Supporting information for: Rapid, accurate, precise and reliable relative free energy prediction using ensemble based thermodynamic integration

Agastya P. Bhati, Shunzhou Wan, David W. Wright, and Peter V. Coveney*

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

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

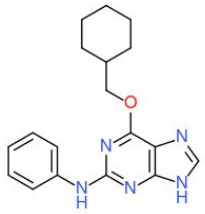
In the main article, we have described our approach which we refer to as thermodynamic integration with enhanced sampling (TIES) to rapidly calculate accurate, precise and reproducible relative binding affinities of ligand-protein complexes. All our results along with the quantification of their accuracy and precision are detailed there. In the Supporting Information, we provide additional details for our results. Figures S1 to S5 shows the chemical structures and experimental binding affinities of all the ligands used in our calculations for the five target proteins. The structures are downloaded from the Supporting Information of an open access article.[S1] The experimental binding affinities for thrombin are taken from the reported ITC experiments,[S2] while for other targets they are calculated with equation 1 using the reported IC_{50} or K_i values.[S3–S7] Tables S1 to S5 provide the predicted and experimental $\Delta\Delta G$ values for all the 55 transformations studied, while Figure S6 shows the

*To whom correspondence should be addressed

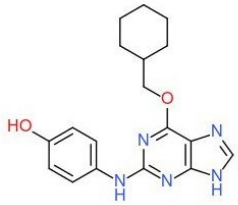
correlation between TIES-predictions and the experimental data for each biomolecular system separately. The $\Delta\Delta G_{TIES}$ and σ_{TIES} are calculated as described in Section 3.3 of the main article.

$$\begin{aligned} \Delta G_{exp} &= RT \ln K_i \\ \Delta G_{exp} &\approx RT \ln(IC_{50}) \end{aligned} \tag{1}$$

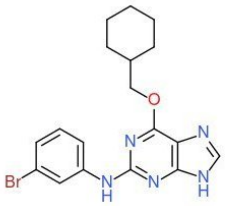
where T=297 K and R is the gas constant.Figure S7 shows the variation of the mean energy derivative with the simulation length for one of the transformations studied providing a justification to our choice of 2 ns long equilibration phase. Thereafter, we provide a more detailed discussion of selected ligand-protein interactions of note. In particular, we discuss how to compute relative binding affinities in situations where there is a significant change in the charge distribution within the pair of ligands studied. Figure S10 displays a comparison of the relative binding affinity predictions from TIES with those from Wang et al. using FEP method[S1] for the 18 transformations found common in both the studies. Table S6 lists the set of ligands forming closed cycles in the transformations studied and the corresponding values of hysteresis. Figure S11 shows the variation of the uncertainties in TIES-predictions with the size of perturbation for the transformations in ligands binding with PTP1B.



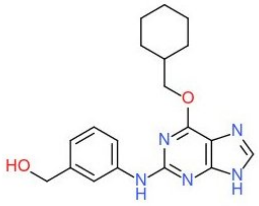
L1Q; ΔG_{exp}=-8.18



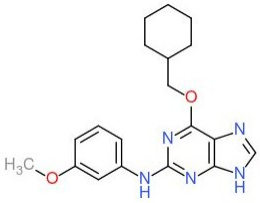
LI9; ΔG_{exp}=-9.74



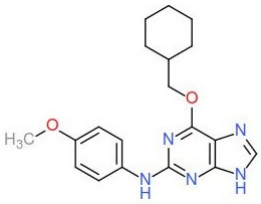
L17; ΔG_{exp}=-7.04



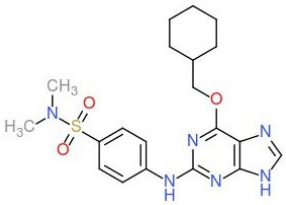
L20; ΔG_{exp}=-8.72



L21; ΔG_{exp}=-7.83

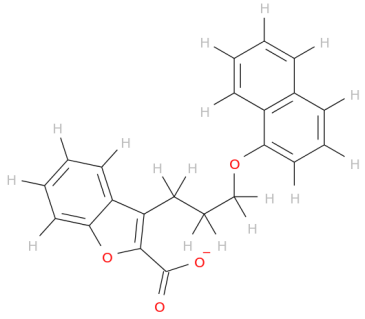


L26; ΔG_{exp}=-8.43



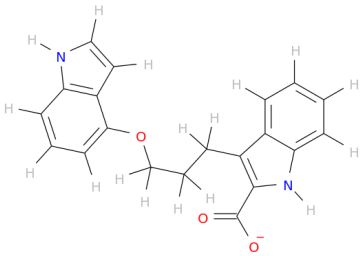
L29; ΔG_{exp}=-9.88

Figure S1: **Chemical structures and experimental binding affinities (kcal/mol) of CDK2 ligands. No errors are available on the values used in this study as quoted above.**



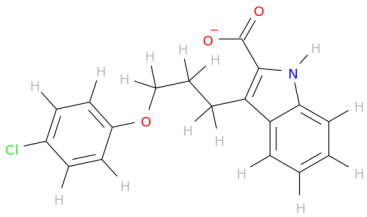
l1.pdb

L1; ΔG_{exp}=-8.24



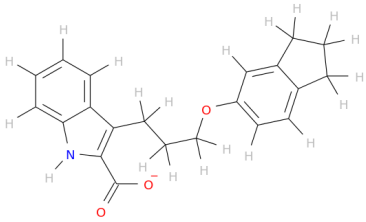
l2.pdb

L2; ΔG_{exp}=-6.66



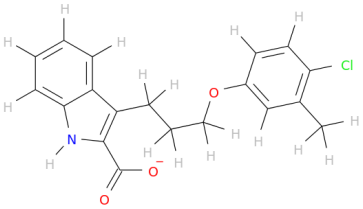
l3.pdb

L3; ΔG_{exp}=-6.88



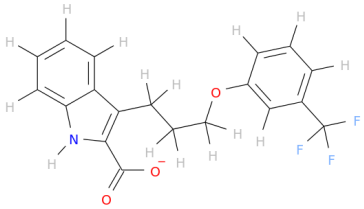
l4.pdb

L4; ΔG_{exp}=-7.60



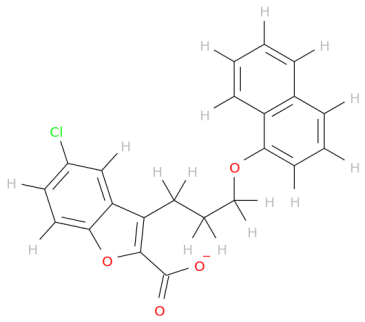
l5.pdb

L5; ΔG_{exp}=-8.81



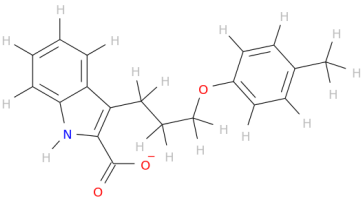
l6.pdb

L6; ΔG_{exp}=-7.92



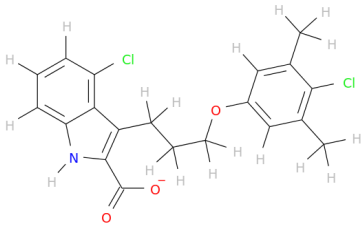
l8.pdb

L8; ΔG_{exp}=-7.69



l9.pdb

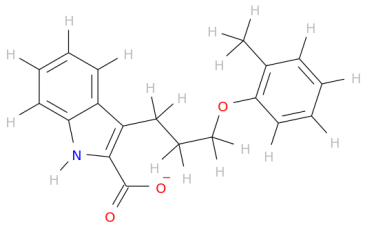
L9; ΔG_{exp}=-6.58



l12.pdb

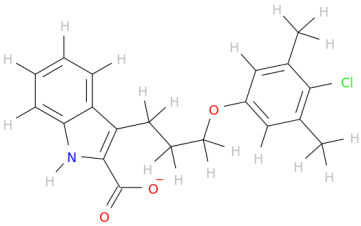
L12; ΔG_{exp}=-9.33

Figure S2: **Chemical structures and experimental binding affinities (kcal/mol) of MCL1 ligands. No errors are available on the values used in this study as quoted above.**



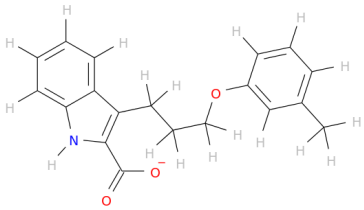
l13.pdb

L13; ΔG_{exp}=-6.62



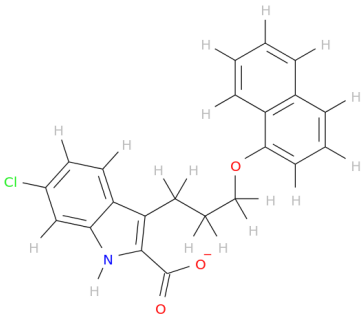
l16.pdb

L16; ΔG_{exp}=-8.95



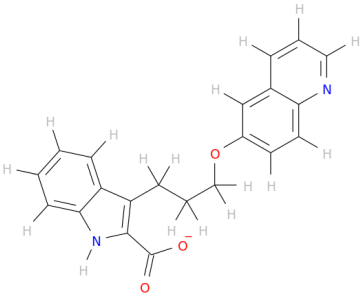
l17.pdb

L17; ΔG_{exp}=-7.85



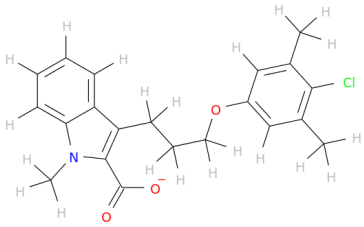
l18.pdb

L18; ΔG_{exp}=-9.78



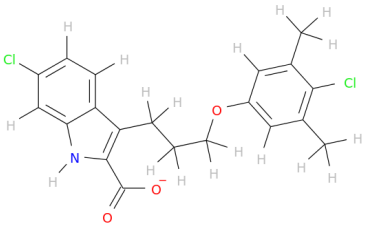
l32.pdb

L32; ΔG_{exp}=-5.78



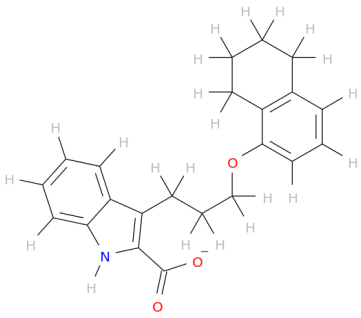
l34.pdb

L34; ΔG_{exp}=-9.26



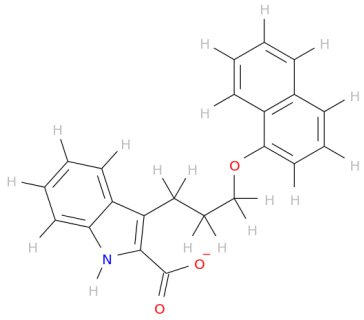
l35.pdb

L35; ΔG_{exp}=-9.96



l38.pdb

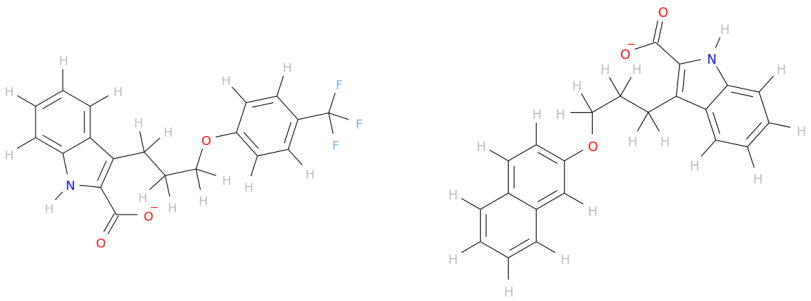
L38; ΔG_{exp}=-8.95



l39.pdb

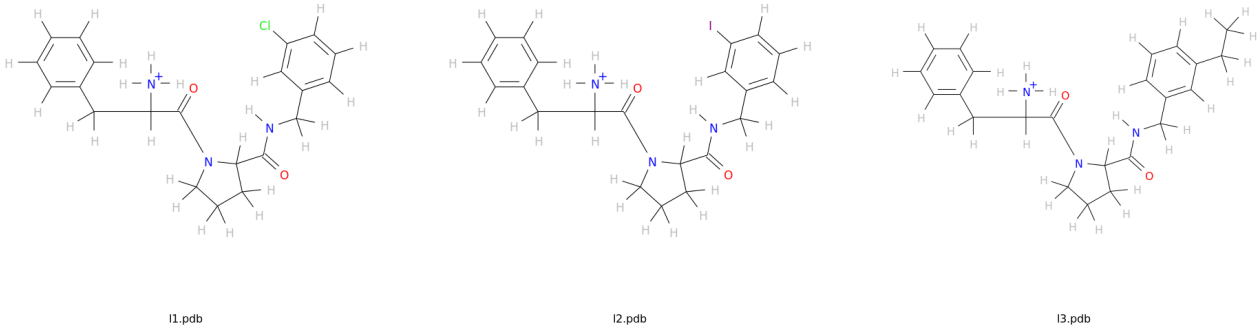
L39; ΔG_{exp}=-8.90

Figure S2: **Chemical structures and experimental binding affinities (kcal/mol) of MCL1 ligands (continued). No errors are available on the values used in this study as quoted above.**



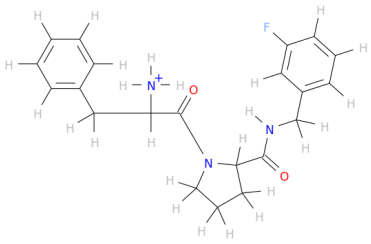
L41; ΔG_{exp}=-6.87 **L42**; ΔG_{exp}=-7.03

Figure S2: **Chemical structures and experimental binding affinities (kcal/mol) of MCL1 ligands (continued). No errors are available on the values used in this study as quoted above.**



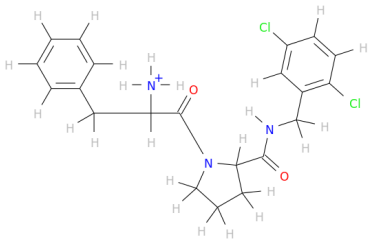
L1; ΔG_{exp}=-8.46 **L2**; ΔG_{exp}=-8.25 **L3**; ΔG_{exp}=-7.86

Figure S3: **Chemical structures and experimental binding affinities (kcal/mol) of thrombin ligands. No errors are available on the values used in this study as quoted above.**



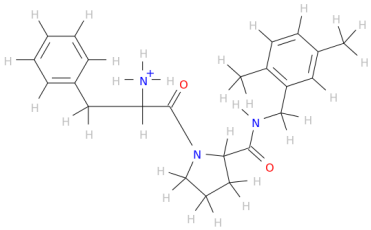
l4.pdb

L4; ΔG_{exp}=-7.48



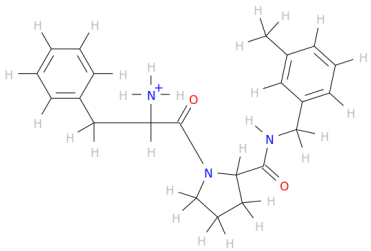
l5.pdb

L5; ΔG_{exp}=-9.18



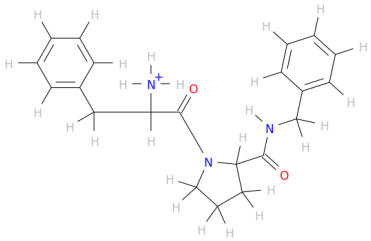
l6.pdb

L6; ΔG_{exp}=-8.22



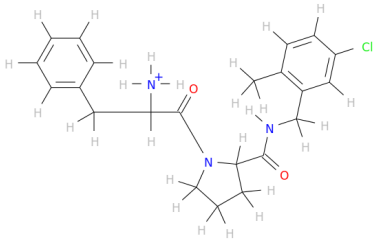
l7.pdb

L7; ΔG_{exp}=-8.32



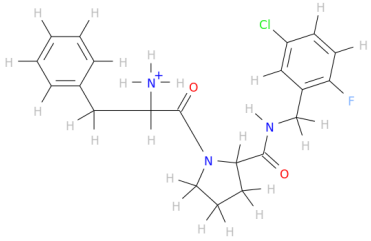
l8.pdb

L8; ΔG_{exp}=-7.58



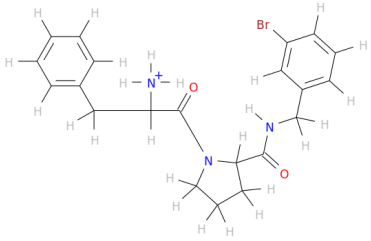
l9.pdb

L9; ΔG_{exp}=-8.89



l10.pdb

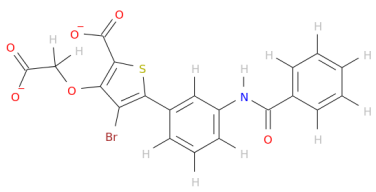
L10; ΔG_{exp}=-8.91



l11.pdb

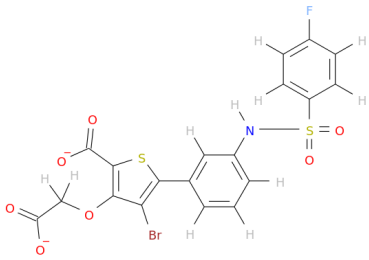
L11; ΔG_{exp}=-8.56

Figure S3: **Chemical structures and experimental binding affinities (kcal/mol) of thrombin ligands (continued). No errors are available on the values used in this study as quoted above.**



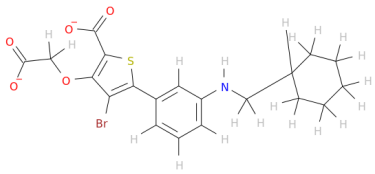
l1.pdb

L1; ΔG_{exp}=-7.72



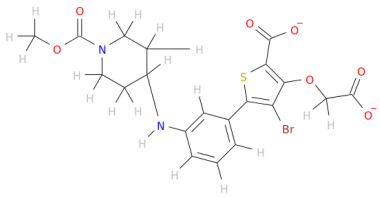
l2.pdb

L2; ΔG_{exp}=-8.65



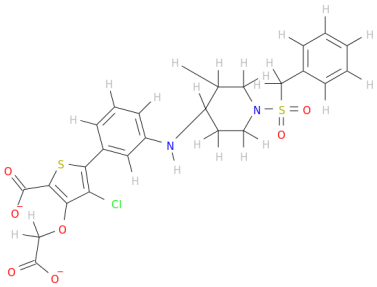
l3.pdb

L3; ΔG_{exp}=-9.11



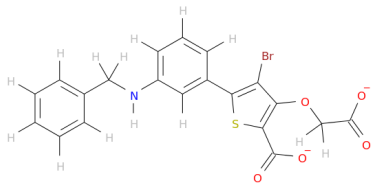
l4.pdb

L4; ΔG_{exp}=-8.72



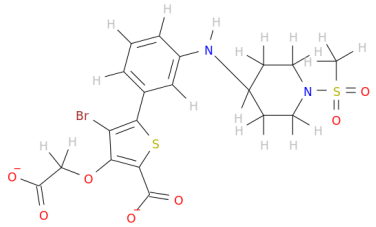
l6.pdb

L6; ΔG_{exp}=-12.47



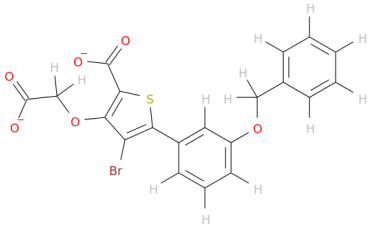
l7.pdb

L7; ΔG_{exp}=-8.61



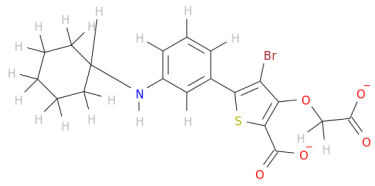
l8.pdb

L8; ΔG_{exp}=-10.01



l10.pdb

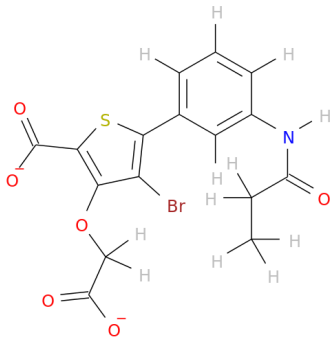
L10; ΔG_{exp}=-8.39



l11.pdb

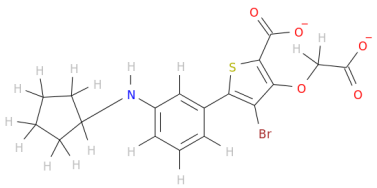
L11; ΔG_{exp}=-9.08

Figure S4: **Chemical structures and experimental binding affinities (kcal/mol) of PTP1B ligands. No errors are available on the values used in this study as quoted above.**



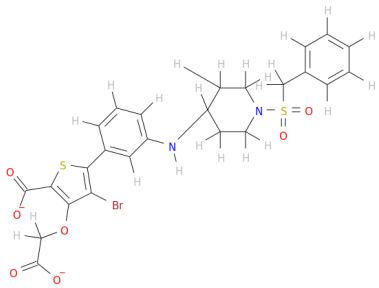
l12.pdb

L12; ΔG_{exp}=-7.75



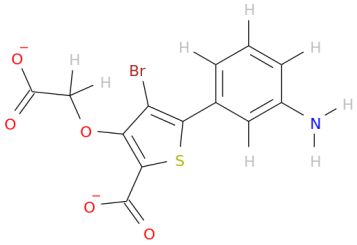
l13.pdb

L13; ΔG_{exp}=-8.72



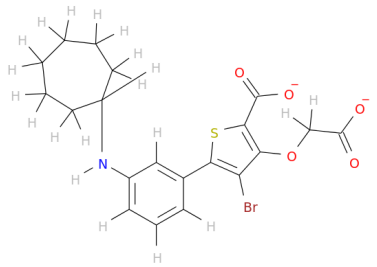
l14.pdb

L14; ΔG_{exp}=-11.42



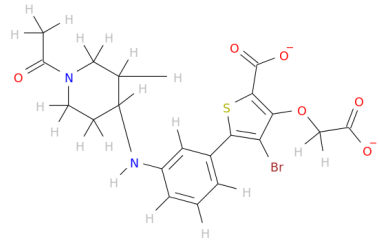
l19.pdb

L19; ΔG_{exp}=-7.85



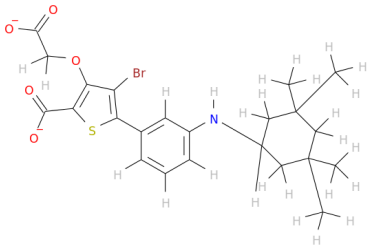
l20.pdb

L20; ΔG_{exp}=-9.41



l22.pdb

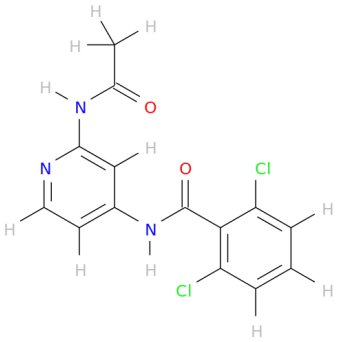
L22; ΔG_{exp}=-9.14



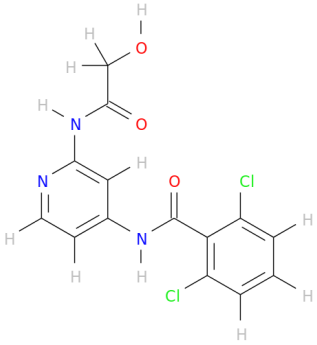
l23.pdb

L23; ΔG_{exp}=-10.12

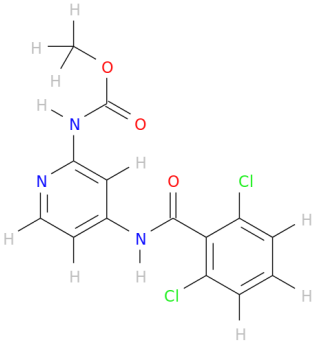
Figure S4: **Chemical structures and experimental binding affinities (kcal/mol) of PTP1B ligands (continued). No errors are available on the values used in this study as quoted above.**



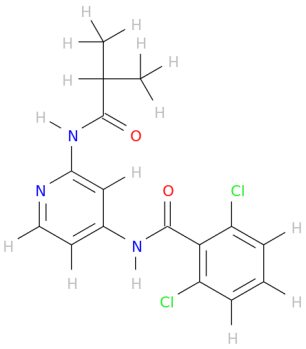
L1; ΔG_{exp}=-9.54



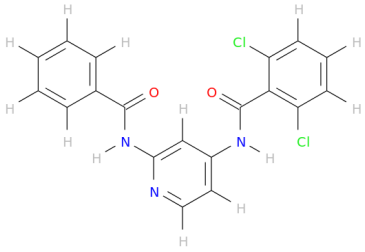
L3; ΔG_{exp}=-8.98



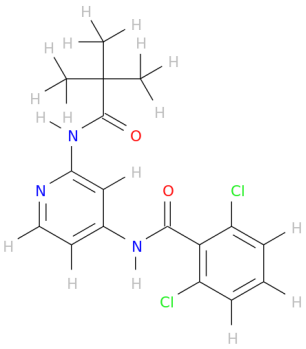
L5; ΔG_{exp}=-9.21



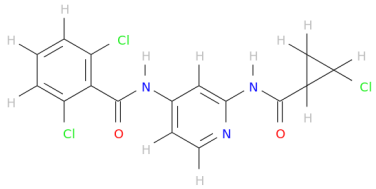
L6; ΔG_{exp}=-8.26



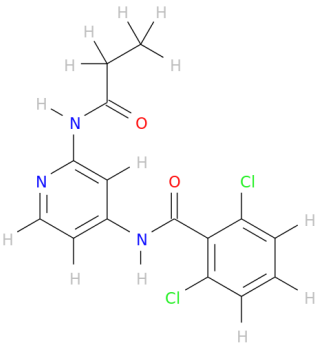
L8; ΔG_{exp}=-7.75



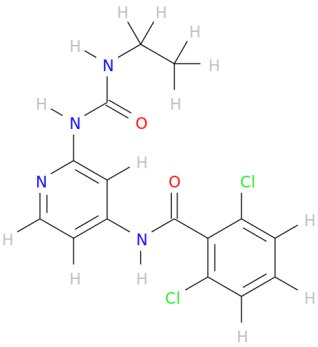
L10; ΔG_{exp}=-7.42



L11; ΔG_{exp}=-11.28



L15; ΔG_{exp}=-9.78



L16; ΔG_{exp}=-10.53

Figure S5: **Chemical structures and experimental binding affinities (kcal/mol) of TYK2 ligands. No errors are available on the values used in this study as quoted above.**

Table S1: $\Delta\Delta G_{TIES}$ for CDK2. All values are in kcal/mol. No errors are available on the experimental data.

Transformation	$\mathbf{\Delta\Delta G_{TIES}}$	$\sigma_{\mathbf{TIES}}$	$\mathbf{\Delta\Delta G_{exp}}$
L1Q-LI9	1.96	0.26	1.56
L1Q-L20	0.59	0.23	0.54
L1Q-L21	-1.33	0.15	-0.35
L20-L21	-1.90	0.26	-0.89
L1Q-L26	1.20	0.19	0.25
L1Q-L29	2.49	0.25	1.70
L1Q-L17	-0.17	0.16	-1.14

Table S2: $\Delta\Delta G_{TIES}$ for MCL1. All values are in kcal/mol. No errors are available on the experimental data.

Transformation	$\mathbf{\Delta\Delta G_{TIES}}$	$\sigma_{\mathbf{TIES}}$	$\mathbf{\Delta\Delta G_{exp}}$
L2-L4	2.97	0.31	0.94
L6-L41	-0.03	0.23	-1.05
L3-L5	1.21	0.21	1.93
L3-L16	2.92	0.17	2.07
L16-L34	-0.33	0.69	0.31
L12-L35	2.57	0.82	0.63
L2-L32	1.77	0.26	-0.88
L32-L42	1.45	0.37	1.25
L38-L42	-3.33	0.29	-1.92
L32-L38	3.59	0.33	3.17
L39-L42	-2.40	0.21	-1.87
L18-L39	-3.50	0.71	-0.88
L1-L8	0.09	0.79	-0.55
L8-L18	3.83	0.69	2.09
L17-L9	-0.17	0.24	-1.27
L13-L17	0.60	0.32	1.23

Table S3: $\Delta\Delta G_{TIES}$ for thrombin. All values are in kcal/mol. No errors are available on the experimental data.

Transformation	$\mathbf{\Delta\Delta G_{TIES}}$	$\sigma_{\mathbf{TIES}}$	$\mathbf{\Delta\Delta G_{exp}}$
L8-L1	1.04	0.24	0.88
L1-L8	-0.61	0.24	-0.88
L6-L7	-0.48	0.26	0.10
L3-L5	2.25	0.23	1.32
L5-L6	-2.12	0.21	-0.96
L1-L4	-1.75	0.35	-0.98
L2-L5	0.38	0.23	0.93
L4-L11	2.18	0.28	1.08
L7-L3	0.41	0.27	-0.46
L1-L9	1.08	0.25	0.43
L4-L10	2.42	0.24	1.43

Table S4: $\Delta\Delta G_{TIES}$ for TYK2. All values are in kcal/mol. No errors are available on the experimental data.

Transformation	$\mathbf{\Delta\Delta G_{TIES}}$	$\sigma_{\mathbf{TIES}}$	$\mathbf{\Delta\Delta G_{exp}}$
L1-L3	0.11	0.07	-0.56
L1-L6	-1.38	0.09	-1.28
L6-L11	2.28	0.28	3.02
L6-L10	-1.15	0.14	-0.84
L15-L16	1.08	0.21	0.75
L1-L8	-0.94	0.08	-1.79
L5-L16	0.19	0.16	1.32
L1-L15	0.07	0.06	-0.33
L15-L6	-1.51	0.09	-1.52
L1-L10	-2.21	0.21	-2.12
L15-L10	-2.59	0.10	-2.36

Table S5: $\Delta\Delta G_{TIES}$ for PTP1B. All values are in kcal/mol. No errors are available on the experimental data.

Transformation	$\mathbf{\Delta\Delta G_{TIES}}$	$\mathbf{\sigma_{TIES}}$	$\mathbf{\Delta\Delta G_{exp}}$
L1-L2	1.54	0.33	0.93
L10-L12	-0.14	0.51	-0.64
L19-L3	1.19	0.29	1.26
L3-L7	0.04	0.15	-0.50
L3-L23	0.80	0.32	1.01
L11-L23	1.54	0.15	1.04
L13-L20	0.68	0.24	0.69
L6-L14	-0.81	0.44	-1.05
L8-L14	1.40	0.34	1.41
L4-L22	-0.61	0.23	0.42

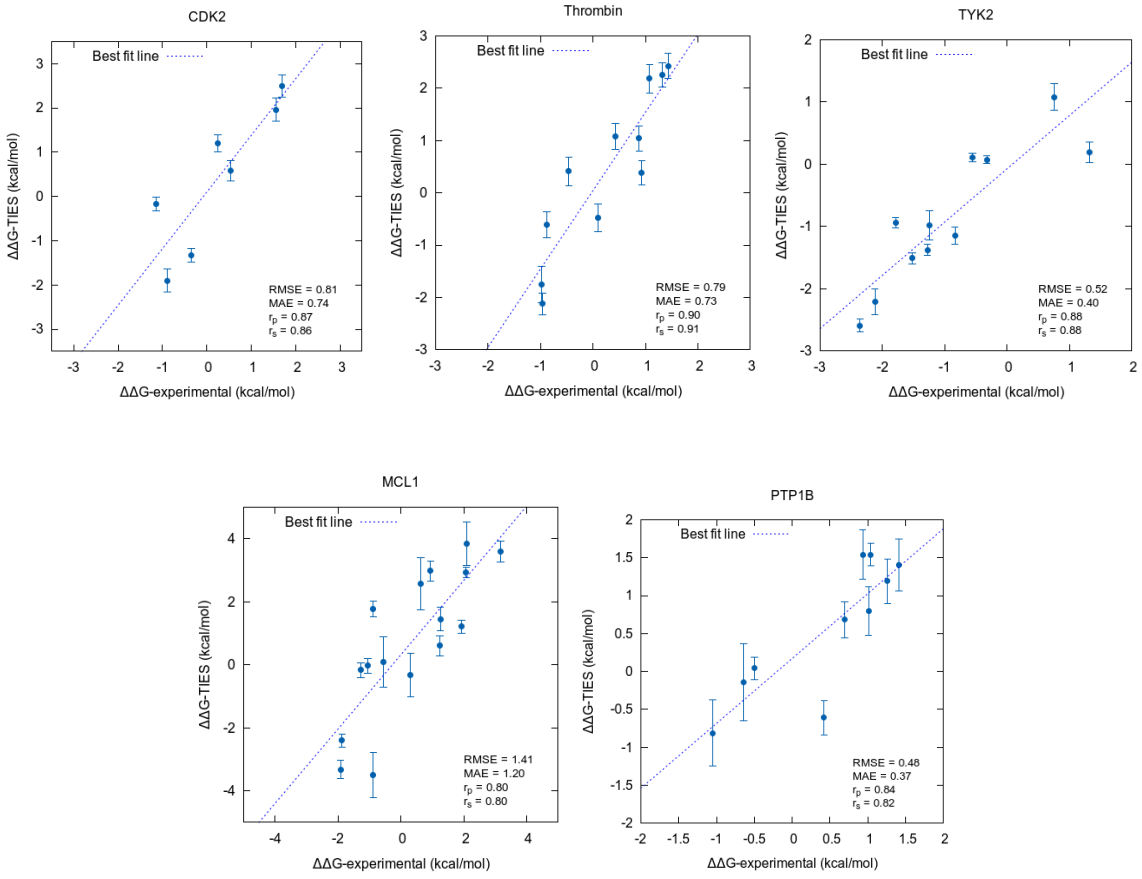


Figure S6: Correlation between TIES-predicted binding affinities and the experimental data for each biomolecular system shown separately. The uncertainties in the TIES predictions are included as error bars. The dashed line is the regression line in each case.

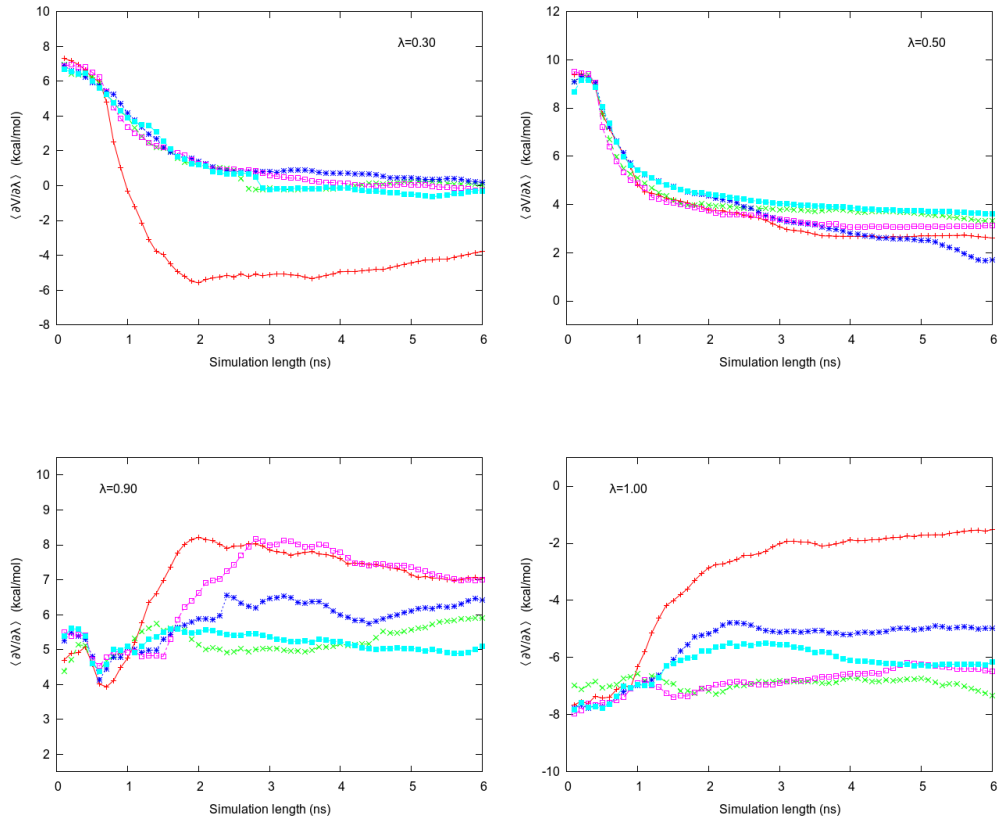


Figure S7: Variation of $\langle \partial V/\partial \lambda \rangle$ with the simulation length (including both equilibration and production phases) for all five replicas at four different λ windows in the case of the transformation from ligand L1 to L4 binding to thrombin. It is clearly visible that all replicas converge at about 2 ns which is the length of the equilibration run in our existing protocol. The variation in the final converged values of $\langle \partial V/\partial \lambda \rangle$ for different replicas at a given λ window as shown above emphasises the advantage of performing ensemble simulation.

Discussion of selected ligand-protein interactions

As noted in the main text in the intermediate λ-windows of TIES calculations the electrostatic interactions are weakly scaled which can result in larger fluctuations of $\partial V/\partial \lambda$ (due to loss or gain of strong electrostatic interactions). This results in larger variation in the stochastic integral (equation 2 in the main article) and hence less precise prediction of $\Delta\Delta G$ using the standard TIES protocol as mentioned in section 3.2 of the main article. In this study, 5 out of the 16 transformations of MCL1 ligands have a charged carboxylate group in the perturbing region, with the uncertainties of their $\Delta\Delta G$ predictions lying in the range of

0.7-0.8 kcal/mol, while the uncertainties of the remaining 11 predictions for MCL1 ligands lie within the range of 0.2-0.4 kcal/mol (see Table 1). In such cases, one can modify the standard protocol described here by, for example, increasing the ensemble size at various λ-windows and/or excluding the charged group from the alchemically mutating part of the ligands to further emphasise this point. In the following paragraph we describe one such case including a comparison of its results with and without charged groups in the perturbing region of the ligand.

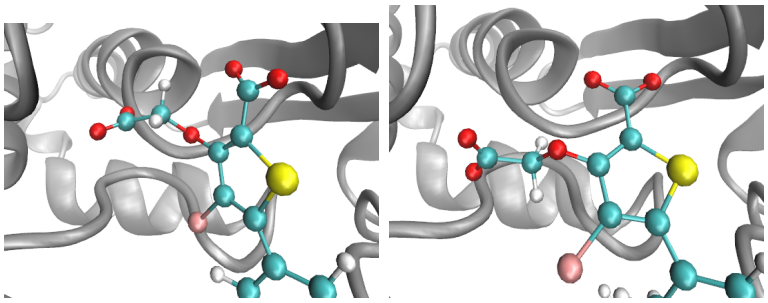


Figure S8: Two different conformations of the flexible carboxylate group of ligand L14 inside the binding pocket of PTP1B taken from the same molecular dynamics simulation. Due to three adjacent rotatable bonds through which the carboxylate group is attached to the ligand, the corresponding dihedral angle changes substantially inside the binding pocket when the electrostatic interactions are scaled to small values. The protein is shown in grey ribbon and ligand atoms are colored by element; hydrogen in white, carbon in cyan, oxygen in red and nitrogen in blue.

The ligands studied with PTP1B all have two carboxylate groups present in the active site of the protein. One of them is attached to the thiophene ring through three rotatable bonds and the binding pocket provides it with enough space to move around. Because of this, the carboxylate group is highly flexible. Figure S8 shows a transformation (L6 to L14) where this carboxylate group forms part of the perturbing region of the ligand within a TIES calculation. Two different conformations of this flexible carboxylate group are shown at an intermediate λ-window where the electrostatic interactions are scaled down to very small values. It is clear that, in the absence of strong electrostatic interactions, such a

charged group is highly flexible, leading to a diminution in both the accuracy and precision of our predictions. In this case, $\Delta\Delta G_{TIES} = 0.6 \pm 1.7$ kcal/mol while the corresponding experimental value is -1.0 kcal/mol. This problem can be dealt with by excluding the charged species from the perturbing region of the ligand. Performing the TIES calculation again for this transformation, after excluding the carboxylate groups from the perturbing regions of the ligand, substantially improves the accuracy and precision of our prediction, yielding a new value of $\Delta\Delta G_{TIES} = -0.8 \pm 0.4$ kcal/mol very close to the experimental relative binding affinity.

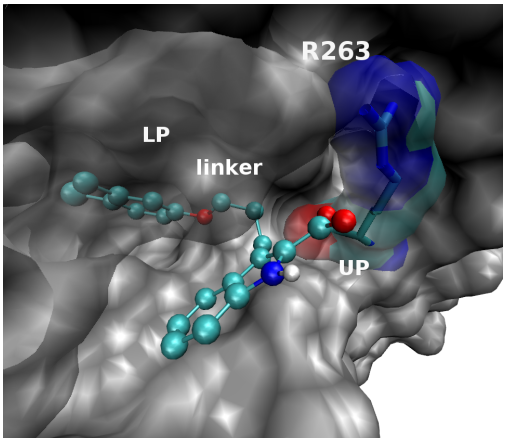


Figure S9: The generic structure of MCL1 ligands with the four membered linker joining the two ends of the ligand containing different functional groups. One end of the ligand is bound to the lower pocket (LP) of the protein and is relatively rigid, while the other end of the ligand is bound to the upper pocket (UP) of the protein and is quite flexible. The carboxylate group present at the second end of the ligand interacts with the side chain of the arginine residue of the protein (R263). The protein surface is shown in grey, the arginine residue (R263) is shown in colored electrostatic surface and ligand atoms are colored by element; hydrogen in white, carbon in cyan, oxygen in red and nitrogen in blue. All non-polar hydrogens have been excluded for clarity.

It is evident from Table 1 and Figure 5 in the main document that the TIES predictions for MCL1 have larger deviations from the experimental results, with the largest RMSE and MAE. Such behavior can be attributed to the highly flexible nature of the ligand. Figure S9 shows the structure of one of the MCL1 ligands. Generically, they contain a 4-membered linker which connects the two ends of the ligand. One end of the ligand is a hydrophobic

aromatic system which is buried deep into the lower pocket (LP) of the protein (Figure S9). Another end of the ligand is a 6,5-fused heterocyclic carboxylic acid which is held at the upper pocket (UP) (Figure S9). The latter interacts with the positively charged side chain of the arginine residue (R263). Since the linker chain has four rotatable bonds, the two ends of the ligand are free to move with respect to each other. The end located in LP is rigid as compared to the other end. Its interaction with R263 becomes too weak to hold it in a stable position at intermediate λ-windows due to the scaling down of electrostatic interactions. This accounts for why the least accurate results occur for the MCL1 predictions as compared to the other targets.

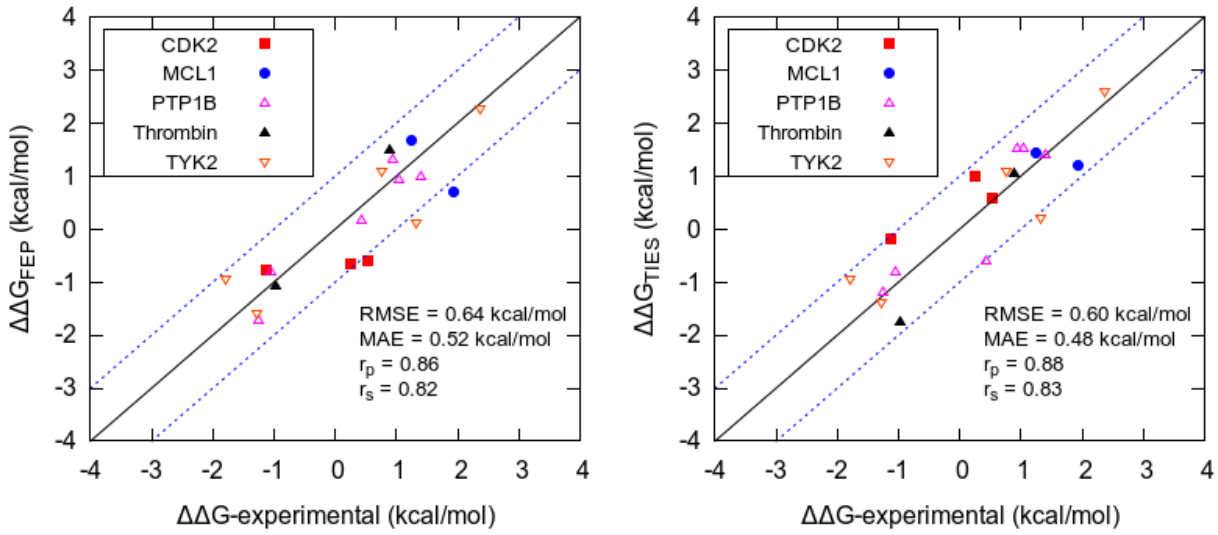


Figure S10: A comparison between the 18 ligand transformations which are common between the ones studied here and those studied by Wang *et al.* using their FEP methodology.[S1] Results from the FEP methodology are shown on the left, while those from TIES are shown on the right. TIES exhibits marginally better accuracy with slightly smaller RMSE and MAE and slightly larger Pearson's r and Spearman's ρ. Only one of the 18 TIES predictions lies outside the 1 kcal/mol window from the experimental data and one directionally disagrees with the experimental value; the corresponding numbers for the FEP predictions are three and two respectively.

Table S6: Hysteresis in the TIES predictions for the closed cycles formed by the ligand transformations studied.

Protein	Ligands forming closed cycles	Hysteresis (kcal/mol)
CDK2	L1Q,L20,L21	0.02
MCL1	L32,L42,L38	1.19
TYK2	L1,L6,L10,L15	0.01
	L1,L6,L10	0.32
	L1,L10,L15	0.31
	L1,L15,L6	0.06

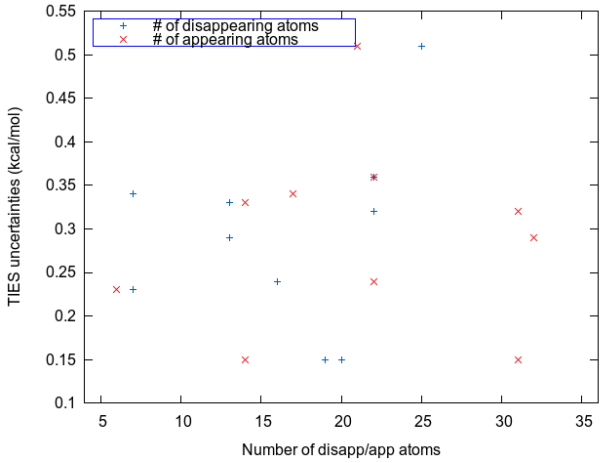


Figure S11: Variation of uncertainties in the TIES-predictions with the size of the perturbation (measured in terms of the number of appearing or disappearing atoms) for the transformations of ligands binding with PTP1B. In this case, and more generally, we observe no clear dependence of the uncertainty on the size of the perturbation.

Bibliography

(S1) 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.; R., A. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695.

(S2) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. More than a simple lipophilic contact: A detailed thermodynamic analysis of nonbasic residues in the S1 pocket of thrombin. *J. Mol. Biol.* **2009**, *390*, 56.

(S3) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Pratt, D. J.; Wang, L.-Z.; ; Whitfield, H. J. N2-substituted O6-cyclohexylmethylguanine derivatives: Potent inhibitors of cyclin-dependent kinases 1 and 2. *J. Med. Chem.* **2004**, *47*, 3710, PMID: 15239650.

(S4) Liang, J.; Tsui, V.; Abbema, A. V.; Bao, L.; Barrett, K.; Beresini, M.; Berezhkovskiy, L.; Blair, W. S.; Chang, C.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liimatta, M.; Mantik, P.; Menghrajani, K.; Murray, J.; Sambrone, A.; Xiao, Y.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Stanley, M.; Ultsch, M.; Zhang, B.; Wu, L. C.; Magnuson, S. Lead identification of novel and selective TYK2 inhibitors. *Eur. J. Med. Chem.* **2013**, *67*, 175.

(S5) Liang, J.; van Abbema, A.; Balazs, M.; Barrett, K.; Berezhkovsky, L.; Blair, W.; Chang, C.; Delarosa, D.; DeVoss, J.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liu, Y.; Lyssikatos, J.; Mantik, P.; Menghrajani, K.; Murray, J.; Peng, I.; Sambrone, A.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Tsui, V.; Ultsch, M.; Wu, L. C.; Xiao, Y.; Yang, W.; Young, J.; Zhang, B.; Zhu, B.-y.; Magnuson, S. Lead optimization of a 4-aminopyridine benzamide scaffold to identify potent, selective, and orally bioavailable TYK2 Inhibitors. *J. Med. Chem.* **2013**, *56*, 4521, PMID: 23668484.

(S6) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-based optimization of protein tyrosine phosphatase 1B

inhibitors: From the active site to the second phosphotyrosine binding site. *J. Med. Chem.* **2007**, *50*, 4681, PMID: 17705360.

(S7) Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J. Med. Chem.* **2013**, *56*, 15, PMID: 23244564.