How Well Does Extended Linear Interaction Energy Method Perform in Accurate Binding Free Energy Calculations

Dongxiao Hao,[1,2] Xibing He,[2] Beihong Ji,[2] Shengli Zhang,[1*] Junmei Wang[2,*]

[1]*School of Physics, Xi'an Jiaotong University, Xi'an 710049*

[2]*Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA 15261, USA*

*To whom correspondence should be addressed. Junmei Wang; E-mail:junmei.wang@pitt.edu or to Shengli Zhang; E-mail: zhangsl@xjtu.edu.cn

ABSTRACT

With continually-increased computer power, molecular mechanics force field-based approaches, such as the endpoint methods of MM-PBSA and MM-GBSA have been routinely applied in both drug lead identification and optimization. However, MM-PB/GBSA method is not accurate as the pathway-based alchemical free energy methods, such as thermodynamic integration (TI) or free energy perturbation (FEP). Although the pathway-based methods are more rigorous in theory, they suffer from slow convergence and computational cost. Moreover, choosing adequate perturbation routes is also crucial for the pathway-based methods. Recently, we proposed a new method, coined ELIE (extended linear interaction energy) method, to overcome some disadvantage of MM-PB/GBSA method to improve the accuracy of binding free energy calculation. In this work, we have systematically assessed this approach using in total 229 protein−ligand complexes for eight protein targets. Our results showed that ELIE performed much better than molecular docking and MM-PBSA method in term of root-mean-square error (RMSE), correlation coefficient (R), predictive index (PI), and Kendall's τ. The mean values of PI, R and τ are 0.62, 0.58 and 0.44 for ELIE calculations. We also explored the impact of the

length of simulation, ranging from 1 ns to 100 ns, on the performance of binding free energy calculation. In general, extending simulation length up to 25 ns could significantly improve the performance of ELIE, while longer MD simulation does not always perform better than short MD simulation. Considering both the computational efficiency and achieved accuracy, ELIE is adequate in filling the gap between the efficient docking methods and computationally demanding alchemical free energy methods. Therefore, ELIE provides a practical solution for routine ranking of compounds in lead optimization.

Keywords: Extended linear interaction energy; Lead optimization; Binding Free Energy; MM-PBSA

1. INTRODUCTION

According to a recent survey, on average the cost of an authorized new drug is estimated to be $2.6 billion, and the developing process takes at least 10 years.[1] In order to alleviate the costly and lengthy process of developing new drugs, many expensive and tedious experiments especially in drug lead discovery and lead optimization processes have being replaced by computational methods.[2-8] Quantitative estimation of the ligand–protein interactions is the foundation of computer-aided drug design (CADD).[9] During the past three decades, various computational methods have been developed to predict protein−ligand binding affinities and achieved different levels of successes. For instance, the docking and scoring methods were developed based on simple molecular mechanics or empirical energy terms or knowledge-based statistical potentials,[10] which could discriminate potentially bioactive ligands from reservoirs of huge amount of candidate compounds and predict most likely ligand binding modes in a defined binding pocket.[11, 12] However, its binding free energies are not of high accuracy, despite with high computational efficiency and low cost.[13,14] On the contrary, alchemical free energy (AFE) methods are theoretically rigorous and of high accuracy.[15] Such as thermodynamic integration (TI), and free energy perturbation (FEP) could generate effective guidance for drug lead optimization.[16] However, the slow convergence of the free energy differences result in huge computational cost and limit its practical applications.[17]

In recent years, physics-based endpoint methods, such as MM-PBSA (molecular mechanics Poisson−Boltzmann surface area) and MM-GBSA (molecular mechanics generalized Born surface area) have been widely used on CADD since those methods achieve a good balance between computational efficiency and accuracy.[18-21] They have been widely used to detect hotspot residues, evaluate docking poses, and further shrink the pool of promising compounds identified using docking methods. However, endpoint approximation methods may not be accurate enough for the drug lead optimization, as the range of binding affinities of compounds with small substitution differences usually spans only 3−4 kcal/mol or less.[22] Another well-

known endpoint approximation method, which also has performance and computational efficiency between docking and FEP/TI methods, is linear interaction energy (LIE).[23] Traditionally, LIE only considers the interaction energies from Coulomb interaction and van der Waals interactions, and apply optimized scaling factors on these two energy terms. And then other variants of LIE with additional terms were developed. For example, Jorgensen et al. proposed to add a third term based on either solvent accessible surface area (SASA) or the cavity energy in a continuum solvent surface generalized Born (SGB) model.[24] In a previous study we proposed a hybrid method, coined extended linear interaction energy (ELIE), which combines MM-PBSA and LIE.[25] This ELIE method was successfully applied to predict binding free energies for Cathepsin S drug target in the third D3R Grand Challenge of blind prediction of protein-ligand binding affinities.[25]

In this work, we conducted a systematic assessment on ELIE method using eight drug targets, namely, Beta-Secretase 1 (BACE1), cyclin-dependent kinase 2 (CDK2), myeloid cell leukemia sequence 1 (MCL1), Thrombin, Tyrosine kinase 2 (TYK2), C-Jun N-terminal kinase 1 (JNK1), P38, and Protein Tyrosine Phosphatase 1B (PTP1B). Each protein system involved a series of congeneric ligands that are structurally similar and only have minor substitution differences. Such scenarios are typical at the stage of drug lead optimization. These eight drug targets were firstly chosen by Wang et al. to demonstrate the capability of the state-of-the-art FEP+ module in the popular commercial Schrodinger Suite software,[16] which is now the de facto standard in pharmaceutical industry. Very recently our lab published a study on the four more difficult systems (BACE1, CDK2, MCL1, PTP1B) among these eight protein systems with GPU-TI module in the academic AMBER program.[22] In this study we compared the performance of protein-ligand binding affinity calculation using various methods, including Glide Docking[26] FEP/TI,[16, 22] regular MM-PBSA and ELIE. We also explored the impact of length of simulation time on the accuracy of both MM-PBSA and ELIE calculations. With those efforts, we hope to establish a common practical protocol of applying ELIE to guide drug lead optimization, which

can achieve a comparable performance to that of pathway-based free energy methods, but with reduced computer resource and time.

2. MATERIALS AND METHODS

2.1. Data Set Preparation

Eight different protein systems studied in this work, including BACE1 (41 ligands),[27] CDK2 (22 ligands),[28] MCL1 (44 ligands),[29] Thrombin (23 ligands),[30] TYK2 (16 ligands),[31] JNK1 (21 ligands),[32] P38 (35 ligands),[33] and PTP1B (27 ligands),[34] have also been studied by pathway-based free energy methods, such as FEP and TI [16, 22]. The chemical structures of ligands for the eight systems were taken from corresponding experimental studies,[27-34] and experimental binding data affinity was calculated using the experimental K_i or IC_{50}.[16] The ranges of K_i or IC_{50} are only two to three orders of magnitude for each protein system. The narrow ranges of binding free energies pose a challenge on ranking those ligands by predicting their binding free energies. However, narrow K_i range is common in drug lead optimization for most drug discovery projects.

2.2. Molecular Docking

The ligands were docked into corresponding protein receptors using Glide Docking[26] implemented in Maestro.[35] The Glide grid file of each receptor was generated with the default values of the van der Waals radius scaling factor and the partial charge. The center of the binding grids was located at the geometric center of the bound ligand, and no constraint or rotatable group was defined. The receptor-based docking was conducted using default settings of Glide plus a reward of intramolecular hydrogen bond formation. The best docking pose for each ligand was selected according to the docking scores.

2.3 Molecular Simulation System Setup

For all the 229 ligands, the atomic partial charges were derived to reproduce the electrostatic potential calculated at the HF/6-31G* level using the RESP (restrained ESP) program.[36] The geometries were optimized at the same level using the Gaussian 16 package.[37] Other force field

parameters of ligands came from GAFF 1.8.[38] The residue topology and force field parameter file were generated using the Antechamber module[39] implemented in AMBER18 software package.[40] The protein receptors were described using the ff14SB force field.[41] The protonation states of the histidine residues are presented in Table S1, which were manually checked and assigned. Since at least one ligand in each of the eight protein systems has crystal structure available and the ligands of each protein system share common scaffold,[27-34] for each protein-ligand complex, a top docking pose which is also close to the binding mode of a native ligand of the same protein system was adopted as the initial configuration for subsequent MD setup and simulation. For each protein-ligand complex, it was solvated in a rectangular TIP3P[42] water box, with the edges of the water box were at least 12 Å away from any protein and ligand atoms. Each protein-ligand complex systems was neutralized with counter ions Na+ or Cl- and extra 0.1 M NaCl ions were added.

2.4. MD simulations

Initially, each MD system was minimized using the steepest-descent (SD) and conjugate gradient (CG) algorithms to remove possible clashes. We adopted a stepwise minimization protocol to gradually relax the system: a 1000-step SD minimization followed by a 1000-step CG minimization were performed with the heavy atoms been restrained using a force constant of 2.0 kcal/mol/Å^2, then the system was relaxed by a 1000-step SD followed by a 1500-step CG full minimizations. The system was then heated from 0 K to 298 K with 2.0 kcal/mol/Å^2 restraints on all heavy atoms in the complex in 50,000 steps. Next, the MD system was further relaxed by 50,000-step restrained MD simulation with the restraint force constant of 2.0 kcal/mol/Å^2.

In the subsequent equilibration phase, the MD system underwent a 1 ns constant pressure MD simulation to produce an NPT ensemble at 298 K. In the sampling phase, a 99-ns MD simulation was performed and snapshots were recorded at a frequency of 100 ps to ensure that sequential snapshots were independent, resulting 990 snapshots for post-analysis. The following are the key settings applied to all the MD simulations: the temperature and pressure were

controlled by Langevin dynamics temperature scheme[43] and isotropic position-scaling algorithm,[44] respectively; a cut-off of 12 Å was employed to generated non-bonded list for calculating the electrostatic and van der Waals interactions; Particle Mesh Ewald (PME) was used for the accurate description of the long-range electrostatic interaction; and the integration of movement is conducted every 2 fs. All the minimization and MD simulations were performed by using the Pmemd program in AMBER18 software package.[40]

2.5. MM-PBSA-WSAS and ELIE Method

The principle of the MM-PBSA-WSAS method has been well described in many references.[18, 45] The MM-PBSA binding free energy for a ligand binding to a receptor to form a complex is expressed as:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{SOL} - T\Delta S \qquad (1)$$

where ΔE_{MM} is the change of molecular mechanics (MM) energy due to complexation in gas-phase, ΔG_{sol} is the change of solvation free energy, and $-T\Delta S$ is the change of conformational entropy upon ligand binding. E_{MM}, is composed of several energy terms including internal energies E_{int} (bond, angle, and dihedral energies), electrostatic energy E_{ele} and van der Waals interaction energy E_{VDW}. ΔG_{sol} can be further decomposed into two parts, the polar contribution (electrostatic solvation energy) which is described by the Poisson-Boltzmann continuum solvation model, and the nonpolar contribution which is described by solvent accessible surface area (SASA). In this work, the value of the exterior dielectric constant was set to 80, and the solute dielectric constant was set to 1. As such, the MM-PBSA binding free energy has five energy terms as shown below when the "Single Trajectory" sampling protocol[46] is applied:

$$\Delta G_{bind} = \Delta E_{VDW} + \Delta E_{elec} + \Delta G_{solv\_pl} + \Delta G_{solv\_np} - T\Delta S \qquad (2)$$

Please note that in Eq. 2, the contribution from internal energies E_{int} is cancelled out in the "Single Trajectory" protocol.[46] In this work, the polar contribution of the solvation free energy is calculated by solving the finite-difference Possion-Boltzmann equation using Delphi 95

software.[47] The nonpolar component is calculated using a linear relationship to SASA, i.e. $\Delta G_{solv\_np} = \gamma\ \Delta SASA + b$, where γ, the surface tension, is set to 0.00542 (kcal/mol/Å^2) and b to 0.92 kcal/mol.[48] The conformational entropy term is calculated using the WSAS program, which has achieved a comparable accuracy as normal mode analysis, but with substantially reduced computational cost.[49]

The original LIE approach only scales the electrostatic and van der Waals interaction energies to fit the experimental values[23]. While in the ELIE model described by Eq. (3), five terms, namely, ΔE_{VDW}, ΔE_{elec}, ΔG_{solv_pl}, ΔG_{solv_np} and $-T\Delta S$ participate restricted fitting.[25]

$$\Delta G_{bind} = c0 + c1 \times \Delta E_{VDW} + c2 \times (\Delta E_{elec} + \Delta G_{solv\_np}) + c3 \times \Delta G_{solv\_np} - c4 \times T\Delta S \quad (3)$$

Where c1, c2, c3, c4 are scaling coefficients, which need to be adjusted by fitting calculated binding free energies to experimental values. We call this procedure a restricted fitting as the weights must have a physical meaning, i.e., all the weights of the energy terms must be positive. Once optimum values are obtained for these scaling coefficients, an ELIE model is created and it can be used to predict the binding free energies of other compounds for the same receptor. Considering different drug targets have different chemical environment at the binding sites, we developed ELIE models for individual drug targets.

3. RESULTS

3.1 Performance of Docking for the ranking prediction

Glide Docking [26] was employed to predict the binding poses, and the best predicted binding pose was used for further analysis. The predicted binding free energies (docking scores) of each ligand with relative receptors are shown in Figure 1. The performance of the docking method was evaluated using root-mean-square error (RMSE), Pearson's correlation coefficient (R) between the predicted and the experimental binding free energies, Kendall's tau (τ), and the predictive index (PI).[50,51] PI is widely used in virtual screening (VS) to evaluate a scoring function or method's ranking power. The RMSE, PI, R, and τ between the experimental binding free energies and the predicted values are summarized in Table 1. It is shown that the performance of Glide docking varies significantly depending on target proteins: there is strong correlation between docking scores and measured activities for TYK2 (PI= 0.85, R= 0.75, τ = 0.65) and PTP1B (PI = 0.72, R = 0.68, τ = 0.52); no correlation for JNK1 (PI = -0.13, R = -0.07, τ = -0.08), CDK2 (PI = 0.25, R = -0.22, τ = -0.14), BACE1 (PI = 0.03, R = 0.04, τ = 0); and mild correlation for MCL1 (PI = 0.42, R = 0.42, τ = 0.28). Overall, the correlation between docking scores and experimental binding free energies is poor for most drug targets.

3.2 Performance of MM-PBSA-WSAS for the ranking prediction

We then calculated the binding free energies using MM-PBSA-WSAS model after protein-ligand complexes underwent 100 ns MD simulations. The detailed binding free energies were listed in Tables S2 to S9. For each complex, the binding free energy was averaged over 195 snapshots from 1 ns to 100 ns. The calculated binding free energies of ligands against the experimental values for the eight systems are shown in Figure 2. As shown in Table 1, the performance of the canonical MM-PBSA-WSAS in ranking protein-ligand binding is similar to that of Glide docking: the mean PI, R and τ values of the eight systems, 0.31, 0.29 and 0.21, are close to the mean values of Glide docking, 0.34 0.27, and 0.21, correspondingly. However, the RMSE of MM-PBSA-

WSAS (2.59 kcal/mol) is higher than that of Glide docking (1.04 kcal/mol). As to the correlation and ranking metrics (PI, R and τ) for individual systems, MM-PBSA-WSAS performs much better than that of docking for Thrombin and CDK2; and it achieved relatively satisfactory performance for TYK2. However, MM-PBSA-WSAS did not perform well for the other systems. It is obvious that the performance of MM-PBSA-WSAS is also receptor-dependent. A key parameter that has a great impact on MM-PBSA-WSAS calculation is the intrinsic dielectric constant in calculating the polar component of solvation free energy using a PB model. Recently, Wang et al. suggested to apply a variable dielectric GB model for predicting protein-ligand binding affinity.[52] It is expected that the same strategy can be applied to a PB model.

3.3 Performance of ELIE for the ranking prediction

Although applying variable dielectrics can further improve the performance of MM-PBSA-WSAS, it is technically more difficult to achieve. On the other hand, ELIE method can much better improve the ranking power with simply restricted fitting. We fitted the ELIE weighting coefficients in Eq. 3 using the experimental data. The parameters and detailed energies were listed in Tables S2 to S9. The calculated binding free energies of ligands against the experimental values are shown in Figure 3. The performance of ligand ranking for the eight systems is summarized in Table 1. With ELIE method, the mean values of PI, R and τ are 0.62, 0.58, and 0.44, respectively, while the mean value of R is 0.74 for FEP calculations,[16] 0.63 for TI calculations.[25] It is encouraging that six out of eight systems have PI values better than 0.6, and the other two have acceptable PI values (~0.42 for BACE and MCL1). The mean RMSE values for ELIE, 0.94 kcal/mol, was slightly less than the RMSE for Glide docking (1.04 kcal/mol), and was much smaller than the RMSE for regular MM-PBSA-WSAS (2.59 kcal/mol). It is interesting to observe that ELIE achieves a good prediction for JNK1 (PI = 0.61, R = 0.60, τ = 0.54), in which both docking (PI = -0.13, R = -0.07, τ = -0.08) and MM-PBSA-WSAS (PI = 0.04, R = 0.04, τ = 0.05) perform very badly.

In principle, ELIE method needs to fit the scaling coefficients from a training set and apply them

to a test set. Due to the very limited size of ligands in the eight protein systems, aforementioned performance of ELIE was based on fitting on full ligand sets. To further estimate of the true performance of ELIE, we randomly selected 5 ligands to enter a test set and the rest enter the training/calibration set for BACE1, MCL1 and P38 systems which have relatively abundant ligands. To eliminate bias, such splitting process was conducted 1000 times for each protein system. The PIs of calibration and test sets for both MMPBSA and ELIE were calculated. The results are shown in Figure 4. For each system, the PI values for the test sets were normally distributed, with a similar median PI value and relatively larger variance compared to the PI values of the training sets, probably due to the size of the test set is quite small. Similar result was found for PI calculated using MMPBSA. As expected, for both calibration set and test set, the ELIE performs better than MMPBSA for all the three protein systems.

As a conclusion, the ELIE method has dramatically improved the predictive index PI and correlation efficient R compared to MM-PBSA-WSAS and even achieved reasonable accuracy for predicting binding free energies comparable to alchemical free energy calculations.

3.4 Impact of the Length of MD Simulations

The impact of MD simulation protocol on MM-PB/GBSA calculation has been studied by Sun et al.[53] Here, we focused on investigating how the length of MD simulations affects the binding free energy calculation with MM-PBSA-WSAS and ELIE. To this end, a comparative study was conducted using 20 different lengths of MD simulations from 1ns to 100ns. The PI, R and τ between the experimental and predicted binding free energies derived from different simulation time are presented in Figure 5. As to MM-PBSA-WSAS, for Thrombin, CDK2 and BACE, extending the simulation time could significantly improve the ranking power predicted by MM-PBSA-WSAS (green lines in Figure 5), up to 30 ns for BACE1, 15 ns for Thrombin and 70 ns for CDK2. For PTP1B, TYK2, P38 and MCL1, extending the simulation time could not improve the ranking performance of MM-PBSA-WSAS (green lines in Figure 5), and the MM-PBSA-WSAS predictions based on relatively short simulation snapshots are slightly better than those based on

longer MD simulations snapshots. It is of note that the shortest MD simulation time considered is 1 ns. It is interesting to observe that the MM-PBSA-WSAS predictions based on 1-5 ns snapshots are significantly better than those based on longer MD simulations snapshots for JNK1.

Similar to MM-PBSA-WSAS predictions, extending the simulation time does not always improve the correlations between the ELIE predicted binding free energies and the experimental values (blue lines in Figure 5). Additionally, we use the scaling coefficients of 1-100 ns to recalculate the PI and R for ELIE predictions (red lines in Figure 5). It was comparable to the performance using scaling coefficients calculated from different simulation length snapshots. The corresponding curves for the averages over all eight systems is also plotted. As shown in Figure S1, 25 ns MD simulation is enough for ELIE binding free energy calculations, and the results was also summarized in Table 1. The mean values of PI, R and τ for 25 ns ELIE calculations (0.59, 0.56, and 0.43, respectively) are close to those values for 100 ns ELIE calculations. The observation that ELIE scaling parameters are independent of MD simulation length is an excellent feature of the ELIE method.

4. DISCUSSION

Computational methods are attractive for drug lead identification and lead optimization. Both Docking methods and MMPBSA could give good predictions for compounds with a wide range of binding free energies. Especially, docking methods could screen potentially bioactive ligands from reservoirs of a huge amount of candidate compounds and are adequate in predicting the correct poses for many drug targets. On the other hand, for shrinking the pool of ligands with a wider range of binding free energy, the MMPBSA is generally more accurate than docking methods, but with the price of much more computing than docking methods. However, in lead optimization phase, the range of experimental binding free energies of compound evolved from the same drug lead is usually very narrow, posing a challenging for the two aforementioned methods to accurately rank those compounds. As indicated in Table 1 and Figures 1-2, the

performances of the two methods in ranking structurally similar compounds vary from system to system. Overall, neither of them is accurate enough to guide lead optimization in practice. Alchemical free energy methods, such as FEP and TI, could achieve an acceptable accuracy to guide lead optimization, however, they are usually very computer-resource and time demanding. Moreover, the system setting up and post-analysis of alchemical free energy calculations is a daunting task for users naïve to molecular modeling.

ELIE is an ideal method to guide drug lead optimization, as long as binding affinities of sample ligands (training set) for the specific receptor target have been measured in this drug development phase from which an ELIE model can be established. This scenario is common in real drug discovery projects. Moreover, our work has indicated that ELIE can achieve performance on ranking structurally similar compounds comparable to FEP and TI using eight drug targets. Our previous work also indicated that ELIE method could significantly improve the correlation between estimated and experimentally values of compounds with wider range binding affinity.[25] To further strengthen the applicability of ELIE in the application of lead optimization stage, we explored how MD simulation length, a key parameter of sampling protocol, affect the ranking power of ELIE, and came out practice guidance on binding free energy calculations with ELIE, i.e. one may conduct ELIE analysis using MD snapshots sampled from 1 to 25 ns for a typical drug target, and make adjustment on sampling protocol whenever necessary. As for TI calculations of one ligand, it costs approximately 195.8, 173.3, 148.5, and 184.5 hours GPU time for BACE1, CDK2, MCL1 and PTP1B, respectively.[22] As for 100-ns MD simulation of one ligand, it costs approximately 53.3, 53.1, 26.1 and 50.3 hours GPU time for BACE1, CDK2, MCL1 and PTP1B, respectively. As for 100-ns MM-PBSA-WSAS calculations, it costs approximately 6.6, 6.6, 3.2 and 6.5 hours CPU time for BACE1, CDK2, MCL1 and PTP1B, respectively. Obviously, 25-ns MMPBSA calculation is highly efficient comprared to TI calculations. However, we should not expect too much for the transferbility and generality of the scaling parameters of LIE and ELIE methods. As the Zhou and Jorgensen suggested that we have

to sacrifice generality to obtain the high accuracy.[24] Thus, the parameters should be developed for the specific receptor, which is especially appropriate for drug lead optimization whose ligands have similar binding modes. It is interesting to note that the the ELIE parameters of BACE1 was very close to the ELIE parameters Thrombin, despite with different binding modes. Thus, our future work will concentrate on the transferability of ELIE.

5. Conclusions

In this work, we conducted a systematic evaluation on how ELIE, which is a hybrid method of MM-PBSA and LIE, performed on ranking series of structurally similar ligands evolved from the same drug leads for eight target proteins. We also make comparison of ELIE to other popular methods, including Glide docking, regular MM-PBSA-WSAS, and FEP/TI. Overall, ELIE achieved a comparable accuracy as that of TI and FEP, and outperformed Glide docking and regular MM-PBSA-WSAS. Moreover, we found that 25 ns MD simulation is enough for optimal performance of ELIE binding free energy calculations for most systems. Considering ELIE achieves a good tradeoff between computational efficiency and accuracy, we expect this method will have a great application in drug lead optimization, especially after the binding affinities of many compounds for the target receptor have been measured.

Funding Sources:

The authors gratefully acknowledge the funding support from the National Science Foundation (NSF) and National Institutes of Health (NIH) to J.W. (NSF 1955260, NIH R01GM079383 and NIH P30DA035778), and from the National Natural Science Foundation of China (No.11774279, 11774280) to S.Z. The authors also thank for the computing resources provided by the Center for Research Computing (CRC) at University of Pittsburgh.

Author contributions:

J. Wang and X. He designed the project. X. He prepared the systems of proteins and ligands for

modelling and simulations. B. Ji performed the docking. D. Hao performed the MD simulations, MM-PBSA-WSAS calculations and ELIE fitting. All authors discussed the project and wrote the manuscript.

References

1. Dimasi, J. A.; Grabowski, H. G.; Hansen, R. W., Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, 47, 20-33.
2. D L Beveridge, a.; DiCapua, F. M., Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, 18, 431-492.
3. Gilson, M. K.; Zhou, H.-X., Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, 36, 21-42.
4. Deng, Y.; Roux, B., Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, 113, 2234-2246.
5. Jorgensen, W. L., The Many Roles of Computation in Drug Discovery. *Science* **2004**, 303, 1813-1818.
6. Jorgensen, W. L., Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, 42, 724-733.
7. Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O., Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, 35, 358-365.
8. Michel, J.; Verdonk, M. L.; Essex, J. W., Protein-Ligand Binding Affinity Predictions by Implicit Solvent Simulations: A Tool for Lead Optimization? *J. Med. Chem.* **2006**, 49, 7427-7439.
9. Gallicchio, E.; Lapelosa, M.; Levy, R. M., The Binding Energy Distribution Analysis Method (BEDAM) for the Estimation of Protein-Ligand Binding Affinities. *J. Chem. Theory Comput.* **2010**, 6, 2961-2977.
10. Huang, S.-Y.; Grinter, S. Z.; Zou, X., Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, 12, 12899-0.
11. Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A., Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, 50, 74-82.
12. Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K., D3R grand challenge 2015: Evaluation of protein–ligand pose and affinity predictions. *J. Comput.-Aided Mol. Des.* **2016**, 30, 651-668.
13. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, 49, 5912-5931.
14. Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K., Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, 32, 742-755.
15. Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V.

S., Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, 21, 150-160.
16. 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., 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.
17. Hansen, N.; van Gunsteren, W. F., Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, 10, 2632-2647.
18. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T., End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, 119, 9478-9508.
19. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, 51, 69-82.
20. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **2015**, 10, 449-461.
21. Gallicchio, E.; Levy, R. M., Advances in all atom sampling methods for modeling protein–ligand binding affinities. *Curr. Opin. Struct. Biol.* **2011**, 21, 161-166.
22. 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.
23. Hansson, T.; Marelius, J.; Åqvist, J., Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, 12, 27-35.
24. Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M., New Linear Interaction Method for Binding Affinity Calculations Using a Continuum Solvent Model. *J. Phys. Chem. B* **2001**, 105, 10388-10397.
25. He, X.; Man, V. H.; Ji, B.; Xie, X.-Q.; Wang, J., Calculate protein–ligand binding affinities with the extended linear interaction energy method: application on the Cathepsin S set in the D3R Grand Challenge 3. *J. Comput.-Aided Mol. Des.* **2019**, 33, 105-117.
26. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, 47, 1739-1749.
27. Cumming, J. N.; Smith, E. M.; Wang, L.; Misiaszek, J.; Durkin, J.; Pan, J.; Iserloh, U.; Wu, Y.; Zhu, Z.; Strickland, C.; Voigt, J.; Chen, X.; Kennedy, M. E.; Kuvelkar, R.; Hyde, L. A.; Cox, K.; Favreau, L.; Czarniecki, M. F.; Greenlee, W. J.; McKittrick, B. A.;

Parker, E. M.; Stamford, A. W., Structure based design of iminohydantoin BACE1 inhibitors: identification of an orally available, centrally active BACE1 inhibitor. *Bioorg. Med. Chem. Lett.* **2012**, 22, 2444-9.
28. 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, 3710-3722.
29. 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, 15-30.
30. 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, 56-69.
31. Akahane, K.; Li, Z.; Etchin, J.; Berezovskaya, A.; Gjini, E.; Masse, C. E.; Miao, W.; Rocnik, J.; Kapeller, R.; Greenwood, J. R.; Tiv, H.; Sanda, T.; Weinstock, D. M.; Look, A. T., Anti-leukaemic activity of the TYK2 selective inhibitor NDI-031301 in T-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* **2017**, 177, 271-282.
32. Szczepankiewicz, B. G.; Kosogof, C.; Nelson, L. T.; Liu, G.; Liu, B.; Zhao, H.; Serby, M. D.; Xin, Z.; Liu, M.; Gum, R. J.; Haasch, D. L.; Wang, S.; Clampit, J. E.; Johnson, E. F.; Lubben, T. H.; Stashko, M. A.; Olejniczak, E. T.; Sun, C.; Dorwin, S. A.; Haskins, K.; Abad-Zapatero, C.; Fry, E. H.; Hutchins, C. W.; Sham, H. L.; Rondinone, C. M.; Trevillyan, J. M., Aminopyridine-based c-Jun N-terminal kinase inhibitors with cellular activity and minimal cross-kinase activity. *J. Med. Chem.* **2006**, 49, 3563-80.
33. Goldstein, D. M.; Soth, M.; Gabriel, T.; Dewdney, N.; Kuglstatter, A.; Arzeno, H.; Chen, J.; Bingenheimer, W.; Dalrymple, S. A.; Dunn, J.; Farrell, R.; Frauchiger, S.; La Fargue, J.; Ghate, M.; Graves, B.; Hill, R. J.; Li, F.; Litman, R.; Loe, B.; McIntosh, J.; McWeeney, D.; Papp, E.; Park, J.; Reese, H. F.; Roberts, R. T.; Rotstein, D.; San Pablo, B.; Sarma, K.; Stahl, M.; Sung, M. L.; Suttman, R. T.; Sjogren, E. B.; Tan, Y.; Trejo, A.; Welch, M.; Weller, P.; Wong, B. R.; Zecic, H., Discovery of 6-(2,4-difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8H-pyrido[2,3-d]pyrimidin-7-one (pamapimod) and 6-(2,4-difluorophenoxy)-8-methyl-2-(tetrahydro-2H-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8H)-one (R1487) as orally bioavailable and highly selective inhibitors of p38α mitogen-activated protein kinase. *J. Med. Chem.* **2011**, 54, 2255-65.
34. 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-

98.
35. 2017-2, S. R. *Maestro*, Schrödinger LLC, New York, NY, 2017.
36. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, 97, 10269-10280.
37. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. B.01*, Wallingford, CT, 2016.
38. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157-1174.
39. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Model.* **2006**, 25, 247-260.
40. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, I., T.E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R. H., N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; D.M., Y.; Kollman, P. A. *AMBER 2018*, University of California, San Francisco, 2018.
41. 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.
42. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys* **1983**, 79, 926-935.
43. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **1992**, 32, 523-535.

44. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, 26, 1668-1688.
45. Massova, I.; Kollman, P. A., Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* **2000**, 18, 113-135.
46. Wang, J.; Hou, T.; Xu, X., Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Curr. Comput.-Aided Drug Des.* **2006**, 2, 287-306.
47. W.; Rocchia; E.; Alexov; B.; Honig, Extending the Applicability of the Nonlinear Poisson−Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001**.
48. Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D., Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, 2, 128-139.
49. Wang, J.; Hou, T., Develop and Test a Solvent Accessible Surface Area-Based Model in Conformational Entropy Calculations. *J. Chem. Inf. Model.* **2012**, 52, 1199-1212.
50. Pearlman, D. A.; Charifson, P. S., Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System †. *J. Med. Chem.* **2001**, 44, 3417-3423.
51. Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L., Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, 6, 3850-3856.
52. Wang, E.; Liu, H.; Wang, J.; Weng, G.; Sun, H.; Wang, Z.; Kang, Y.; Hou, T., Development and Evaluation of MM/GBSA Based on a Variable Dielectric GB Model for Predicting Protein–Ligand Binding Affinities. *J. Chem. Inf. Model.* **2020**.
53. Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **2014**, 16, 16719.

Table 1.

Overall Performance of Docking, MM-PBSA-WSAS and ELIE on Eight Protein Systems

System	BACE1	CDK2	JNK1	MCL1	P38	PTP1B	Thrombin	TYK2	average
no. of ligands	41	22	21	44	35	27	23	16	29
The null hypothesis RMSE[a]	0.78	1.13	0.84	1.06	0.99	1.27	1.39	1.26	1.09
Docking RMSE	0.86	1.42	1.02	0.97	0.94	0.95	1.33	0.84	1.04
Docking PI-value	0.03	0.25	-0.13	0.42	0.30	0.72	0.27	0.85	0.34
Docking R-value	0.04	-0.22	-0.07	0.42	0.24	0.68	0.30	0.75	0.27
Docking τ-value	0.00	-0.14	-0.08	0.28	0.22	0.52	0.23	0.65	0.21
MMPBSA-WSAS(1-25ns) RMSE	1.37	2.13	1.87	3.63	3.16	3.88	1.98	1.74	2.47
MMPBSA-WSAS(1-25ns) PI-value	0.23	0.35	-0.10	0.24	0.12	-0.23	0.68	0.29	0.20
MMPBSA-WSAS(1-25ns) R-value	0.11	0.28	-0.02	0.17	0.05	-0.10	0.50	0.15	0.14
MMPBSA-WSAS(1-25ns) t-value	3.63	3.63	3.63	3.63	3.63	3.63	3.63	3.63	3.63
ELIE(1-25ns) RMSE	0.88	0.80	1.30	0.98	1.30	0.88	1.06	1.12	1.04
ELIE(1-25ns) PI-value	0.39	0.73	0.55	0.40	0.62	0.75	0.75	0.51	0.59
ELIE(1-25ns) R-value	0.35	0.71	0.55	0.41	0.58	0.73	0.67	0.51	0.56
ELIE(1-25ns) τ-value	0.20	0.57	0.47	0.27	0.47	0.53	0.55	0.35	0.43
MMPBSA-WSAS(1-100ns) RMSE	1.40	1.59	1.85	3.70	2.30	6.22	2.13	1.52	2.59
MMPBSA-WSAS(1-100ns) PI-value	0.29	0.56	0.04	0.28	0.23	-0.09	0.68	0.51	0.31
MMPBSA-WSAS(1-100ns) R-value	0.29	0.52	0.04	0.33	0.22	-0.25	0.62	0.54	0.29
MMPBSA-WSAS(1-100ns) τ-value	0.16	0.39	0.05	0.20	0.14	-0.06	0.50	0.33	0.21
ELIE(1-100ns) RMSE	0.84	0.75	0.76	0.99	0.92	1.13	1.15	1.00	0.94
ELIE(1-100ns) PI-value	0.42	0.80	0.61	0.42	0.68	0.64	0.70	0.64	0.62
ELIE(1-100ns) R-value	0.39	0.75	0.60	0.41	0.67	0.57	0.63	0.62	0.58
ELIE(1-100ns) τ-value	0.25	0.57	0.54	0.29	0.51	0.40	0.53	0.43	0.44
FEP RMSE[b]	1.03	1.11	1.00	1.41	1.03	1.22	0.93	0.93	1.08
FEP R-value[b]	0.78	0.48	0.85	0.77	0.65	0.80	0.71	0.89	0.74
TI RMSE[c]	0.92	1.01	**	0.98	**	0.92	**	**	0.96
TI R-value[c]	0.48	0.64	**	0.65	**	0.75	**	**	0.63

Note: (a) The null hypothesis RMSEs were calculated by setting all predicted affinities to the average of experimental affinities for each protein system; (b) FEP RMSEs and FEP R-values are taken from Ref 16; (c) TI RMSEs and TI R-values are taken from Ref 22.

Figure captions

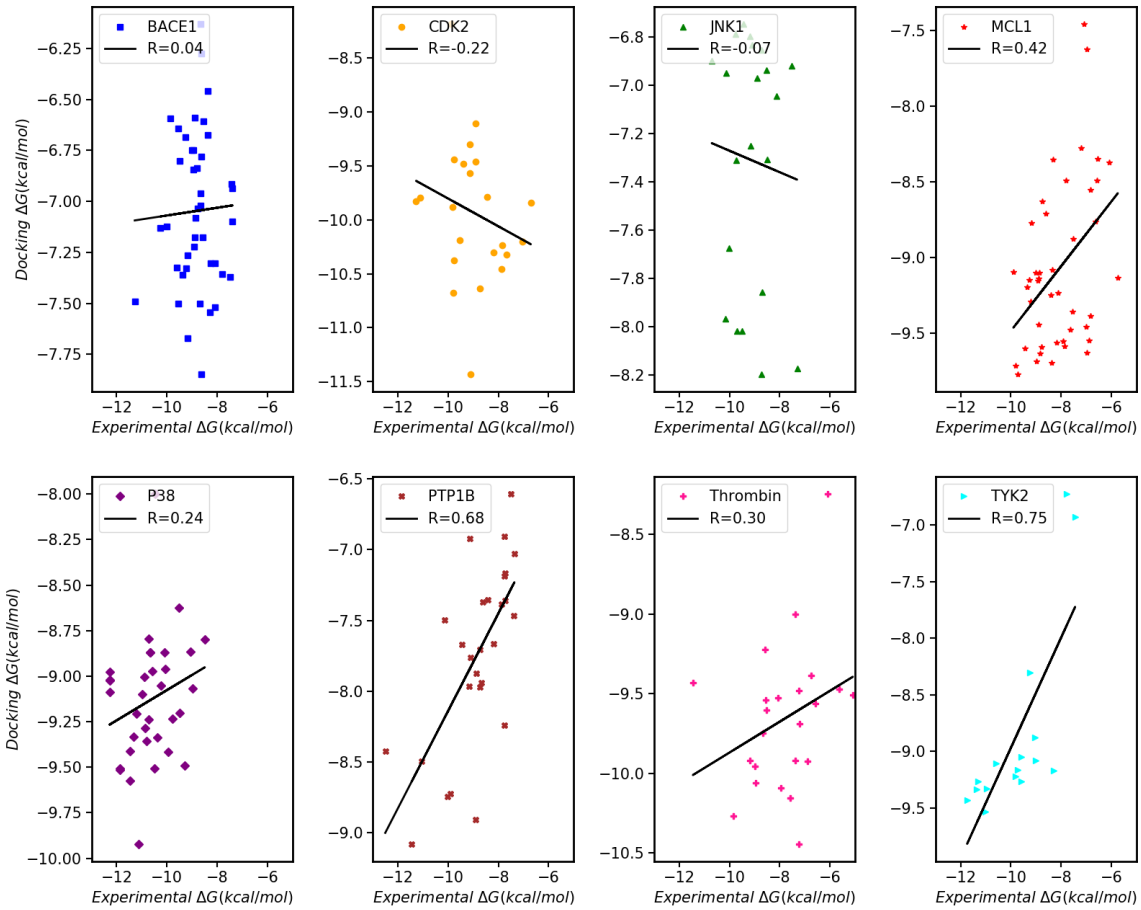


FIGURE 1 Absolute binding free energies predicted by docking method versus experimental values for the ligands of the eight systems.

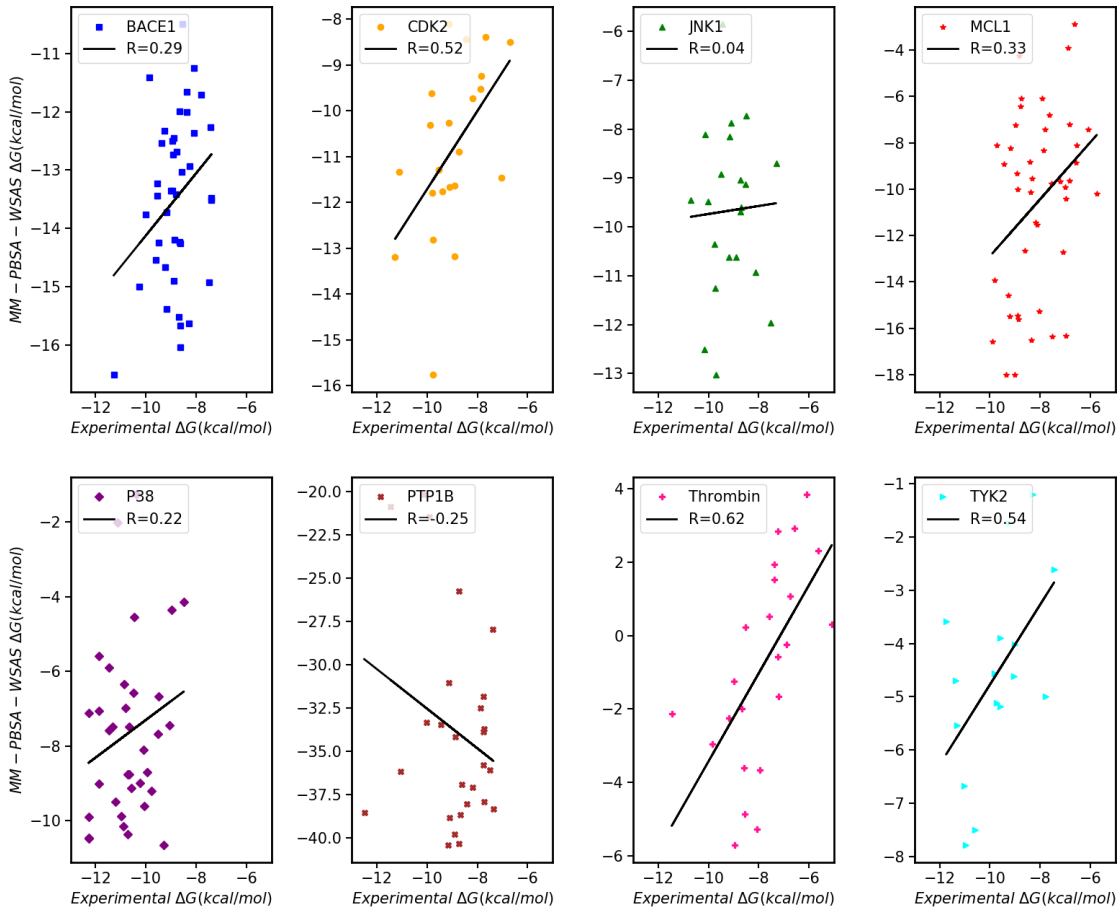


FIGURE 2 Absolute binding free energies predicted by MM-PBSA-WSAS versus experimental values for the ligands of the eight systems.

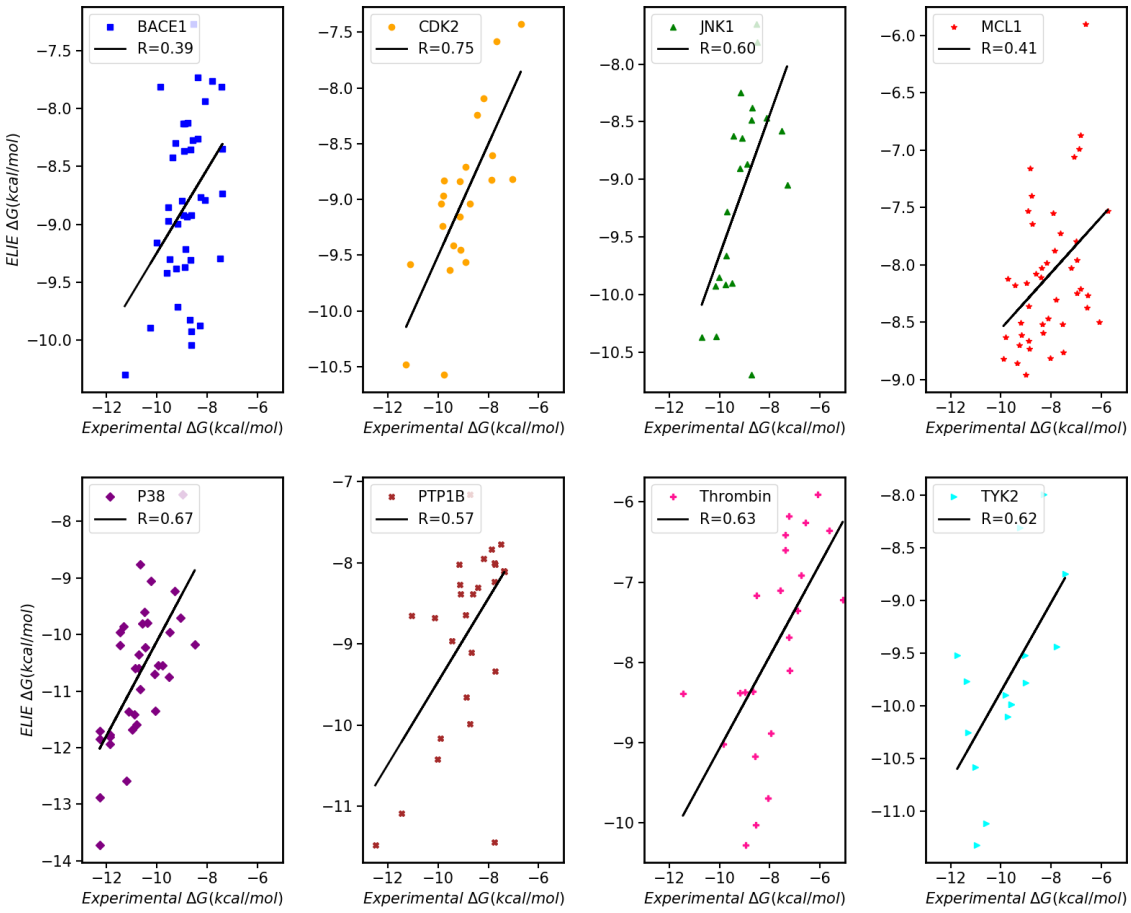


FIGURE 3 Relative binding free energies predicted by ELIE versus experimental values for the ligands of the eight systems.

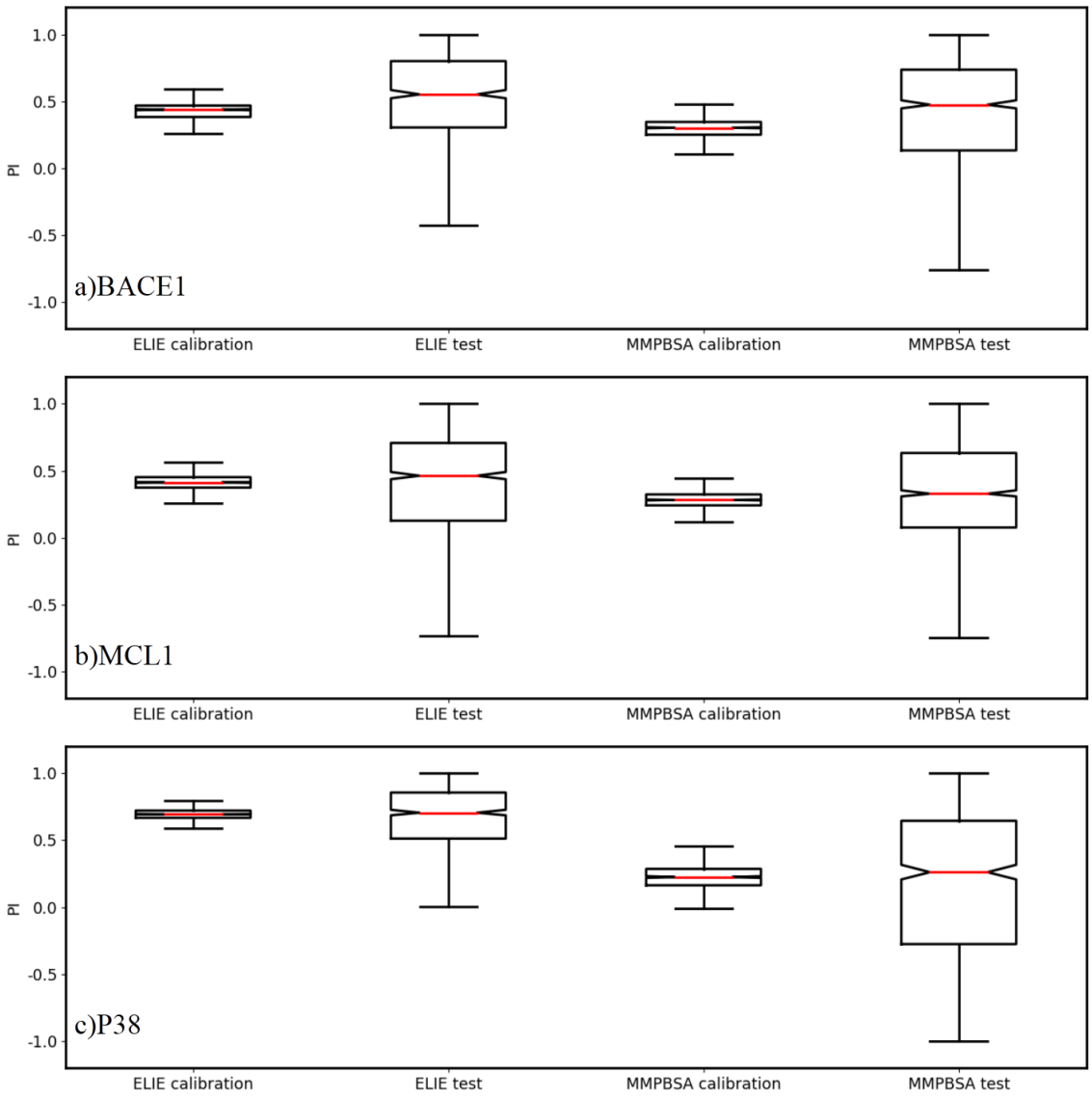


FIGURE 4 Perfomance of ELIE for test sets.

a)BACE1, b) MCL1, c) MCL1

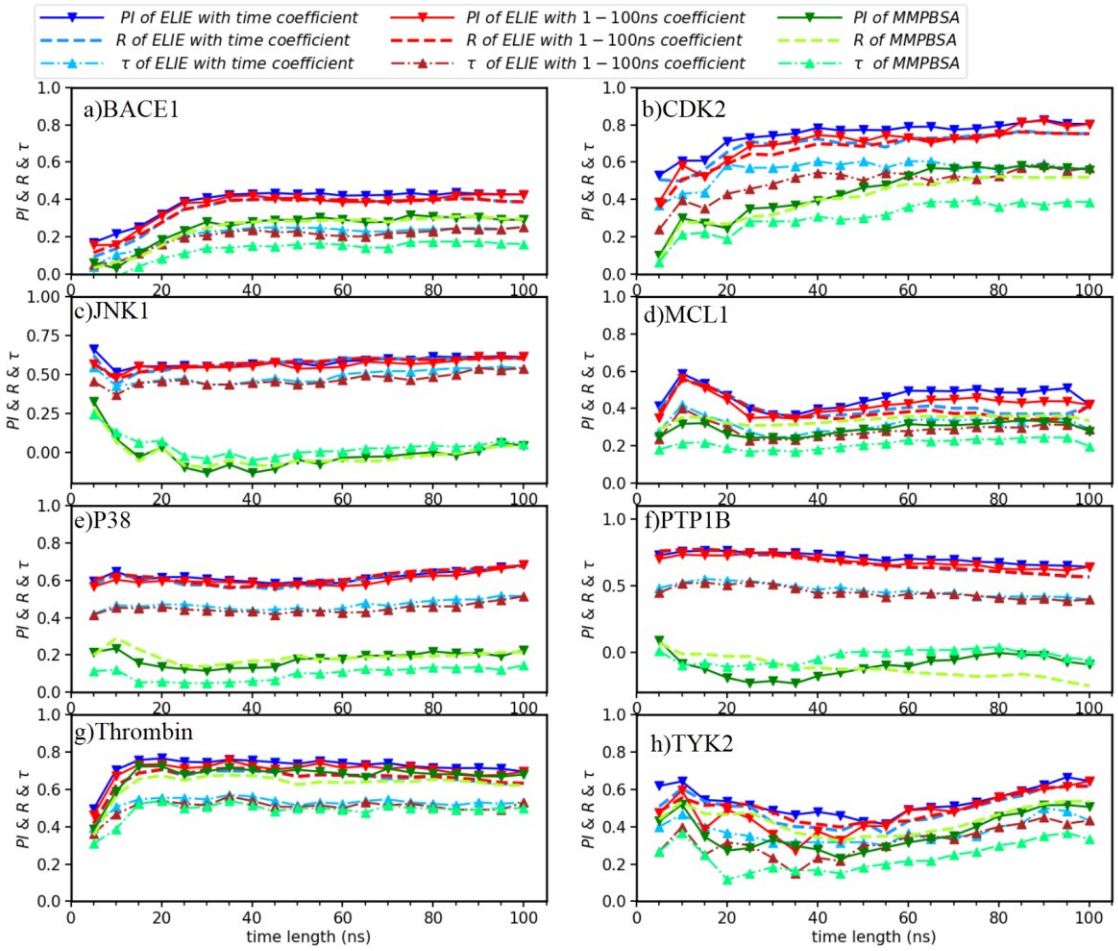
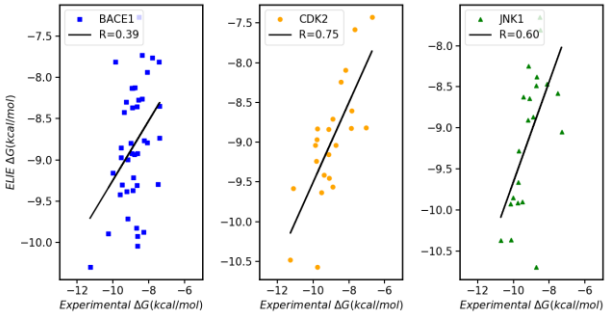


FIGURE 5 PI & R & τ between the experimental and the calculated binding free energies using 20 different lengths of MD simulation.

(a)BACE1, (b) CDK2, (c) JNK1, and (d)MCL1, (e)P38, (f)PTP1B, (g) Thrombin，(h)TYK2

TOC



Size (3.2 × 1.68 inches)