Free	R-Value Work
CECR2	I	3NXB	1.83	0.214	0.176
FALZ	I	3UV2	1.58	0.209	0.162
PCAF	I	3GG3	2.25	0.245	0.207
BRD2(1)	II	4ALG	1.60	0.202	0.176
BRD2(2)	II	4MR5	1.63	0.170	0.136
BRD3(1)	II	3S91	2.06	0.233	0.195
BRD3(2)	II	3S92	1.36	0.186	0.131
BRD4(1)	II	2OSS	1.35	0.174	0.149
BRD4(2)	II	2OUO	1.89	0.216	0.182
BRDT(1)	II	4KCX	2.00	0.253	0.209
CREBBP	III	4NYX	1.10	0.135	0.119
EP300	III	3I3J	2.33	0.275	0.229
BRD1	IV	5AME	1.58	0.204	0.181
BRD9	IV	4XY8	1.70	0.186	0.148
BRPF1B	IV	4LC2	1.65	0.220	0.174
BAZ2A	V	4QBM	1.65	0.205	0.185
TIF1	V	4YBM	1.46	0.196	0.163
TAF1(1)	VII	3UV5	2.03	0.230	0.190
TAF1(2)	VII	3UV4	1.89	0.225	0.191
TAF1L(2)	VII	3HMH	2.05	0.252	0.181
PB1(5)	VIII	3MB4	1.66	0.219	0.197
SMARCA4	VIII	2GRC	1.50	0.264	0.242

Figure S1

Torsion scans and derived MM parameters for the biaryl dihedral angles in the ligands. The fragment used for the scan and parameter fitting is highlighted in black in the chemical structure, while the rest of the molecule is in gray; the torsion angles parameterized are highlighted in red. The parameter search was performed with a genetic algorithm in which the fitness function was the root mean square deviation between the QM and MM relative energies, using the python script available at http://www.ub.edu/cbdd/?q=content/small-molecule-dihedrals-parametrization.



RVX-OH/RVX-208

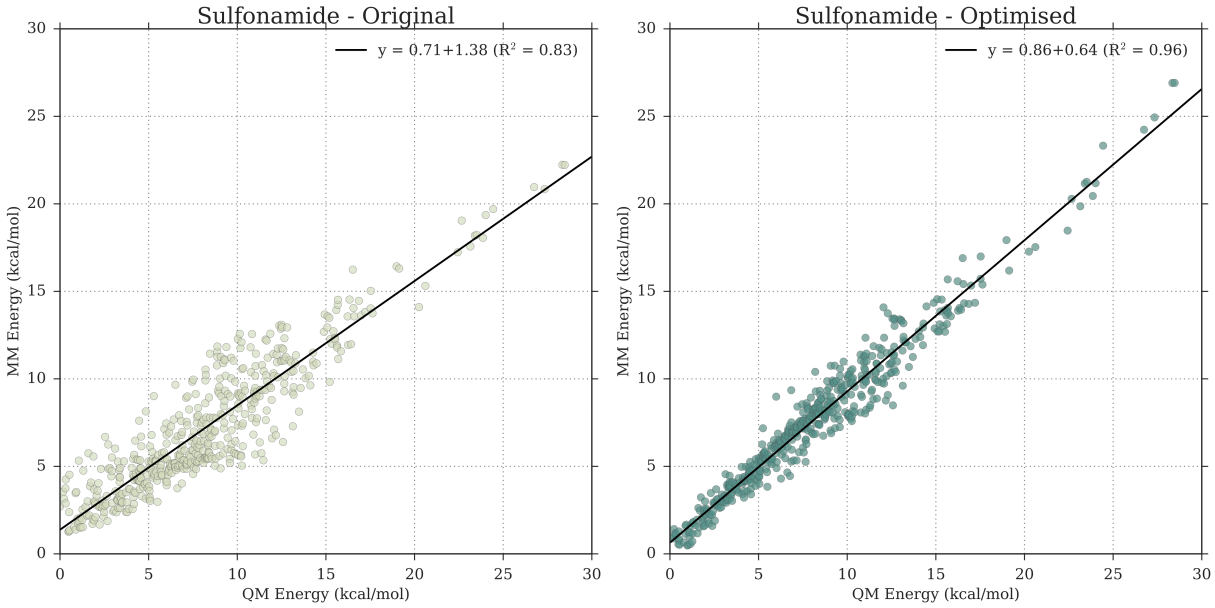
Dihedral	IDIVF	PK (kcal/mol)	PHASE (°)	PN
ca-ca-cc-n	4	3.30	180.0	2.0
ca-ca-cc-n	4	0.40	0.0	4.0
ca-ca-cc-nd	4	3.30	180.0	2.0
ca-ca-cc-nd	4	0.40	0.0	4.0

Bromosporine

Dihedral	IDIVF	PK (kcal/mol)	PHASE (°)	PN
ca-ca-cc-cc	4	3.90	180.0	2.0
ca-ca-cc-cc	4	0.30	180.0	6.0
ca-ca-cc-nd	4	3.90	180.0	2.0
ca-ca-cc-nd	4	0.30	180.0	6.0

Figure S2

New benzensulfonamide torsional parameters for bromosporine. Shown is the fit between MM and QM relative energies before and after the parameterization procedure. The torsions that have been reparameterized are highlighted in red within the bromosporine structure in the table.



Bromosporine (benzensulfonamide)

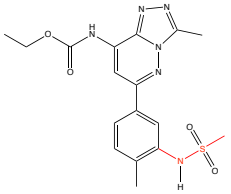
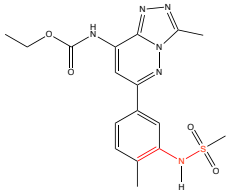
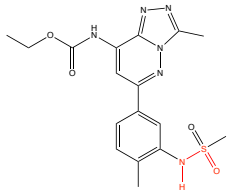
	Dihedral	IDIVF	PK (kcal/mol)	PHASE (°)	PN
	c3-s6-nh-ca	1	0.1126	115.8277	4.0
	c3-s6-nh-ca	1	0.5594	15.0293	3.0
	c3-s6-nh-ca	1	1.5296	5.9420	2.0
	c3-s6-nh-ca	1	-1.1272	-20.9470	1.0
	ca-ca-nh-s6	1	0.0772	237.9864	4.0
	ca-ca-nh-s6	1	-1.6048	207.5177	3.0
	ca-ca-nh-s6	1	-1.1204	-11.8563	2.0
	ca-ca-nh-s6	1	2.7114	-44.7267	1.0
	hn-nh-s6-o	1	-1.8576	28.7191	4.0
	hn-nh-s6-o	1	0.1574	62.1178	3.0
	hn-nh-s6-o	1	1.0255	175.3778	2.0
	hn-nh-s6-o	1	1.5400	4.1667	1.0

Figure S3

Scatter plots of experimental binding free energies versus predicted free energies using a machine learning scoring function (CSM-Lig)[1]. (a) Predicted affinities for RVX-208 and RVX-OH. (b) Predicted affinities for bromosporine. The shaded gray areas indicate where the 1 and 2 kcal/mol error boundaries lie.

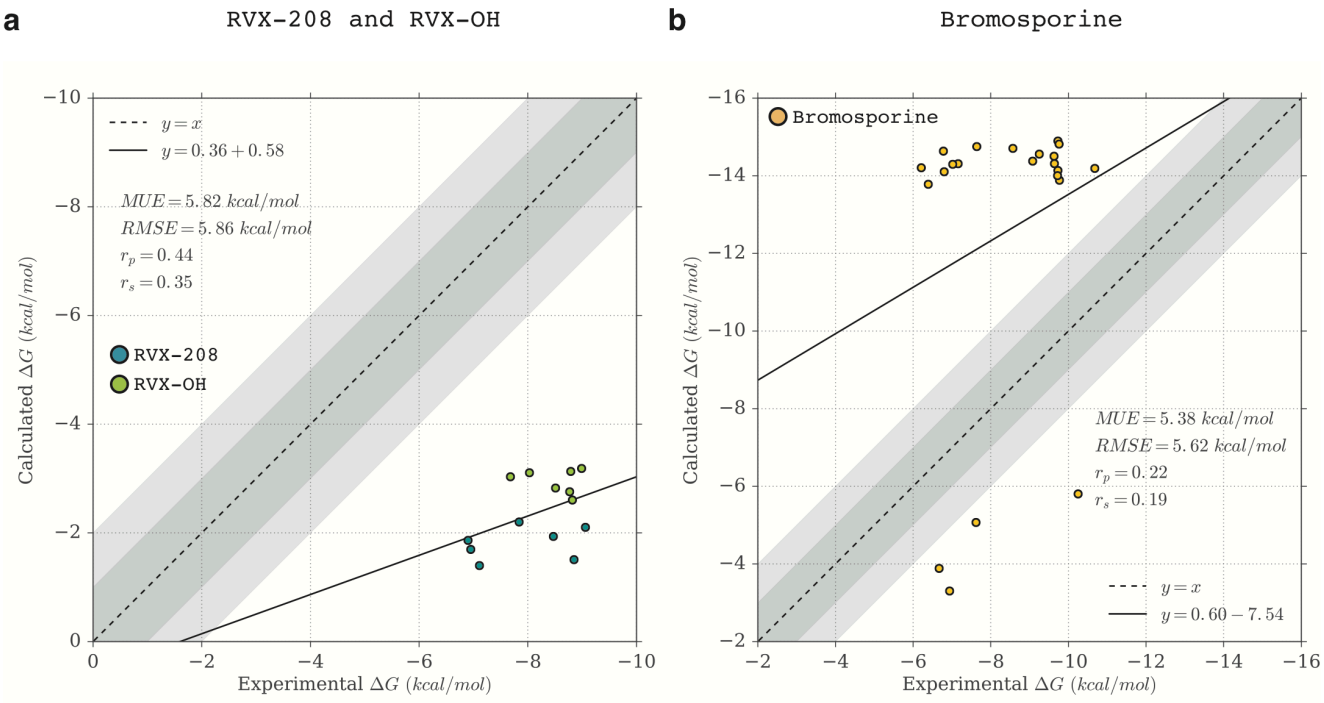


Table S2

Full breakdown of free energy components for the calculations involving the ligands RVX-OH and RVX-208. All free energies are in kcal/mol, and all uncertainties are one standard deviation. "Pose" refers to the ligand orientation evaluated in the calculation for which the data are shown on the same row.

Protein	Ligand	Pose	$\Delta G^{solv}_{elec+vdw}$	ΔG^{solv}_{restr}	$\Delta G^{prot}_{elec+vdw+restr}$	$\Delta\Delta G_{EXP-LR}$
BRD2(1)	RVX-OH	H	82.189 ± 0.025	6.901	-98.523 ± 0.110	-0.403
BRD2(2)	RVX-OH	H	82.189 ± 0.025	7.340	-96.732 ± 0.118	-0.361
BRD3(1)	RVX-OH	H	82.189 ± 0.025	7.205	-95.696 ± 0.114	-0.383
BRD3(2)	RVX-OH	H	82.189 ± 0.025	7.344	-95.850 ± 0.170	-0.309
BRD4(1)	RVX-OH	H	82.189 ± 0.025	6.922	-98.510 ± 0.077	-0.501
BRD4(2)	RVX-OH	H	82.189 ± 0.025	7.254	-94.052 ± 0.104	-0.427
BRDT(1)	RVX-OH	H	82.189 ± 0.025	7.247	-96.707 ± 0.119	-0.408
BRD2(1)	RVX-OH	V	82.189 ± 0.025	6.587	-94.963 ± 0.097	-0.332
BRD2(2)	RVX-OH	V	82.189 ± 0.025	6.569	-96.417 ± 0.056	-0.251
BRD3(1)	RVX-OH	V	82.189 ± 0.025	6.596	-94.386 ± 0.101	-0.118
BRD3(2)	RVX-OH	V	82.189 ± 0.025	6.522	-95.695 ± 0.090	-0.384
BRD4(1)	RVX-OH	V	82.189 ± 0.025	6.573	-95.019 ± 0.101	-0.342
BRD4(2)	RVX-OH	V	82.189 ± 0.025	6.585	-96.869 ± 0.078	-0.405
BRDT(1)	RVX-OH	V	82.189 ± 0.025	6.610	-94.821 ± 0.079	-0.324
BRD2(1)	RVX-208	V	89.772 ± 0.023	6.521	-103.147 ± 0.070	-0.385
BRD2(2)	RVX-208	V	89.772 ± 0.023	6.584	-103.502 ± 0.078	-0.380
BRD3(1)	RVX-208	V	89.772 ± 0.023	6.573	-101.583 ± 0.132	-0.288
BRD3(2)	RVX-208	V	89.772 ± 0.023	6.562	-104.737 ± 0.126	-0.312
BRD4(1)	RVX-208	V	89.772 ± 0.023	6.546	-102.727 ± 0.120	-0.387
BRD4(2)	RVX-208	V	89.772 ± 0.023	6.575	-105.663 ± 0.087	-0.365
BRDT(1)	RVX-208	V	89.772 ± 0.023	6.595	-102.617 ± 0.080	-0.228

Table S3

Full breakdown of free energy components for the ABFE calculations involving bromosporine. All free energies are in kcal/mol, and all uncertainties are one standard deviation.

Protein	Family	$\Delta G^{solv}_{elec+vdw}$	ΔG^{solv}_{restr}	$\Delta G^{prot}_{elec+vdw+restr}$	$\Delta\Delta G_{EXP-LR}$
CECR2	I	167.545 ± 0.227	6.725	-185.349 ± 0.123	-0.405
FALZ	I	167.545 ± 0.227	6.734	-183.669 ± 0.089	-0.412
PCAF	I	167.545 ± 0.227	6.852	-184.576 ± 0.126	-0.494
BRD2(1)	II	167.545 ± 0.227	6.736	-181.497 ± 0.122	-0.462
BRD2(2)	II	167.545 ± 0.227	6.808	-183.664 ± 0.133	-0.349
BRD3(1)	II	167.545 ± 0.227	6.850	-181.284 ± 0.155	-0.338
BRD3(2)	II	167.545 ± 0.227	6.534	-185.243 ± 0.105	-0.497
BRD4(1)	II	167.545 ± 0.227	6.879	-185.162 ± 0.114	-0.537
BRD4(2)	II	167.545 ± 0.227	6.919	-184.046 ± 0.127	-0.436
BRDT(1)	II	167.545 ± 0.227	6.851	-183.287 ± 0.196	-0.422
CREBBP	III	167.545 ± 0.227	6.844	-184.771 ± 0.087	-0.440
EP300	III	167.545 ± 0.227	6.937	-183.897 ± 0.147	-0.382
BRD1	IV	167.545 ± 0.227	6.918	-184.916 ± 0.100	-0.368
BRD9	IV	167.545 ± 0.227	6.878	-186.011 ± 0.097	-0.446
BRPF1B	IV	167.545 ± 0.227	6.892	-185.495 ± 0.113	–0.423
BAZ2A	V	167.545 ± 0.227	6.829	-184.998 ± 0.129	-0.414
TIF1	V	167.545 ± 0.227	6.760	-179.415 ± 0.153	-0.269
TAF1(1)	VII	167.545 ± 0.227	6.832	-182.439 ± 0.147	-0.437
TAF1(2)	VII	167.545 ± 0.227	6.577	-184.492 ± 0.104	-0.286
TAF1L(2)	VII	167.545 ± 0.227	6.829	-184.453 ± 0.065	-0.402
PB1(5)	VIII	167.545 ± 0.227	6.738	-180.942 ± 0.101	-0.338
SMARCA4	VIII	167.545 ± 0.227	6.754	-180.654 ± 0.117	-0.365

Table S4

Full breakdown of free energy components for the relative calculations in which the effect of the torsional parameters of the benzensulfonamide in bromosporine was evaluated. All free energies are in kcal/mol, and all uncertainties are one standard deviation.

Protein	Family	ΔG^{solv}_{dih}	ΔG^{prot}_{dih}	$\Delta\Delta G_{calc}$
CECR2	I	-2.038 ± 0.019	-1.162 ± 0.003	+0.876 ± 0.019
FALZ	I	-2.038 ± 0.019	-1.462 ± 0.004	+0.576 ± 0.019
PCAF	I	-2.038 ± 0.019	-0.920 ± 0.003	+1.118 ± 0.019
BRD2(1)	II	-2.038 ± 0.019	-0.935 ± 0.001	+1.103 ± 0.019
BRD2(2)	II	-2.038 ± 0.019	-1.163 ± 0.002	+0.875 ± 0.019
BRD3(1)	II	-2.038 ± 0.019	-0.935 ± 0.005	+1.103 ± 0.020
BRD3(2)	II	-2.038 ± 0.019	-1.343 ± 0.001	+0.695 ± 0.019
BRD4(1)	II	-2.038 ± 0.019	-0.538 ± 0.003	+1.500 ± 0.019
BRD4(2)	II	-2.038 ± 0.019	-1.220 ± 0.004	+0.818 ± 0.019
BRDT(1)	II	-2.038 ± 0.019	-0.908 ± 0.002	+1.130 ± 0.019
CREBBP	III	-2.038 ± 0.019	-1.500 ± 0.004	+0.538 ± 0.019
EP300	III	-2.038 ± 0.019	-1.463 ± 0.004	+0.575 ± 0.019
BRD1	IV	-2.038 ± 0.019	-1.570 ± 0.001	+0.468 ± 0.019
BRD9	IV	-2.038 ± 0.019	-1.096 ± 0.001	+0.942 ± 0.019
BRPF1B	IV	-2.038 ± 0.019	-2.279 ± 0.003	-0.241 ± 0.019
BAZ2A	V	-2.038 ± 0.019	-1.034 ± 0.002	+1.004 ± 0.019
TIF1	V	-2.038 ± 0.019	-1.671 ± 0.003	+0.367 ± 0.019
TAF1(1)	VII	-2.038 ± 0.019	-1.223 ± 0.001	+0.815 ± 0.019
TAF1(2)	VII	-2.038 ± 0.019	-1.506 ± 0.002	+0.532 ± 0.019
TAF1L(2)	VII	-2.038 ± 0.019	-1.678 ± 0.003	+0.360 ± 0.019
PB1(5)	VIII	-2.038 ± 0.019	-1.499 ± 0.005	+0.539 ± 0.020
SMARCA4	VIII	-2.038 ± 0.019	-0.894 ± 0.004	+1.144 ± 0.019

Table S5

Scoring function (CSM-Lig)[1] affinity predictions for the ligands RVX-OH and RVX-208. All free energies are in kcal/mol. "Pose" refers to the ligand orientation evaluated in the calculation for which the data are shown on the same row. ΔG_{bind} has been calculated from the pKi at T = 298.15 K using $\Delta G_{bind} = RT\ln(10^{-pKi})$, where R is the gas constant.

Protein	Ligand	Pose	pKi	ΔG_{bind}
BRD2(1)	RVX-OH	H	2.071	-2.825
BRD2(2)	RVX-OH	H	2.021	-2.757
BRD3(1)	RVX-OH	H	2.277	-3.106
BRD3(2)	RVX-OH	H	1.809	-2.468
BRD4(1)	RVX-OH	H	2.335	-3.186
BRD4(2)	RVX-OH	H	2.295	-3.131
BRDT(1)	RVX-OH	H	2.223	-3.033
BRD2(1)	RVX-OH	V	1.628	-2.221
BRD2(2)	RVX-OH	V	1.772	-2.417
BRD3(1)	RVX-OH	V	1.641	-2.239
BRD3(2)	RVX-OH	V	1.909	-2.604
BRD4(1)	RVX-OH	V	1.978	-2.698
BRD4(2)	RVX-OH	V	1.749	-2.386
BRDT(1)	RVX-OH	V	1.605	-2.190
BRD2(1)	RVX-208	V	1.365	-1.862
BRD2(2)	RVX-208	V	1.418	-1.935
BRD3(1)	RVX-208	V	1.025	-1.398
BRD3(2)	RVX-208	V	1.105	-1.507
BRD4(1)	RVX-208	V	1.613	-2.201
BRD4(2)	RVX-208	V	1.541	-2.102
BRDT(1)	RVX-208	V	1.243	-1.696

Table S6

Scoring function (CSM-Lig)[1] affinity predictions for the ligand bromosporine. All free energies are in kcal/mol. ΔG_{bind} has been calculated from the pKi at T=298.15 K using $\Delta G_{bind} = RTln(10^{-pKi})$, where R is the gas constant.

Protein	Family	pKi	ΔG_{bind}
CECR2	I	10.401	-14.190
FALZ	I	10.338	-14.104
PCAF	I	10.478	-14.295
BRD2(1)	II	10.671	-14.558
BRD2(2)	II	10.491	-14.312
BRD3(1)	II	10.538	-14.376
BRD3(2)	II	10.631	-14.503
BRD4(1)	II	10.918	-14.895
BRD4(2)	II	10.177	-13.884
BRDT(1)	II	10.860	-14.816
CREBBP	III	10.813	-14.752
EP300	III	10.727	-14.634
BRD1	IV	3.715	-5.068
BRD9	IV	10.361	-14.135
BRPF1B	IV	10.779	-14.705
BAZ2A	V	10.489	-14.310
TIF1	V	2.848	-3.885
TAF1(1)	VII	2.421	-3.303
TAF1(2)	VII	4.254	-5.804
TAF1L(2)	VII	10.264	-14.003
PB1(5)	VIII	10.101	-13.780
SMARCA4	VIII	10.413	-14.206

Text T1: Discussion on the precision of the calculations.

The errors shown for the binding free energy estimates reflect the statistical uncertainty of the free energy estimator only. The error due to limited sampling was not taken into account and this likely resulted in an underestimate of the true uncertainty of each binding free energy. The uncertainty for decoupling the ligand from the solvent was assessed through five calculations started from different ligand conformations, thereby providing an estimate of uncertainty that includes information about finite sampling. The sample standard deviation of the decoupling free energy from these five calculations was 0.23 kcal/mol, about four-fold the size of the error provided by the statistical estimator (0.05 kcal/mol) for each individual calculation. In order to gauge the magnitude of the uncertainty due to sampling for the complex simulations, the calculation for BRD4(1) was repeated three times, obtaining a standard deviation for decoupling the ligand from the protein of 0.54 kcal/mol, about four to five times the statistical uncertainty given by the estimator for each individual simulation (0.11 kcal/mol, 0.15 kcal/mol, and 0.16 kcal/mol, respectively, for the three repeats). Based on the overall estimates of the binding affinities (-11.28 ± 0.25 kcal/mol, -10.36 ± 0.27 kcal/mol, -11.46 ± 0.28 kcal/mol), the sample standard deviation obtained was 0.59 kcal/mol, about two-fold the uncertainties of each single repeat, which were derived taking into account the five calculations for the free ligand in solution but a single one for the ligand in the complex. For all other proteins, a single repeat of the complex simulation was used due to the computational cost of the calculations, however, based on what discussed above, the reader should bear in mind that the real uncertainty in the calculations could be the double of the ones reported. Finally, we note that the use of ensemble calculations as very

recently described by Bhati *et al*.[2] provide a route to more precise and reproducible calculations. However, there is of course an additional computational cost.

Text T2: Discussion on the potential sources of error for bromosporine.

Firstly, we note that it is very difficult know exactly what is causing the inaccuracies seen in the predictions for bromosporine as long as the issue itself cannot be resolved. Possible sources of error include: 1) the presence of alternative unknown binding poses due to the ability of different BRDs to modulate the binding mode; 2) convergence issues; 3) force field deficiencies. Working by exclusion, we would estimate model deficiencies to be the most likely cause, followed by convergence issues, due to the following reasons.

Alternative binding poses

(1) Currently, four structures of human BRDs (from three different families) in complex with bromosporine are available (PDB IDs: 5C7N, 5IGK, 5IGL, 5IGM), and they all show the same conserved binding mode.

(2) Assuming the existence of an alternative unknown pose for some BRDs, the fact that the majority of binding affinities are overestimated still points toward force field deficiencies. In fact, a hypothetical alternative binding pose would need to have a higher calculated affinity than the pose we considered in order to be predicted to be the most stable pose. If it were calculated to have higher affinity, then the error in affinity would be even larger than the one reported (due to either force field - or convergence - issues). If it were not calculated to have larger affinity than the pose we considered, it would erroneously be considered a secondary high-energy (less

stable) pose, which would still point towards issues with the force field or severe convergence problems.

(3) Free energy calculations for BRD9 and BRPF1B overestimate the binding free energy despite the fact that X-ray structures confirm the correctness of the binding pose used.

(4) There is still a chance that alternative binding poses for bromosporine depending on the BRD considered do exist, and that the force field is in fact accurate, but there are systematic and severe convergence problems with most of the calculations. However, given the previous points and what discussed below, we think this is an unlikely scenario, and still one that would need another source of error affecting the results in order to explain the inaccuracy of some predictions.

(5) It is also possible this issue may affect only the predictions that underestimate the binding free energy, in which case the errors can be explained without having to also postulate force fields or convergence problems. On the other hand, the errors when the predictions overestimate affinity would still remain unexplained.

Convergence

(1) Repeated BRD4(1) calculations indicate the results are quite precise, so that the uncertainty alone cannot justify large errors. It is true that BRD4(1) is one of the BRDs for which the complex has been resolved by X-ray, however, we still used the docking pose for the calculations. On the other hand, because we started form the same coordinates for the repeats, precision does not guarantee the calculations have visited all relevant states both for the apo and holo simulations (corresponding to a fully coupled or decoupled ligand). In fact, we might

be sampling very well the phase space around the initial structure, accessible within the timescales simulated (hence the precise results), while there might be other conformations that are important for determining affinity that are not reached with 15 ns windows.

(2) Following from point 1, we ran some additional tests with PCAF, a BRD among the ones showing a large deviation from ITC data. Before the re-parametrization of the sulfonamide, the calculations returned a binding free energy of -10.7 ± 0.3 kcal/mol, while ITC measured -7.0 ± 0.1 kcal/mol. We extended the calculations for PCAF to 30 ns per window, to see whether the prediction would change substantially from the estimate using 15 ns windows. However, the longer calculations returned a binding free energy of -10.5 ± 0.2 kcal/mol, within the uncertainty of the results obtained with the shorter calculations.

(3) As an additional test, we ran the calculations for PCAF using the crystallographic pose of bromosporine (PDB ID 5C7N) rather than the docking one. This was done in order to check whether the slight difference in structure between the X-ray binding pose and the pose returned by docking could have an unexpectedly large effect on the results. However, also in this case, the predicted binding free energy was not significantly different from the original calculation (-10.4 ± 0.2 kcal/mol).

(4) Intuitively, when the affinity is overestimated due to convergence issues, one could expect this to be caused by a failure to sample low energy states for the apo simulation (i.e. decoupled ligand), which would not be compatible with high affinity binding. The failure to sample such states in the apo protein would in fact allow the ligand to bind without an energy penalty (due to, for instance, a side chain that needs to move from its most stable orientation when the ligand is present) that would reduce its binding free energy. For the proteins, we used a mixture of apo and holo X-ray structures to initialize the calculations. Therefore, we hypothesized that

when starting from a holo structure, there might be a bias towards higher affinity binding (and the opposite when starting from apo structures) due to limited sampling. However, no significant difference between the errors obtained was observed when using apo versus holo X-ray structures. Apo structures returned largely overestimated binding affinities just as the holo ones did.

Text T3: Additional considerations and information useful for new users.

The methodology used here is transferable to other ligand-protein systems, and it is based on freely available software and force fields. However, as for all MD simulations, each system might present its own challenges. Firstly, convergence may be an issue when dealing with flexible proteins or ligands. While ligand flexibility might affect the precision of the calculations, sampling limitations in this case should often be identifiable and alleviated, or possibly resolved, with longer simulations or enhanced sampling schemes. On the other hand, convergence issues due to very slow degrees of freedom in the protein, such as large conformational changes upon binding, might not be easily resolved and could have a large impact on the predicted binding free energies. If conformational changes are known, there might be ways to tackle the issue. For instance, Mobley *et al*.[3] used Potential of Mean Force (PMF) calculations in order to take into account the impact of a valine rearrangement on the binding free energy. Lin *et al*.[4] have too used PMF calculations to take into account the large loop movement involved in the binding of type-II kinase inhibitors. However, when conformational changes are unknown, and not sampled, they can have a large effect on the results of the calculations. The position of conserved solvent molecules might be important as well. If conserved water molecules are not modelled correctly, and diffusion to their binding site is hindered during the course of the simulations, this too can result in convergence issues and biased results. Therefore, when water molecules are resolved crystallographically, we generally prefer not to remove them (unless there are specific reasons to do so). General considerations on model quality as per standard MD simulations apply here too: the quality of the protein structure and the suitability of the force field for the molecules studied can both affect the resulting binding free energies. Finally, charged compounds need to be treated with special

care, due to the effects of having a non-neutral simulation box with a periodic treatment of electrostatics (see the work of Rocklin *et al*.[5] for a detailed explanation of the issue).

New users starting to learn how to run absolute free energy calculations might find themselves in front of a steep learning curve. The theory behind the calculations is considerable and detailed practical information within single research publications might appear either scarce or overwhelming. Below are some key online tutorials and resources that essentially implement the wisdom acquired from theoretical foundations[6-9] and suggested good practice[10-13] and technical considerations[5,14-17] over the years. Several excellent reviews have appeared[18-21] which make reference to previous efforts[22-27] along these lines.

Online tutorials and resources:

- http://www.alchemistry.org
- http://www.alchemistry.org/wiki/Tutorials
- http://www.ks.uiuc.edu/Training/Tutorials/freenergy-index.html
- http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/free_energy/

Supplementary References

(1) Pires, D. E. V.; Ascher, D. B. *Nucleic Acids Res.* **2016**, *44*, W557.
(2) Bhati, A. P.; Wan, S.; Wright, D. W.; Coveney, P. V. *J. Chem. Theor. Comput.* **2016**, *Just Accepted Manuscript*.
(3) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theor. Comput.* **2007**, *3*, 1231.
(4) Lin, Y.-L.; Meng, Y.; Jiang, W.; Roux, B. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1664.
(5) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. *J. Chem. Phys.* **2013**, *139*, 184103.
(6) Chipot, C.; Pohorille, A. *Free energy calculations: Theory and applications in chemistry and biology*; Springer, **2007**; Vol. 86.
(7) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(8) Wereszczynski, J.; McCammon, J. A. *Quart. Rev. Biophys.* **2012**, *45*, 1.
(9) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.
(10) Hansen, N.; van Gunsteren, W. F. *J. Chem. Theor. Comput.* **2014**, *10*, 2632.
(11) Klimovich, P.; Shirts, M.; Mobley, D. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397.
(12) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235.
(13) Shirts, M. R. In *Computational drug discovery and design*; Baron, R., Ed. **2012**; Vol. 819, 425.
(14) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *J. Phys. Chem. B.* **2003**, *107*, 9535.
(15) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125*, 084902.
(16) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
(17) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B.* **2007**, *111*, 13052.
(18) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struc. Biol.* **2011**, *21*, 150.
(19) Mobley, D. L.; Dill, K. A. *Structure* **2009**, *17*, 489.
(20) Mobley, D. L.; Klimovich, P. V. *J. Chem. Phys.* **2012**, *137*, 230901.
(21) Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A. *Curr. Opin. Sruct. Biol.* **2016**, *36*, 25.
(22) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. *Chem. Sci.* **2016**, *7*, 207.
(23) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. *J Mol Biol* **2009**, *394*, 747.
(24) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*.
(25) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J Mol Biol* **2007**, *371*, 1118.
(26) Rocklin, G. J.; Boyce, S. E.; Fischer, M.; Fish, I.; Mobley, D. L.; Shoichet, B. K.; Dill, K. A. *J. MOl. Biol.* **2013**, *425*, 4569.
(27) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. *J. Am. Chem. Soc.* **2015**, *137*, 2695.