

# Validation of the Alchemical Transfer Method for the Estimation of Relative Binding Affinities of Molecular Series

Francesc Sabanés Zariquiey,<sup>†</sup> Adrià Pérez,<sup>‡</sup> Maciej Majewski,<sup>‡</sup> Emilio Gallicchio,<sup>\*,¶</sup> and Gianni De Fabritiis<sup>\*,†,§,||</sup>

<sup>†</sup>*Computational Science Laboratory, Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C Dr. Aiguader 88, 08003, Barcelona, Spain*

<sup>‡</sup>*Acellera Labs, C Dr Trueta 183, 08005, Barcelona, Spain*

<sup>¶</sup>*Department of Chemistry, Brooklyn College of the City University of New York; PhD Program in Chemistry, Graduate Center of the City University of New York; PhD Program in Biochemistry, Graduate Center of the City University of New York, USA*

<sup>§</sup>*Acellera, Devonshire House 582 Honeypot Lane Stanmore Middlesex, HA7 1JS United Kingdom*

<sup>||</sup>*Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain*

E-mail: [egallicchio@brooklyn.cuny.edu](mailto:egallicchio@brooklyn.cuny.edu); [gianni.defabritiis@upf.edu](mailto:gianni.defabritiis@upf.edu)

## Abstract

The accurate prediction of protein-ligand binding affinities is crucial for drug discovery. Alchemical free energy calculations have become a popular tool for this purpose. However, the accuracy and reliability of these methods can vary depending on the methodology. In this study, we evaluate the performance of a relative binding free energy protocol based on the alchemical transfer method (ATM), a novel approach based on a coordinate transformation that swaps the positions of two ligands. The results show that ATM matches the performance of more complex free energy perturbation (FEP) methods in terms of Pearson correlation, but with marginally higher mean absolute errors. This study shows that the ATM method is competitive compared to more traditional methods in speed and accuracy and offers the advantage of being applicable with any potential energy function.

## 1 Introduction

The ability to accurately predict the binding free energy of a ligand to a protein can provide crucial information for drug discovery, as it allows for the identification of compounds that have a higher likelihood of binding to a target. Alchemical free energy calculations have become the leading tools in this field.<sup>1-3</sup> Free energy approaches are especially relevant in hit-to-lead and lead optimization stages of drug design while dealing with a series of similar ligands. Both commercial and free tools for free energy calculations have been developed over the past few years, with extensive use in both academia and the pharmaceutical industry.<sup>4-8</sup>

One of the most common approaches to alchemical calculations is Free Energy Perturbation (FEP). This method involves many distinct equilibrium MD simulations for all states along a  $\lambda$  coordinate that modifies a ligand A into a ligand B alchemically. Commonly these simulations are split between 12 or more  $\lambda$ -

intermediates where the two ligands are interchanged.<sup>9</sup> One of the most common methodologies that use this approach is Schrödinger’s FEP+.<sup>7</sup> Another way to approach alchemical calculations is via Thermodynamic Integration (TI), as in the Amber implementation.<sup>10</sup> The main difference with FEP is that TI calculates the free energy difference by integrating the derivative of the Hamiltonian with respect to the alchemical progress parameter  $\lambda$ . The pmx<sup>5</sup> protocol implements a similar strategy based on non-equilibrium trajectories. Although different, FEP and TI share a few common traits such as the adoption of a double-decoupling process that obtains the relative binding free energy from the difference of the alchemical free energies from separate solution and receptor legs, and the requirement of softcore potentials to avoid clashes and instabilities<sup>11–13</sup>. Custom alchemical topologies and the need for multiple simulations of distinct systems (receptor complex and ligand in solvent) tend to require more user expertise. Furthermore, they are usually not suited for ligand pairs with different net charges<sup>14</sup>, leading to potential issues with the treatment of long-range electrostatic interactions and artifacts in the free energy estimates, unless complex correction factors are introduced.<sup>15</sup>

Recently, a novel approach to performing alchemical calculations has been proposed. The Alchemical Transfer Method (ATM) is a protocol for the estimation of relative binding free energies based on a coordinate transformation that swaps the positions of two ligands. The method performs the calculation in a single solvent box and, unlike double-decoupling free energy perturbation approaches,<sup>7,10,16</sup> avoids the split of the binding free energy calculation into receptor and solvation legs. Furthermore, ATM does not require the implementation of soft-core pair potentials. ATM is implemented in the free and open-source OpenMM<sup>17</sup> molecular simulation package, allowing a simple and easy route to large-scale automated deployments and flexibility to employ any potential energy function. In spirit, ATM is similar to the separated topologies method<sup>18</sup> with the difference that the latter achieves the transfer by decoupling

the first ligand while coupling the second by modifying the force field parameters. Whereas in ATM, the perturbation is implemented as a coordinate displacement that swaps the position of the two ligands. Recently, this approach was reintroduced to be used in GROMACS.<sup>19</sup> ATM can handle both relative (ATM-RBFE) and absolute (ATM-ABFE) binding free energy calculations. In this work we focus on testing the accuracy and feasibility of the RBFE approach.

All the different free energy estimation methods have their pros and cons and can vary in their accuracy and reliability. Therefore it is important to rigorously evaluate their performance against large and diverse benchmark datasets.

In this work, we aim to evaluate the performance of ATM,<sup>20,21</sup> using the dataset of Wang *et al.*,<sup>7</sup> one of the most popular benchmarks for evaluating relative binding free energy protocols. We use ATM to calculate the difference in binding free energies for 330 ligand pairs across 8 different protein systems. We also compared our results with state-of-the-art methodologies such as FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>22</sup>. We show that ATM, a methodology that requires less expertise and preparation than alternative protocols, performs as well as other existing tools and even better from a correlation point of view.

## 2 Methods

The aim of this study is to further expand the benchmarking of ATM to a series of targets tested in other similar methodologies and evaluate whether it can provide accurate and reliable estimates of relative binding free energies for these systems. To address this question, we conducted a computational study in which we applied ATM to the dataset of Wang *et al.*<sup>7</sup> This benchmark includes eight targets relevant to pharmaceutical research (MCL-1, TYK2, MCL-1, JNK1, PTP1B, BACE, Thrombin and p38) with a total of 330 ligand pairs.

ATM is based on a displacement coordinate transformation that swaps the positions of two ligands, one of which is initially placed in

the binding site of the receptor and the other into the solvent bulk.<sup>21</sup> The potential energies of the system before and after the displacement are combined into a  $\lambda$ -dependent potential function, such that the system is progressively transformed from the state in which the first ligand is bound to the receptor and the second is in solution, to the reversed situation in which the second ligand is bound to the receptor and the first is not. ATM protocol does not require soft-core pair potentials or modifications of the energy routines of the molecular dynamics engine, and it does not require splitting the binding free energy calculation into receptor and solvation legs. ATM is implemented as an OpenMM plugin.<sup>23</sup> Further details of the methodology can be found in previous work.<sup>20,21,24</sup>

We used the AToM-OpenMM package<sup>25</sup> to set up and run the alchemical calculations. The AToM-OpenMM workflow (Fig. 1) prepares the complex systems for simulation using the LEaP program in AmberTools19.<sup>26</sup> Amber ff14SB parameters<sup>4,27</sup> were assigned to the receptors while GAFF2/AM1-BCC<sup>28,29</sup> were used for the ligands. Each complex system built in LEaP consists of the receptor and a pair of aligned ligands. One of the ligands is selected to be translated along the diagonal of the solvent box so it is placed outside the receptor, ensuring at least three layers of water molecules in between. A restraining potential is also introduced to maintain geometrical alignment between the two ligands aimed at enhancing the rate of convergence of the free energy estimate. The alignment restraints are based on the relative position and orientation of the coordinate frames of the two ligands defined by three chosen reference atoms.<sup>21</sup> More information on the reference atoms selected for each system can be found in the Supporting Information (Supporting Figure 3).

Each complex system was solvated with a 10 Å solvent buffer and with sufficient sodium and chloride ions to neutralize the system. The solvated complexes are minimized and thermalized at 300 K. Next, the system was annealed from the bound state ( $\lambda = 0$ ) to the symmetric alchemical intermediate ( $\lambda = 0.5$ ) for 250 ps.



Figure 1: The AToM-OpenMM workflow used in this work. (1) Starting from the protein-ligand complexes from the Wang dataset<sup>7</sup> the ligands are parametrised with GAFF2 and protein topologies prepared in the Amber ff4SB forcefield. (2) System complexes are build with Ambertools and ligand B is displaced based on a vector. (3) Energy minimisation and equilibration is performed. Later an annealing and equilibration at  $\lambda=1/2$  is performed. (4) Aysncronous replica exchange simulations are performed until a total sampling of at least 50ns is achieved. (5) After the simulations were finished, these were analyzed with the UWHAM package to obtain the calculated  $\Delta\Delta G$  estimates.

This step facilitates the creation of an initial configuration of the system without strong re-

pulsive interactions at the alchemical intermediate state, which serves as the starting point for the subsequent Hamiltonian replica exchange<sup>24</sup> molecular dynamics that computes the free energies of the two ATM legs connecting the alchemical intermediate to physical end states at  $\lambda = 0$  and  $\lambda = 1$ . C $\alpha$  atoms of the protein receptor were restrained using a flat-bottom harmonic restraining potential with a tolerance of 1.5 Å. Additionally, we apply binding site restraints as a flat-bottom distance restraint between the geometrical centres of sets of receptor atoms that surround the binding site and ligand atoms. This defines the binding site volume as required by the quasi-chemical statistical mechanics formulation of molecular binding.<sup>16</sup> The restraints of the receptor and the ligands are optional and employed here to limit the conformational space that needs to be explored to reach convergence of the binding free energy estimate.

The softplus alchemical potential<sup>30</sup> was used for all calculations with 11  $\lambda$ -states distributed between  $\lambda = 0$  and 0.5 for each of the two ATM legs (Supporting Table 1). AToM-OpenMM performs asynchronous Hamiltonian replica exchanges in  $\lambda$ -space using the method described by Gallicchio *et al.*<sup>24</sup> Exchanges were performed every 10 ps. To maintain a temperature of 300 K, a Langevin thermostat with a time constant of 2 ps was employed. Each ligand pair was simulated for a minimum of 50 ns per  $\Delta\Delta G$  estimate. The sampling time has been chosen in order to be comparable to the aforementioned works, FEP+ studies sample between 36 and 60 ns per  $\Delta\Delta G$  estimate whereas Lee *et al.*<sup>10</sup> employed a total of 48 ns per ligand pair in Amber. In the case of pmx,<sup>22</sup> calculations were carried out for 50 ns per pair for two force fields, GAFF and CGenFF. Since we performed ATM calculations on wall time rather than simulation time, as at the time of performing these calculations there was no support for simulation time-based runs, the sampled simulation time is similar but not identical for all ligand pairs. Binding free energies and their corresponding uncertainties were calculated from the perturbation energy samples using the UWHAM method.<sup>31</sup> The obtained

relative binding free energies ( $\Delta\Delta G$ ) were compared to experimental measurements in terms of the mean absolute error (MAE), root mean square error (RMSE), and Pearson correlation coefficient. The obtained values are compared to the corresponding values reported in the literature.<sup>7,10,22</sup>

The parallel replica exchange alchemical molecular dynamics simulations were performed with the OpenMM 7.7<sup>17</sup> MD engine and the ATM Meta Force plugin<sup>23</sup> using the CUDA platform on NVIDIA RTX 2080 Ti cards.

## 2.1 Results

We conducted a comparison of ATM’s relative binding free energies ( $\Delta\Delta G$ ) estimates for the benchmarking dataset of Wang *et al.*<sup>7</sup> with those of commercial and open-source alchemical approaches: FEP+,<sup>7</sup> Amber,<sup>10</sup> and pmx.<sup>22</sup> In addition to differences in the methodology, these comparisons test variations in protein force fields, ligand parameterization techniques, and differences in the behavior of MD packages as potential sources of deviation in the obtained results. In this study, we have chosen to maintain the parameters described in the previous ATM publications as we believe it would provide a fair and consistent comparison.

The results of the simulations are displayed in Figures 3 and 4, which highlight the relative (Pearson correlation) and absolute (MAE) performance of the method. Table 1 contains a comparison against other free energy methods. Comparison against the other mentioned methodologies can be found in Supporting Figures 1, 2 and 4. We observe that ATM performs similarly to the other approaches in overall Pearson correlation (0.59), with values for specific systems ranging from 0.42 to 0.71. ATM’s Pearson correlation coefficients are particularly good for the MCL-1, JNK1, and Thrombin datasets, where it outperforms the other methods albeit only by relatively small margins. For the other protein targets, Pearson correlation metrics fall within those of the other methodologies. For instance, in the case of p38, the observed correlation is 0.71 for ATM, which is the lowest when compared to the other ap-

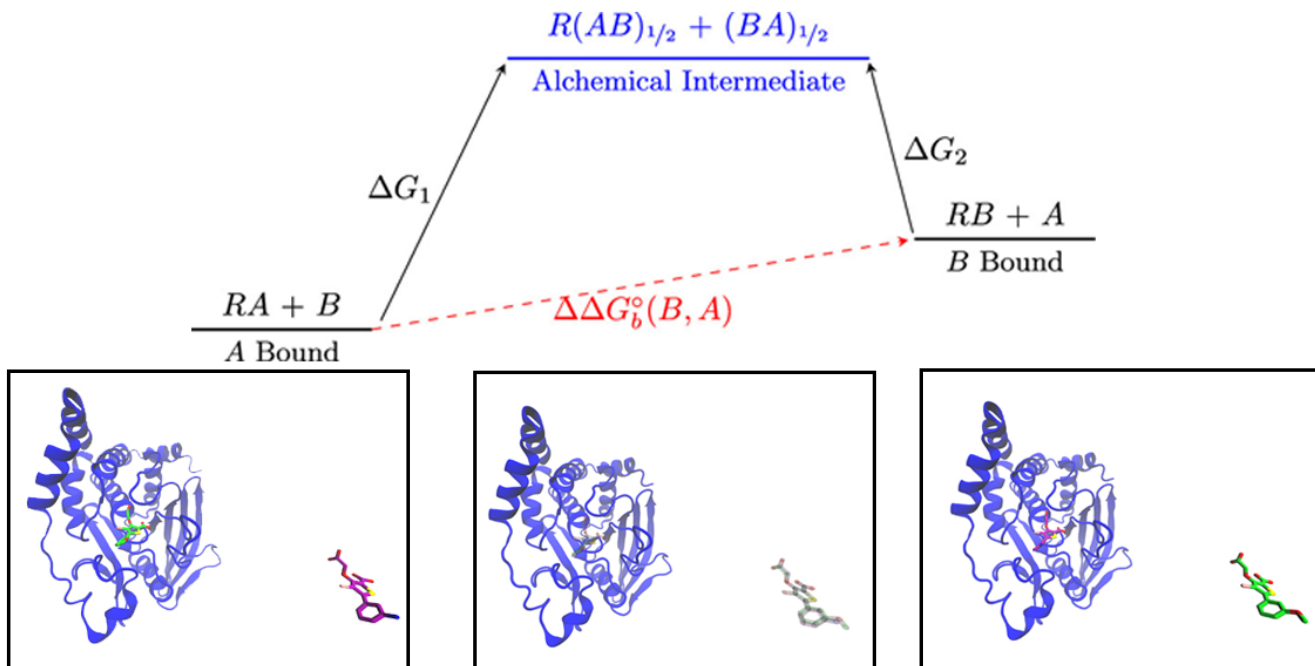


Figure 2: Top: Free energy diagram for an ATM-RBFE calculation, consisting of two independent legs connected to single alchemical intermediate state. The first leg starts at  $\lambda=0$  where ligand A is bound to the receptor R's binding site, and ligand B is present in the solvent bulk. Leg 1 ends at  $\lambda=1/2$  where both ligands A and B are simultaneously present at 50% both in the binding site and in the solvent bulk. The second leg starts with ligand B bound to the binding site and ligand A in the solvent bulk and ends in the same alchemical intermediate. Bottom: Graphical representations for the described events.



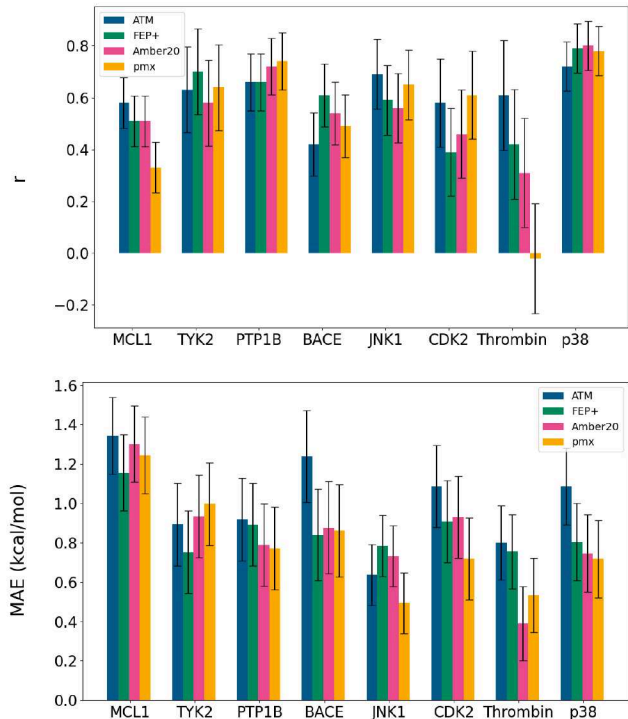


Figure 3: (Top) Pearson correlation ( $r$ ) and (bottom) Mean Absolute Error (MAE) for each protein-ligand system calculated with ATM and reported estimates using the alternative methodologies FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>5</sup>.

proaches. However, the difference is not significant as the results for all methods fall within the measurement error. Despite these positive observations, we did encounter some difficulties with the BACE dataset, as we obtained the lowest correlation (0.42) among the methods which is significantly lower than the best correlation value obtained using the FEP+ methodology (0.61). It is worth mentioning that the range of  $\Delta\Delta G$  values of the BACE pairs is quite narrow and covers only 3.5 kcal/mol, while comparable in-size datasets cover a wider range of values of at least 5 kcal/mol. In effect inaccuracies of the method, as well as experimental measurements, get amplified.

When considering absolute deviations from the experimental references, ATM displayed consistently poorer performance than the other methodologies. ATM’s MAE metric is the highest among the three methods considered in the comparison. However, the difference between ATM and other methodologies is not very

high in most cases. The exception to this is BACE, where, consistent with the previously mentioned results, the differences are the highest.

In terms of convergence, we observed that 50 to 60 ns per  $\Delta\Delta G$  estimate (2.3-2.8 ns per  $\lambda$ ) tends to be sufficient. Convergence analysis over time shows good convergence for most cases as illustrated in Figure (Figure 5). The variance of the predicted  $\Delta\Delta G$  tends to level off for a majority of cases around 50 ns per  $\Delta\Delta G$  estimate. We also performed longer simulations for a series of ligand pairs to evaluate if that extended sampling time was causing any drift on the predicted  $\Delta\Delta G$  (Supporting Figure 12) where we observed that obtained values were stable. Furthermore, we analysed the perturbation energy distributions for every  $\lambda$ -state (Supporting Figures 5,6,7). This analysis helps to determine if the system has converged or will converge in a reasonable amount of time. A poor overlap between perturbation energy distributions are indicative of unreliable relative free energy estimates that could require the design of alternative alchemical routes. We have observed that in some of the ligand pairs that showed poor correlation with both experimental and calculated values from other approaches, there tends to be a poor overlap between perturbation energy distributions of nearby  $\lambda$ -states, even after conducting multiple replicates. From these results we can say that we obtain an analogous convergence performing a similar sampling time than the other mentioned methodologies in this work. The issue of convergence of binding free energy calculations is a very complex topic that we intend to investigate in future work.

We have observed that ATM’s performance metrics are significantly skewed by poor relative binding free energy predictions involving a relatively small number of problematic ligands (Supporting Figure 11). These ligands might be affected by force field parameterization issues or some specific aspects of the ATM methodology. In terms of ligand force fields, OPLS3 was used for FEP+, GAFF2 was used in two of the approaches (ATM and Amber), and a consensus of GAFF2 with CGenFF in pmx. Given that all

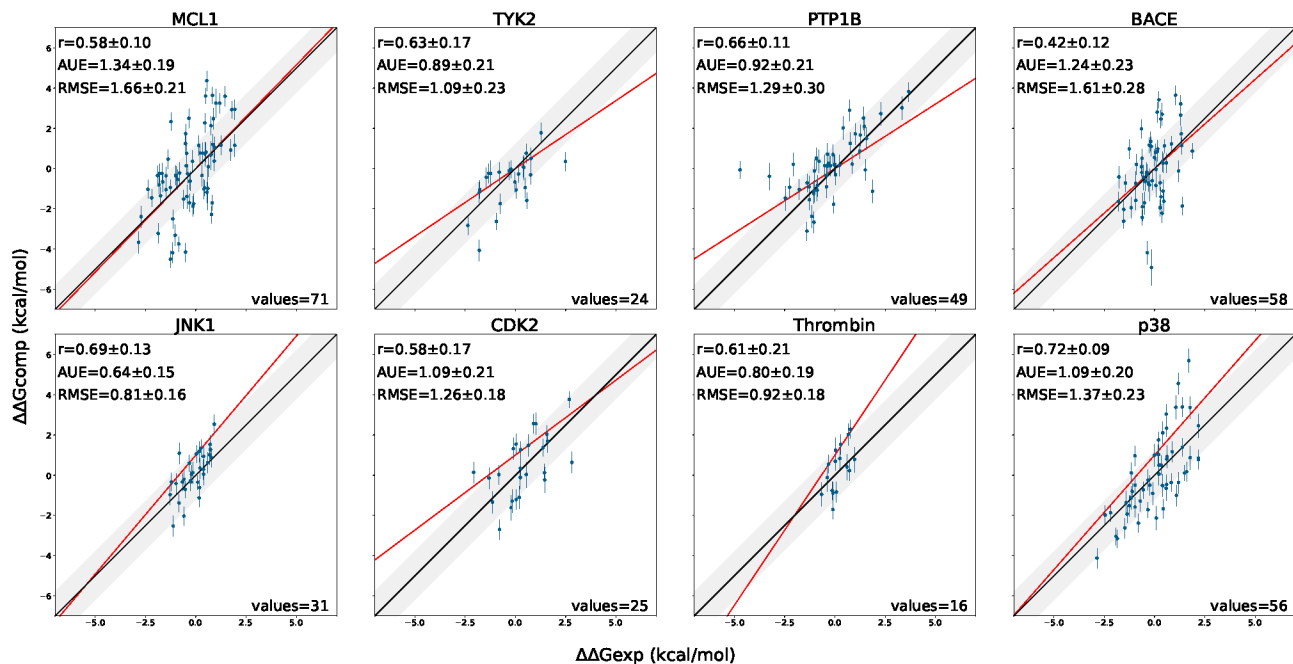


Figure 4: Performance of the Alchemical Transfer Method (ATM) for each protein-ligand system studied. The calculated  $\Delta\Delta G$  estimates are plotted against their corresponding experimental values. MAE is in kcal/mol,  $r$  is Pearson correlation and values refer to the number of ligand pairs evaluated for each system.

Table 1: Comparison of the performance of free energy methods. Pearson correlation ( $r$ ) and Mean Absolute Error (MAE) in kcal/mol for the 8 tested Protein Targets.

	ATM		FEP+ <sup>7</sup>		Amber <sup>10</sup>		pmx <sup>22</sup>	
	r	MAE	r	MAE	r	MAE	r	MAE
MCL1	<b>0.58±0.10</b>	1.7±0.2	0.51±0.10	<b>1.4±0.2</b>	0.51±0.10	1.3±0.2	0.32±0.11	1.2±0.2
TYK2	0.63±0.17	0.9±0.2	<b>0.70±0.15</b>	<b>0.7±0.2</b>	0.58±0.17	0.9±0.2	0.64±0.16	1.0±0.2
JNK1	<b>0.69±0.13</b>	0.6±0.1	0.59±0.15	0.8±0.1	0.59±0.15	0.7±0.2	0.65±0.14	<b>0.5±0.1</b>
PTP1B	0.66±0.11	0.9±0.2	0.66±0.11	0.9±0.2	0.72±0.10	0.8±0.2	<b>0.74±0.10</b>	<b>0.8±0.2</b>
CDK2	0.58±0.17	1.0±0.2	0.39±0.19	0.9±0.2	0.46±0.19	0.9±0.2	<b>0.61±0.17</b>	<b>0.7±0.2</b>
Thrombin	<b>0.61±0.21</b>	0.8±0.2	0.42±0.24	0.8±0.2	0.31±0.25	<b>0.4±0.2</b>	-0.02±0.27	0.5±0.2
p38	0.71±0.09	1.0±0.2	0.79±0.08	0.8±0.2	<b>0.80±0.08</b>	0.8±0.2	0.78±0.08	<b>0.7±0.2</b>
BACE	0.42±0.12	1.2±0.2	<b>0.61±0.11</b>	<b>0.8±0.2</b>	0.54±0.11	0.9±0.2	0.49±0.12	0.9±0.2
ALL	0.59±0.11	1.0±0.1	<b>0.60±0.11</b>	0.9±0.1	0.58±0.11	0.9±0.1	0.56±0.11	<b>0.8±0.1</b>

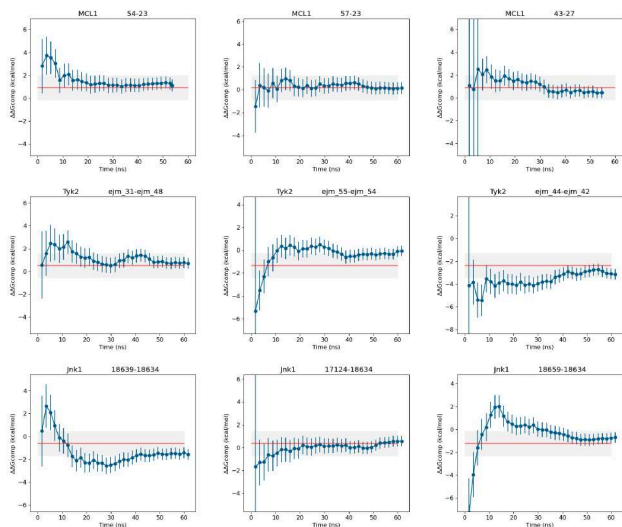


Figure 5: Free energy convergence as a function of time for a series of ligand pairs of MCL-1, TYK2 and JNK1. The red line corresponds to the experimental  $\Delta\Delta G$  value.

methods perform similarly, the accuracy of the forcefield is of the same order as the precision of the methods. In relation to this aspect, Merck published a series of guidelines for FEP calculations with the requirement of an RMSE lower than 1.3 kcal/mol in the validation phase.<sup>32</sup> As we can observe in Table 1 ATM fulfills this requirement for most of the analyzed systems in this study.

One major difference between ATM and the other methods is that ATM models explicitly binding/unbinding processes in the alchemical space. For example, a flexible ligand with different conformational propensities when bound vs when in solution, will undergo an actual conformational transition. In double-decoupling instead, the transformation is applied to the bound and solution conformations individually and a conformational transition is not necessarily required to reach convergence. We believe that this is both a strength and a weakness of ATM. When conformational changes are important, especially if there are differences in the conformational transition between the ligands, ATM is expected to be superior to other methods as it explicitly models the transitions. When considering rigid ligands or ligands with similar transitions, the conformational rearrangements will cancel out and the extra work

ATM needs to do is unnecessary and might hurt convergence. We intend to study these aspects in more detail in future work.

### 3 Conclusion

In this study, we evaluated the performance of the Alchemical Transfer Method (ATM), a novel approach for predicting protein-ligand binding affinities. We benchmarked it against the dataset of Wang *et al.*<sup>7</sup>, one of the most popular data sets on the evaluation of free binding energy methodologies. Our results showed that ATM is a competitive approach for predicting binding affinities, matching or even surpassing the performance of other state-of-the-art methods in terms of Pearson correlation. While mean absolute errors were slightly higher compared to other methods, ATM is a promising approach for the estimation of relative binding free energies.

Unlike other methods, ATM does not require splitting of binding free energy calculations into receptor and solvation legs or the use of softcore pair potentials. Furthermore, ATM is implemented in the open-source OpenMM MD engine, which is freely available. Its flexibility opens up the possibility for further improvement of the method through the use of new force fields, such as neural network potentials.

### 4 Data and software availability

The calculated free energy values, ligand and protein structures as well as preparation scripts are available at: [https://github.com/compsciencelab/ATM\\_benchmark](https://github.com/compsciencelab/ATM_benchmark)

### 5 Acknowledgement

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 823712; and the project PID2020-116564GB-I00 has been funded by MCIN /



AEI / 10.13039/501100011033; the Torres-Quevedo Programme from the Spanish National Agency for Research (PTQ2020-011145 / AEI / 10.13039/501100011033). Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under award number GM140090. E. G. acknowledges support from the United States' National Science Foundation (NSF CAREER 1750511). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## 6 Supporting Information

### Supporting Methods

Table 1: Alchemical schedule of the Solftplus Alchemical Potential for the two legs for the alchemical transformations.  $\alpha$  values are in  $(\text{kcal/mol})^{-1}$  and  $u_0$  and  $w_0$  are in kcal/mol

$\lambda$	$\lambda_1$	$\lambda_2$	$\alpha$	$u_0$	$w_0$
0.00	0.00	0.00	0.10	110	0
0.05	0.00	0.10	0.10	110	0
0.10	0.00	0.20	0.10	110	0
0.15	0.00	0.30	0.10	110	0
0.20	0.00	0.40	0.10	110	0
0.25	0.00	0.50	0.10	110	0
0.30	0.10	0.50	0.10	110	0
0.35	0.20	0.50	0.10	110	0
0.40	0.30	0.50	0.10	110	0
0.45	0.40	0.50	0.10	110	0
0.50	0.50	0.50	0.10	110	0

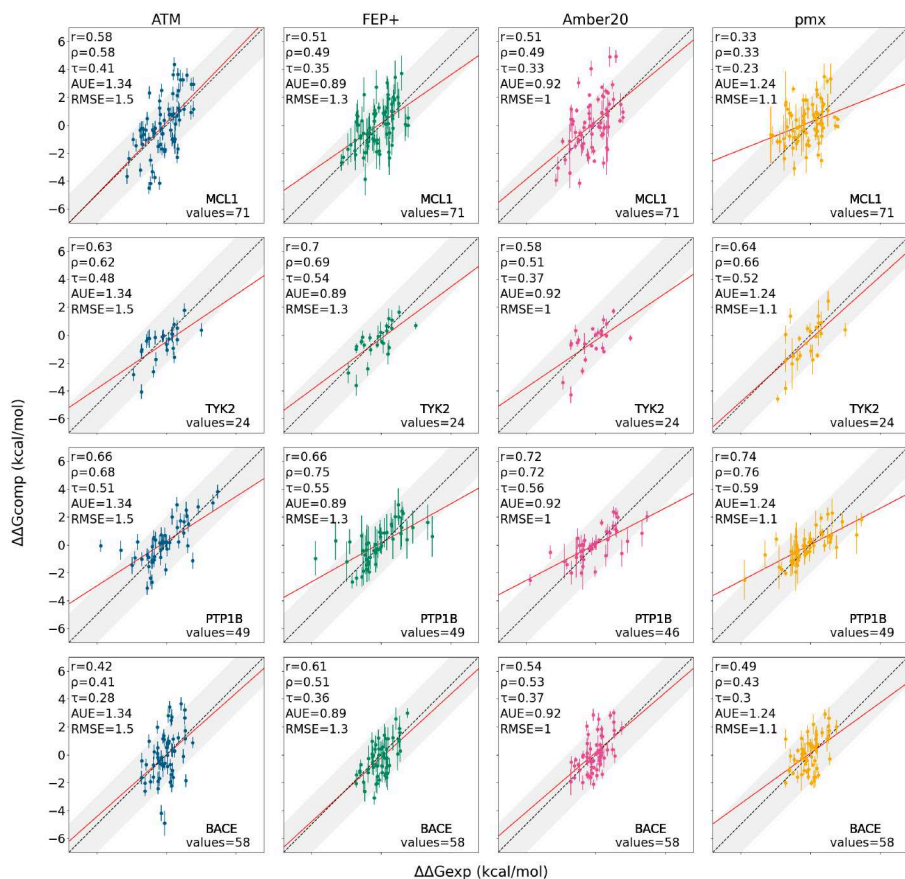


Figure 1: Scatterplots for the calculated  $\Delta\Delta G$  estimated against the experimental measurements and compared to other methodologies for MCL-1, TYK2, PTP1B and BACE systems. The first column represents calculations performed with ATM. The other columns contain data from benchmark studies of FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>22</sup>.

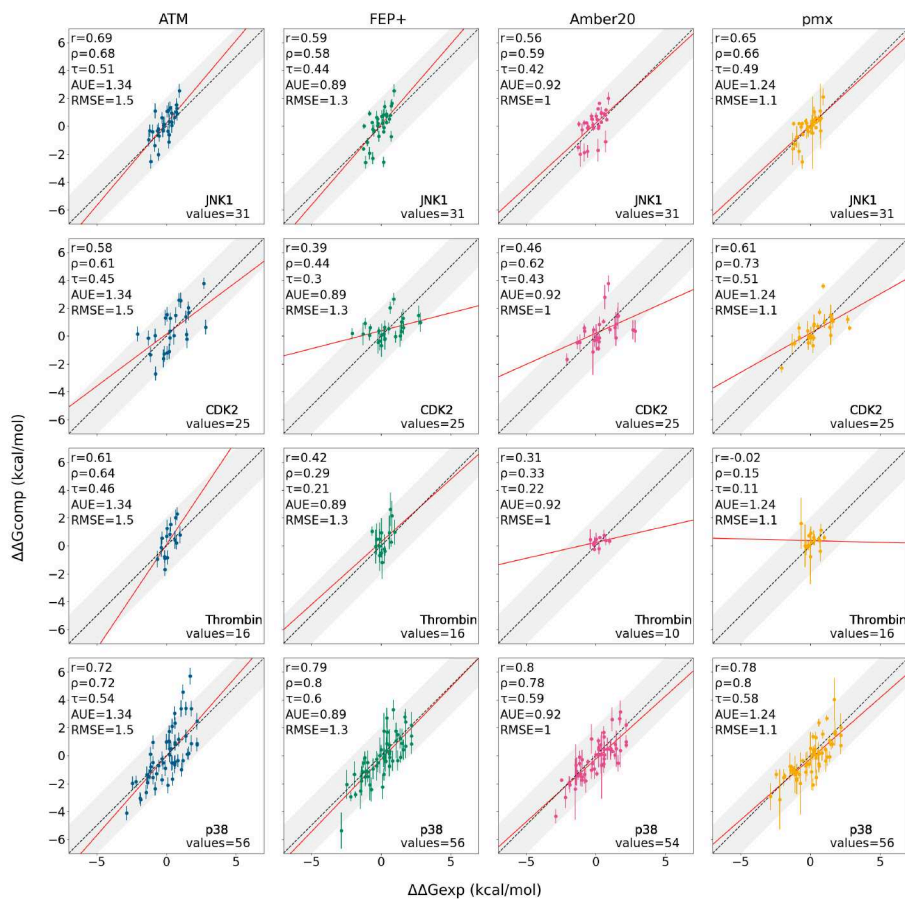


Figure 2: Scatterplots for the calculated  $\Delta\Delta G$  estimated against the experimental measurements and compared to other methodologies for JNK1, CDK2, Thrombin and p38 systems. The first column represents calculations performed with ATM. The other columns contain data from benchmark studies of FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>22</sup>.



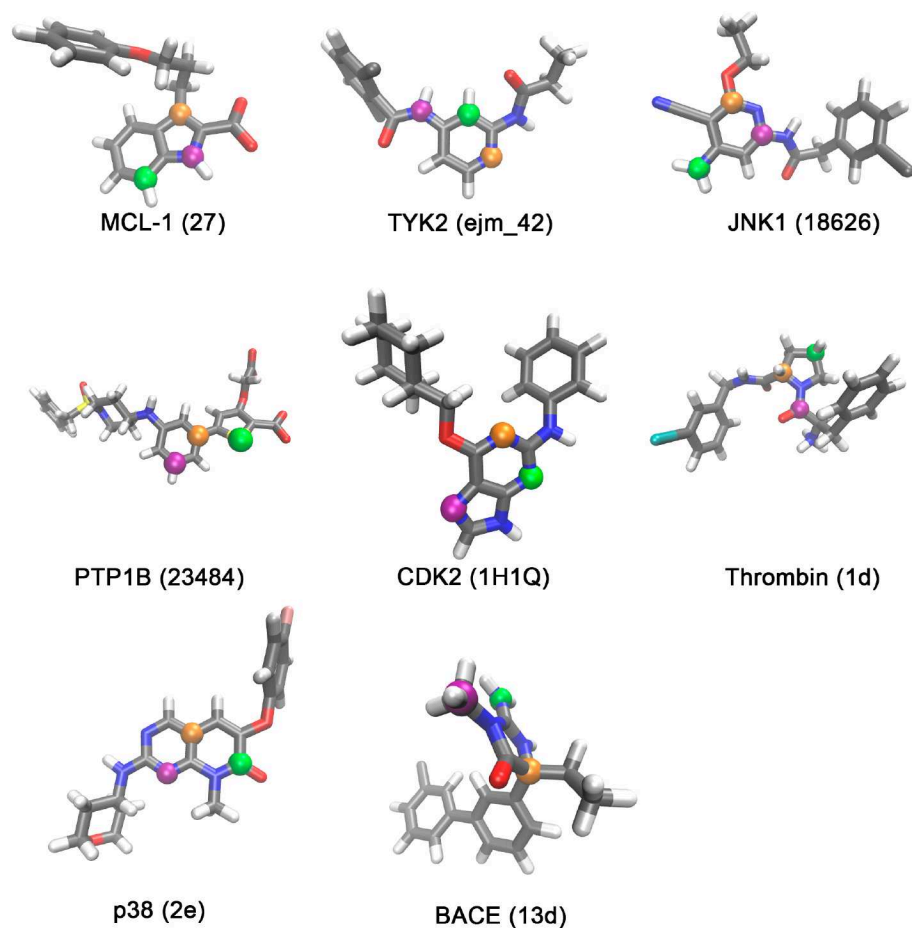


Figure 3: Reference atoms selected for the applied restraining potential in every system. selected atoms of each molecule define a cartesian coordinate system with the orange atom at the origin, a  $z$  axis along the orange to green direction, and the purple atom oriented at the  $xz$  plane

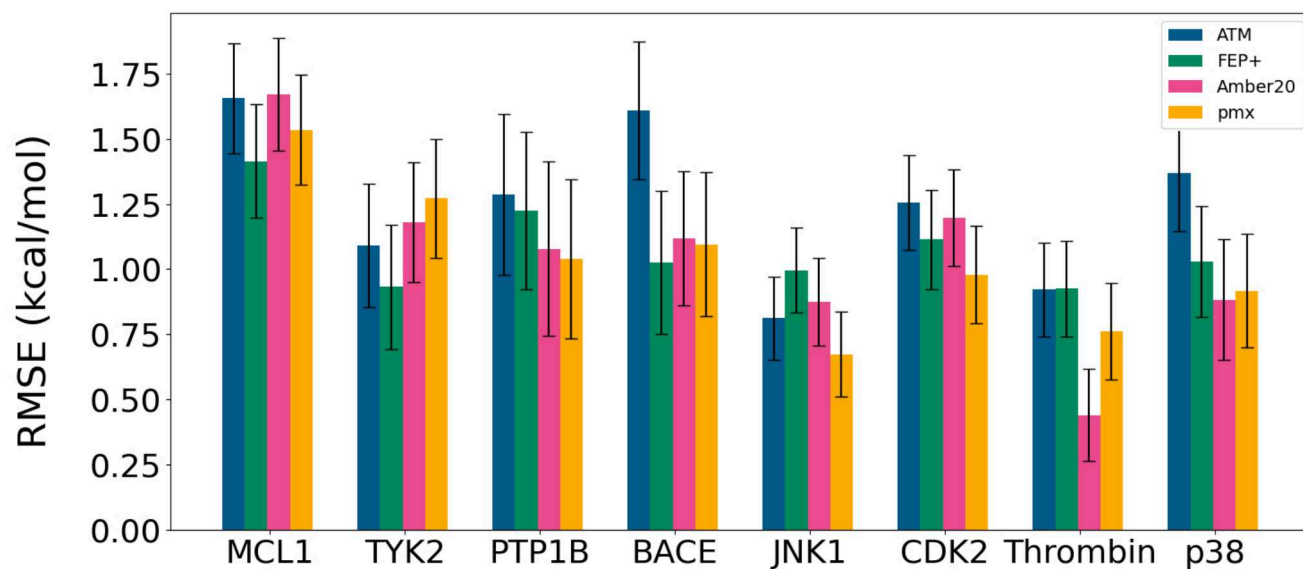


Figure 4: Root Mean Square Error (RMSE) for each protein-ligand system calculated with ATM and reported through other methodologies such as FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>5</sup>.

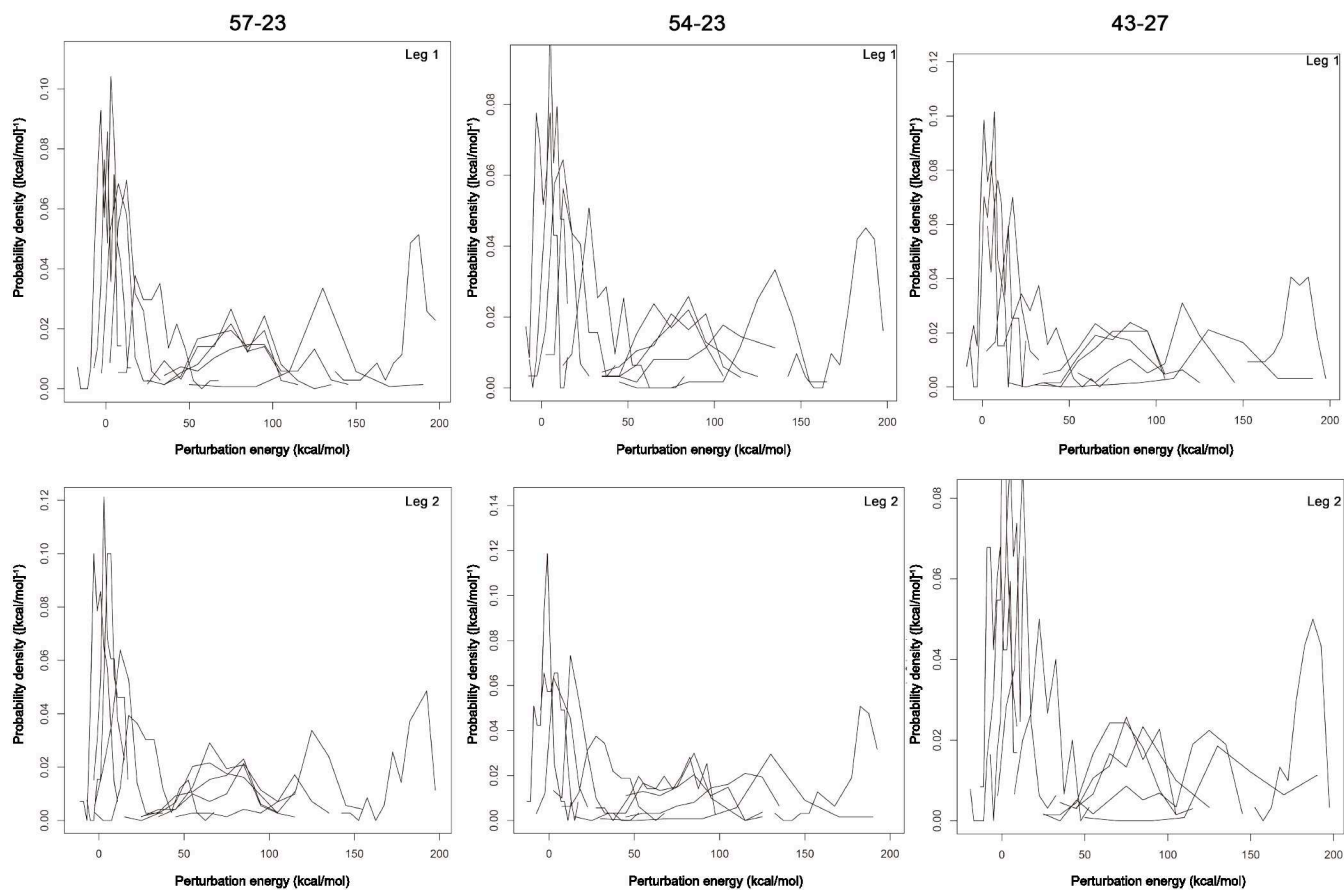


Figure 5: Distribution of the perturbation energies for a series of ligand pairs corresponding to the MCL-1 system. Each line corresponds to a  $\lambda$ . Top row refers to the first leg and bottom row to the second one.

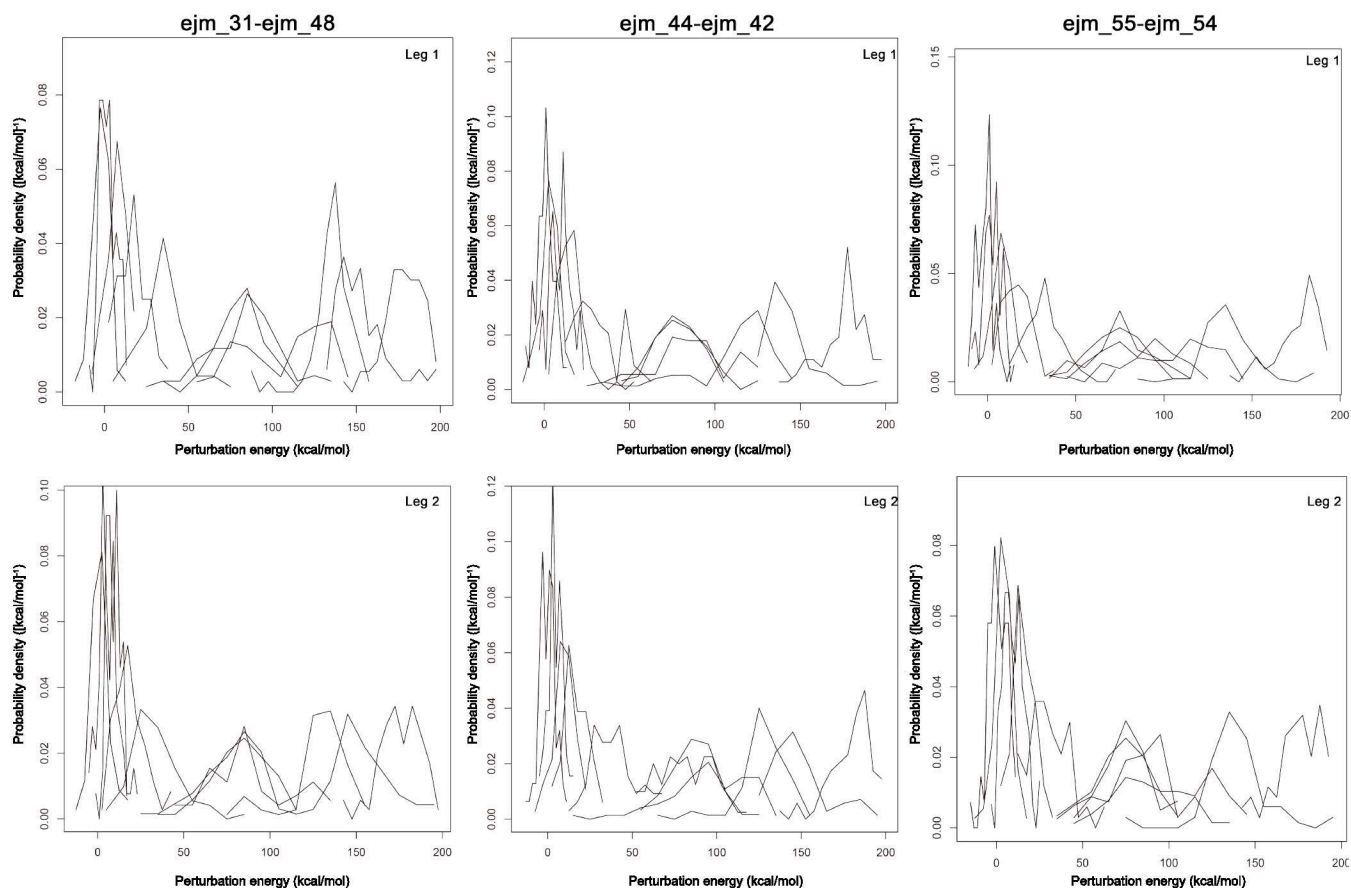


Figure 6: Distribution of the perturbation energies for a series of ligand pairs corresponding to the TYK2 system. Each line corresponds to a  $\lambda$ . Top row refers to the first leg and bottom row to the second one.

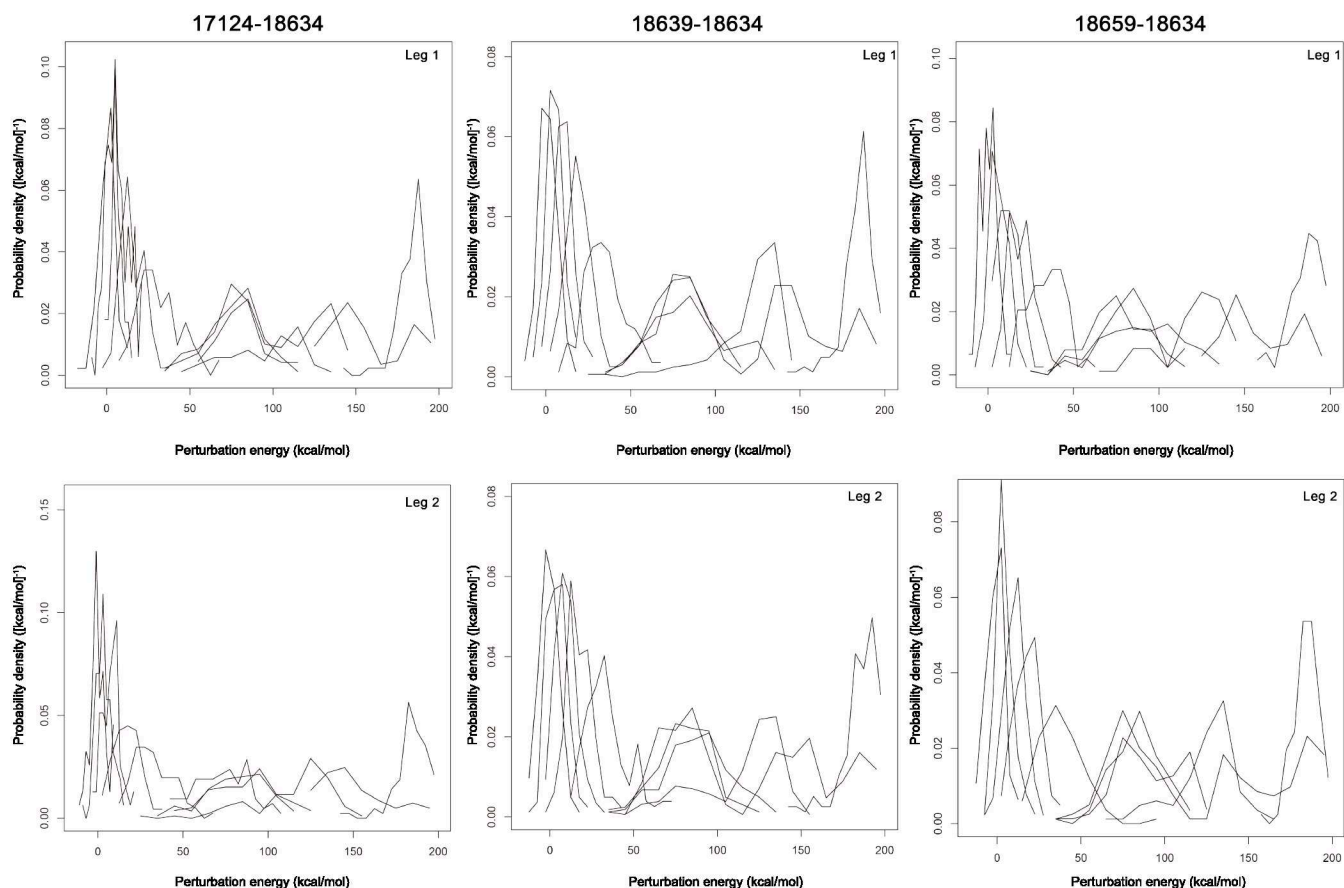


Figure 7: Distribution of the perturbation energies for a series of ligand pairs corresponding to the JNK1 system. Each line corresponds to a  $\lambda$ . Top row refers to the first leg and bottom row to the second one.

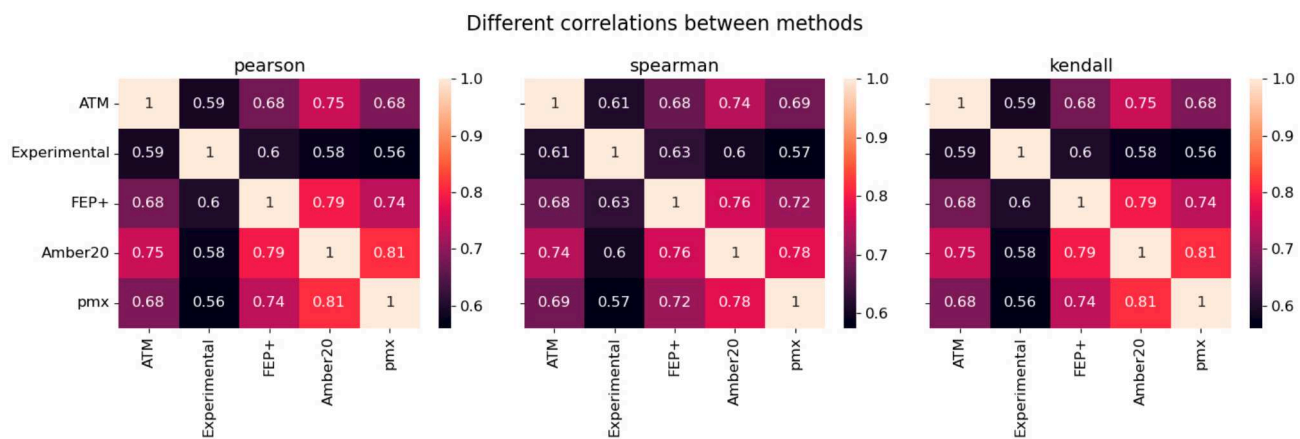


Figure 8: Pearson, spearman and kendall correlations for all data points of ATM against experimental data and the other compared methodologies: FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>22</sup>



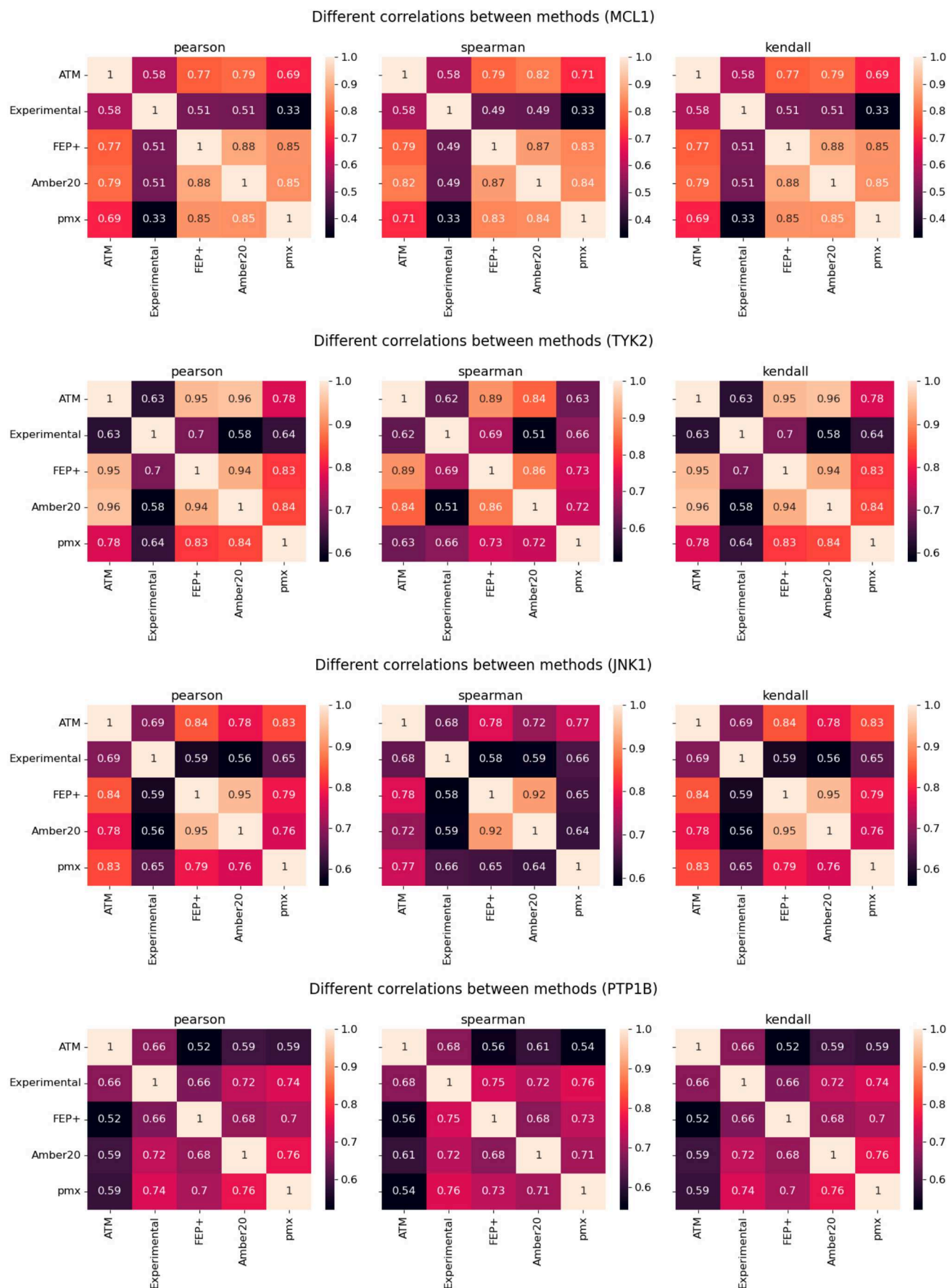
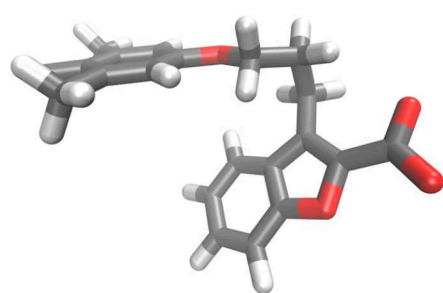


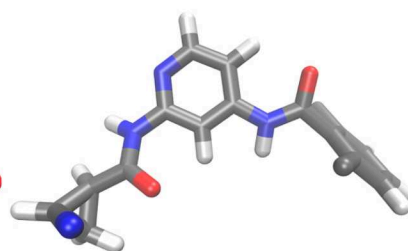
Figure 9: Pearson, spearman and kendall correlations for the MCL1, TYK2, JNK1 and PTP1B systems of ATM against experimental data and the other compared methodologies: FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>22</sup>



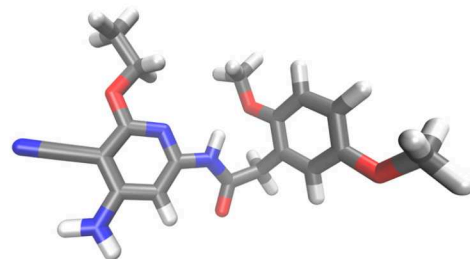
Figure 10: Pearson, spearman and kendall correlations for the BACE, p38, CDK2 and Thrombin systems of ATM against experimental data and the other compared methodologies: FEP+<sup>7</sup>, Amber<sup>10</sup> and pmx<sup>22</sup>



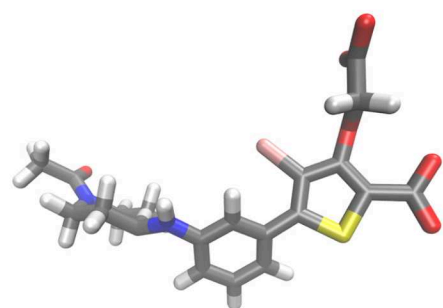
**MCL-1 (67)**



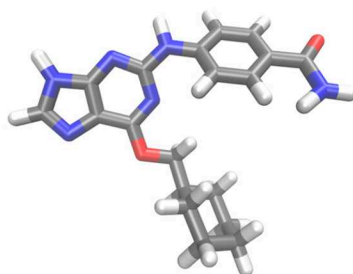
**TYK2 (jmc\_30)**



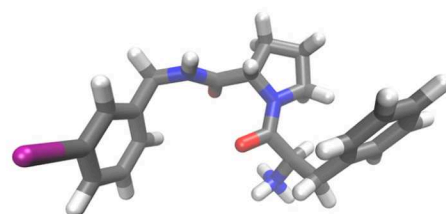
**JNK1 (18634)**



**PTP1B (23479)**



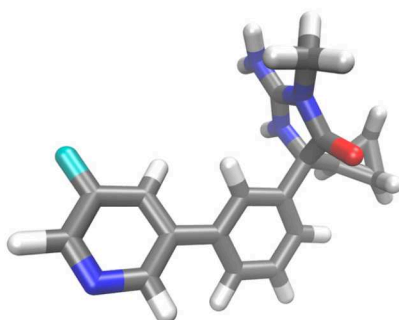
**CDK2 (10IY)**



**Thrombin (1d)**



**p38 (3fln)**



**BACE (17g)**

Figure 11: Structures of some of the ligands that showed poor correlation with experimental values in two or more instances

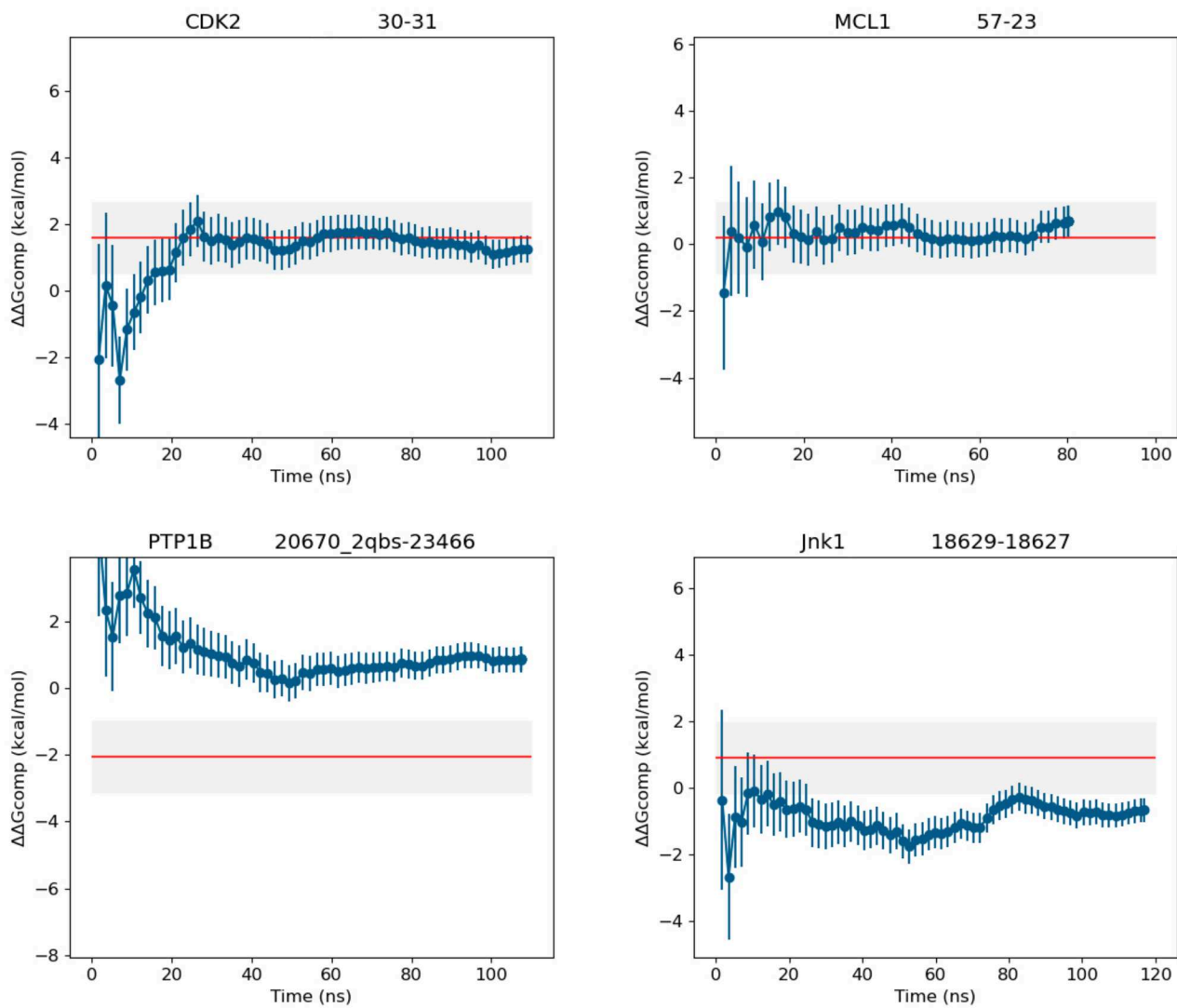


Figure 12: Free energy convergence as a function of time for a series of ligand pairs of CDK2, MCL-1, PTP1B and JNK1 with a greater sampling than 50ns. The red line corresponds to the experimental  $\Delta\Delta G$  value.



## References

- (1) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
- (2) Abel, R.; Wang, L.; Harder, E. D.; Berne, B.; Friesner, R. A. Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632.
- (3) Armacost, K. A.; Riniker, S.; Cournia, Z. Novel Directions in Free Energy Methods and Applications. *J. Chem. Inf. Model.* **2020**, *60*, 1–5, PMID: 31983210.
- (4) Zou, J.; Tian, C.; Simmerling, C. Blinded prediction of protein–ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 1021–1029.
- (5) Gapsys, V.; Michielssens, S.; Seeliger, D.; De Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **2015**, *36*, 348–354.
- (6) Bieniek, M. K.; Bhati, A. P.; Wan, S.; Coveney, P. V. TIES 20: Relative binding free energy with a flexible superimposition algorithm and partial ring morphing. *J. Chem. Theory Comput.* **2021**, *17*, 1250–1265.
- (7) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J., et al. 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*, 2695–2703.
- (8) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S.; Mackey, M.; Michel, J. Assessment of binding affinity via alchemical free-energy calculations. *J. Chem. Inf. Model.* **2020**, *60*, 3120–3130.
- (9) Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378.
- (10) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee Jr, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y., et al. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595–5623.
- (11) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; Van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (12) Zacharias, M.; Straatsma, T.; McCammon, J. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *The Journal of Chemical Physics* **1994**, *100*, 9025–9031.
- (13) Gapsys, V.; Seeliger, D.; de Groot, B. L. New soft-core potential function for molecular dynamics based alchemical free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2373–2382.
- (14) Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate calculation of relative binding free energies between ligands with different net charges. *J. Chem. Theory Comput.* **2018**, *14*, 6346–6358.

- (15) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *The Journal of Chemical Physics* **2013**, *139*, 11B606\_1.
- (16) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047–1069.
- (17) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
- (18) Rocklin, G. J.; Mobley, D. L.; Dill, K. A. Separated topologies—A method for relative binding free energy calculations using orientational restraints. *The Journal of Chemical Physics* **2013**, *138*, 02B614.
- (19) Baumann, H. M.; Dybeck, E.; McClendon, C. L.; Pickard IV, F. C.; Gapsys, V.; Pérez-Benito, L.; Hahn, D. F.; Tresadern, G.; Mathiowetz, A. M.; Mobley, D. L. Broadening the scope of binding free energy calculations using a Separated Topologies approach. **2023**,
- (20) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical transfer approach to absolute binding free energy estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309–3319.
- (21) Azimi, S.; Khuttan, S.; Wu, J. Z.; Pal, R. K.; Gallicchio, E. Relative binding free energy calculations for ligands with diverse scaffolds with the alchemical transfer method. *J. Chem. Inf. Model.* **2022**, *62*, 309–323.
- (22) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Gary, T.; de Groot, B. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. 2020.
- (23) Gallicchio, E. The ATM Meta Force Plugin for OpenMM. <https://github.com/Gallicchio-Lab/openmm-atmmetaforce-plugin>, 2022; (accessed March 16th 2023).
- (24) Gallicchio, E.; Xia, J.; Flynn, W. F.; Zhang, B.; Samlalsingh, S.; Menten, A.; Levy, R. M. Asynchronous replica exchange software for grid and heterogeneous computing. *Comput. Phys. Commun.* **2015**, *196*, 236–246.
- (25) AToM-OpenMM. <https://github.com/Gallicchio-Lab/AToM-OpenMM>, 2022; (accessed March 16th 2023).
- (26) Case, D.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.; Gilson, M.; Gohlke, H.; Götz, A.; Harris, R.; Izadi, S.; Izmailov, S.; Jin, C.; Kasavajhala, K.; Kollman, P. Amber 2019. 2019.
- (27) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (28) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

- (29) He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J. Fast, accurate, and reliable protocols for routine calculations of protein–ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5*, 4611–4619.
- (30) Khuttan, S.; Azimi, S.; Wu, J. Z.; Gallicchio, E. Alchemical transformations for concerted hydration free energy estimation with explicit solvation. *J. Chem. Phys.* **2021**, *154*, 054103.
- (31) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *The Journal of Chemical Physics* **2012**, *136*, 04B608.
- (32) Schindler, C. E.; Baumann, H.; Blum, A.; Bose, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D., et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.