Computation of relative binding free energy for an inhibitor and its analogs binding with Erk kinase using thermodynamic integration MD simulation

Kuan-Wei Wu · Po-Chin Chen · Jun Wang · Ying-Chieh Sun

Abstract In the present study, we carried out thermodynamic integration molecular dynamics simulation for a pair of analogous inhibitors binding with Erk kinase to investigate how computation performs in reproducing the relative binding free energy. The computation with BCC-AM1 charges for ligands gave -1.1 kcal/mol, deviated from experimental value of -2.3 kcal/mol by 1.2 kcal/mol, in good agreement with experimental result. The error of computed value was estimated to be 0.5 kcal/mol. To obtain convergence, switching vdw interaction on and off required approximately 10 times more CPU time than switching charges. Residue-based contributions and hydrogen bonding were analyzed and discussed. Furthermore, subsequent simulation using RESP charge for ligand gave $\Delta\Delta G$ of -1.6 kcal/mol. The computed results are better than the result of -5.6 kcal/mol estimated using PBSA method in a previous study. Based on these results, we further carried out computations to predict $\Delta\Delta G$ for five new analogs, focusing on placing polar and nonpolar functional groups at the meta site of benzene ring shown in the Fig. 1, to see if these ligands have better binding affinity than the above ligands. The computations resulted that a ligand with polar –OH group has better binding affinity than the previous examined ligand by ~ 2.0 kcal/mol and two other ligands have better affinity by ~ 1.0 kcal/mol. The predicted better inhibitors of this kind should be of interest to experimentalist for future experimental enzyme and/or cell assays.

Abbreviations
Przp Pyrazolylpyrrole
TI   Thermodynamic integration
MD   Molecular dynamics

Introduction

Kinases are important drug targets for the treatment of a number of diseases, including cancers, neurodegenerative diseases, and diabetes; indeed, they are considered the second most important target protein group of pharmaceutical interest [1]. For example, several kinases in the Egfr/Raf/Mek/Erk cell signaling pathway, which is associated with cell survival, proliferation, and apoptosis, are target proteins of approved drugs and/or inhibitors in clinical trials. The development of kinase inhibitors is, therefore, an attractive challenge for drug design.

Recent advances in computational chemistry have significantly enhanced the ability to predict the binding free energies of ligand–protein complexes. The levels of accuracy and the advantages and disadvantages of the various computational methods have been summarized in several review papers [2–6]. In general, the demand for CPU time increases upon proceeding from docking and Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) to Thermodynamic Integration/Free Energy Perturbation (TI/FEP) and absolute binding free energy methods. The

K.-W. Wu · P.-C. Chen · J. Wang · Y.-C. Sun (✉)
Department of Chemistry, National Taiwan Normal University, 88, TingChow Road Section 4, Taipei 116, Taiwan
e-mail: sun@ntnu.edu.tw

Fig. 1 Structures of Przp and its analog. 3-dimensional structures are shown in *color* (in magenta and *grey*, respectively). The units of Przp that differ from those of its analog are highlighted in *orange*; these atoms are also labeled. *Other color codes* are the standard ones of Chimera, the graphic software used to produce this figure. Here, Przp is the reference ligand ($\lambda = 0$); its analog is the mutant ligand ($\lambda = 1$)

Przp ligand scaffold   R group of Przp ligand   R group of its analog
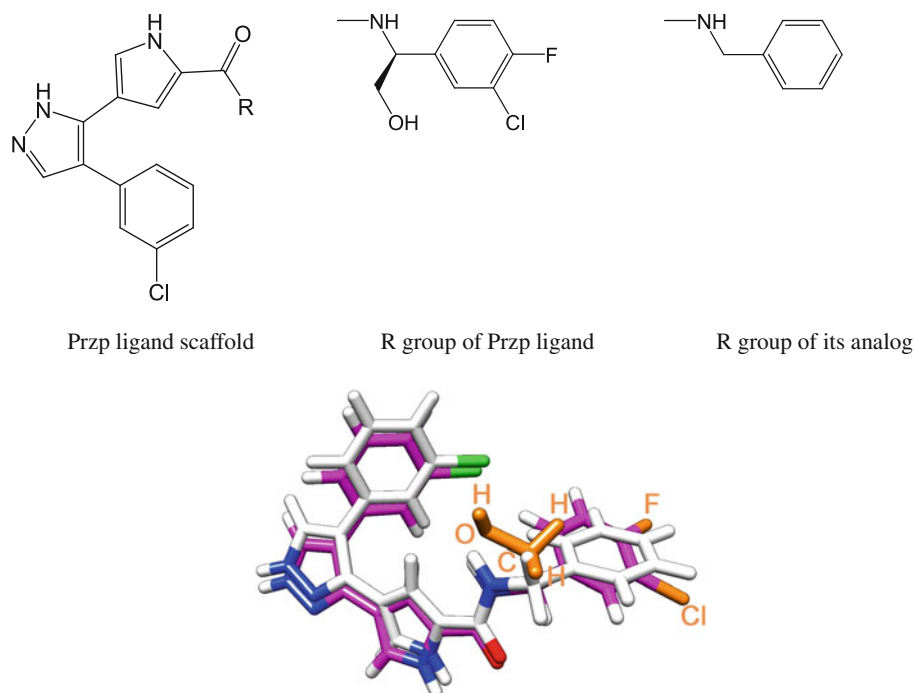
latter two methods are more-rigorous first-principle methods that provide higher accuracies than do the former two. Significant advance has been made in obtaining accurate binding free energy of biomolecule-ligand complexes using the latter two methods in a number of simulations using various computational protocols (see literatures [7–37] and references [2, 4]). For example, earlier simulation with partial solvation at binding site and structural restraints gave excellent agreement with experimental values at accuracy of ~0.5 kcal/mol for ligands binding with a kinase [7]. As computer power increases, it is natural to explore computation with full solvated model and without structural restraints. Although this type of computation is increasingly common, the number of simulations using this protocol remains small due to high demand of cpu time. Accuracy was estimated in average to be 1.4–2.5 and <3.0 kcal/mol based on investigated systems in two review papers, respectively [2, 4]. One notable example of high accuracy at ~0.5 kcal/mol is the computation for trypsin-ligand complexes [33]. Compared with absolute binding free energy calculation method, TI and FEP are more affordable first-principle methods. These approaches are, however, limited to determining only the relative binding free energy for two similar inhibitors interacting with a protein. For these two methods, significant progress has been made to enhance the accuracy of the calculated results when applied to various molecular systems [25–36]. For treatment of disappearing/appearing atoms, soft core potential has been developed recently and applied to a number of free energy calculations; using this treatment has been successful for a number of molecular systems [25–33]. Nevertheless, number of simulations using this newly-developed treatment for a pair of analogous ligands binding with proteins with fully solvated model and without structural restraints remains small. In this present study, we applied this approach to investigate how this protocol performs in reproducing the relative binding free energy of two similar kinase–inhibitor complexes of pharmaceutical interest.

Among the experimentally characterized kinase–inhibitor complexes, those of Erk kinase and its inhibitor pyrazolylpyrrole (Przp) and its analog are ideal for TI computation because of the availability of crystal structures and binding affinity data and because of their similar binding modes [38, 39]. This system has been investigated using the MM-PBSA approach [40, 41]. In addition to its great suitability for TI computation, the protein Erk kinase is of high pharmaceutical interest as a biomarker for a number of cancers, including lung and colorectal cancers, and as a potential target protein for drug discovery. Figure 1 displays the structures of the two inhibitors. Przp differs from its analog by the presence of a CH_2OH group and F and Cl atoms on the benzene ring close to the glycine-rich loop of Erk kinase [38]. The calculated $\Delta\Delta G$ is in good agreement with experimental value, showing success of the present computational procedure. In addition to these two ligands, we employed the same protocol to predict binding affinity for several new analogous ligands in order to find better inhibitors. It was found that ligands having polar functional group like –OH at the benzene ring shown in the figures below, which was not investigated in previous experimental study [38], enhances affinity of the ligand binding with the Erk

kinase. This prediction should be of interest to experimentalists for future study. Computational method and procedure are described in the section "Method and computation". Section "Results and discussion" contains results and discussion. Concluding remarks are given in the section "Concluding remarks".

Method and computation

Formulation

The fundamental formulation and computational procedure for free energy calculation can be found in a review paper [42]. The relative binding free energy ($\Delta\Delta G$) of a compound and its derivative binding with a protein can be calculated using the thermodynamic cycle displayed in Fig. 2. The relative binding free energy, $\Delta G_{inh} - \Delta G_{der}$, is equal to

$$\Delta G_{mut,s} - \Delta G_{mut,p}$$

where the subscripts "inh" and "der" denote the inhibitor and its derivative, respectively, "mut" is the mutational change, and "p" and "s" represent the protein bound and solution states, respectively. The mutational free energy change is calculated using the equation

$$\Delta G_{mut} = \int_0^1 d\lambda \langle dV/d\lambda \rangle_\lambda$$

where λ is the mutational variable switching from 0 (reference state) to 1 (mutated state); $< >_\lambda$ is the ensemble average of the calculated quantities sampled at the hybrid Hamiltonian at the specific value of λ; and $<dV/d\lambda>_\lambda$ is calculated using MD simulation. In practice, selected λ points are computed in MD simulations to give integrants at those λ points and are used in subsequent numerical integration to obtain ΔG_{mut}. In the present computation, the values were computed at values of λ of 0.1, 0.2, …, and 0.9. The values at values of λ of 0.0 and 1.0 were obtained through extrapolation, following the procedure described in the Amber11 manual.

Computation

To compute $\Delta G_{mut,p}$ and $\Delta G_{mut,s}$, the modeling systems were prepared using the Amber11 package [43]; ΔG_{mut} is the mutational free energy change, with p and s denoting protein and solution states, respectively. For the protein bound state, the Erk–Przp structure obtained from the protein databank (pdb; code: 2ojj) was used as the initial structure. For the ligands, the coordinates of the atoms were obtained from their pdb files (codes: 2ojj and 2oji,
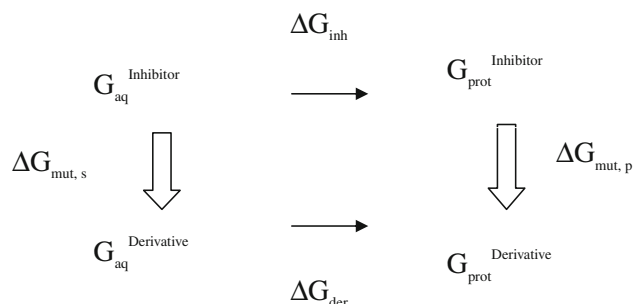
Fig. 2 Thermodynamic cycle of binding

respectively) and the atoms of both ligands in the files were re-arranged such that common atoms were "matched" and the atoms in difference were placed at the end of atom lists. The resolutions of crystal strcutures (codes: 2ojj and 2oji) are 2.4 and 2.6 Å, respectively. The atom charges of both ligands were derived using the AM1-BCC method with the antechamber module [44]. Other force field parameters were obtained from the gaff force field and the parchmk module. The Amber ff03.r1 force field was used for the protein. Three Na^+ ions were added to neutralize the net charge of the complex. A rectangular box of TIP3P water molecules was added such that the minimum distance of any solute atom to the walls of the box was 12 Å. In total, the protein state simulation involved 54,646 atoms. The topology and coordinate files for the analog were prepared following the procedure in the Amber11 manual [43]. We employed a similar procedure for the solution state. The mutation of Przp into its analog was divided into three steps, following the Amber11 manual [43]. First, we switched off the electrostatic interactions of the atoms of Przp that differ from those of its analog by turning the partial charges to zero. Second, the atoms of Przp differing from those of its analog were "disappeared" and the atoms of its analog differing from those of Przp were "appeared," using the soft core potential scheme of van der Waals (vdw) interactions [43]. Third, we switched on the partial charges of the atoms of the analog differing from those of Przp to complete the mutation.

We subjected the system to an energy minimization of 4000 steps of steepest descent and 1000 steps of conjugate gradient, followed by an NTV MD simulation for 10 ps with an initial temperature of 10 K and a target temperature of 300 K, using the Amber default random number seed for atom velocity assignment. We performed a subsequent NTP MD simulation for 200 ps as an equilibration run. MD simulations for a further 600 ps were performed as production runs and used to calculate $<dV/d\lambda>_\lambda$ and the resulting values of ΔG_{mut} and perform residue-based free energy component analysis. The value of $dV/d\lambda$ was collected every 0.2 ps for each run, resulting in a total of 3,000 data points for further analysis. In MD simulations, all bonds associated

with hydrogen atoms were constrained at their equilibrium bond lengths with a time step of 2 fs, except for the second mutational step for which there were no bond constraints and the time step was 1 fs, due to the use of the soft core potential for the disappearing and appearing atoms [43]. The structure of the complex in the MD simulations remained close to the crystal structure with root-mean-square deviations (RMSDs) ranging from approximately 1.2–2 Å. To investigate the residue-based contributions to the free energy difference, the free energy decomposition option was turned on during simulation; the analyzed results are reported below. As to the cpu time used, in total (see the total simulation length described below), the simulations took approximately 100,000 cpu core hours.

Investigation of the convergence of results revealed that the second mutational step in the protein state did not provide a converged result after a 600-ps production run (see Fig. 3). To pursue convergence, we extended this run to 1.2 ns for the three mutational steps in the protein state. The results of the first and third mutational steps revealed good convergence, but the second mutational step did not; it displayed large uncertainty. To pursue this matter further, we ran more independent trajectories for the second mutational step in the protein state with different random number seeds for the initial atom velocity assignment and then examined the convergence. We obtained good convergence for up to six trajectories (Fig. 3; step 2); the averages of the results of these six trajectories are reported and discussed herein. We estimated the error bars of the calculated values of $\langle dV/d\lambda \rangle_\lambda$ by examining the variation of calculated values at long time points on simulation time. Values at 75, 150, 300, and 600 ps are shown in the Fig. 4; for the protein state, values at 1,200 ps are also calculated. It was seen that calculated values in the solution state have variations of ∼0.1 kcal/mol only, showing good convergence due to the homogeneous environment in the solution state. For protein states, the first and third steps have variation of ∼0.1 kcal/mol but variation of the second mutational step is ∼0.5 kcal/mol. With these, the error bar is therefore estimated to be 0.5 kcal/mol. In addition to this error estimation, another way of estimating uncertainty is through confidence interval analysis [45], which is given in the supportive information to demonstrate how error bars decreases as simulation time increases.

Results and discussion

Calculated $\Delta\Delta G$ with BCC-AM1 charge

Table 1 and Fig. 4 summarize the calculated results for the three mutational steps and the effect of the simulation length. Figure 5 displays calculated $\langle dV/d\lambda \rangle_\lambda$ and its
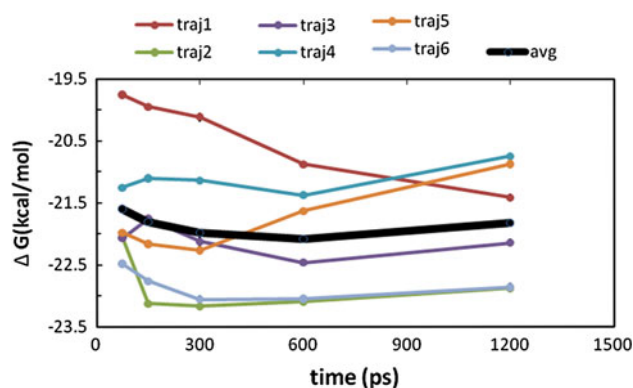

Fig. 3 Simulation length dependence of calculated ΔG of the second mutational step in protein state (AM1-BCC charge). traj1 denotes the 1st trajectory etc. avg denotes the average of these trajectories
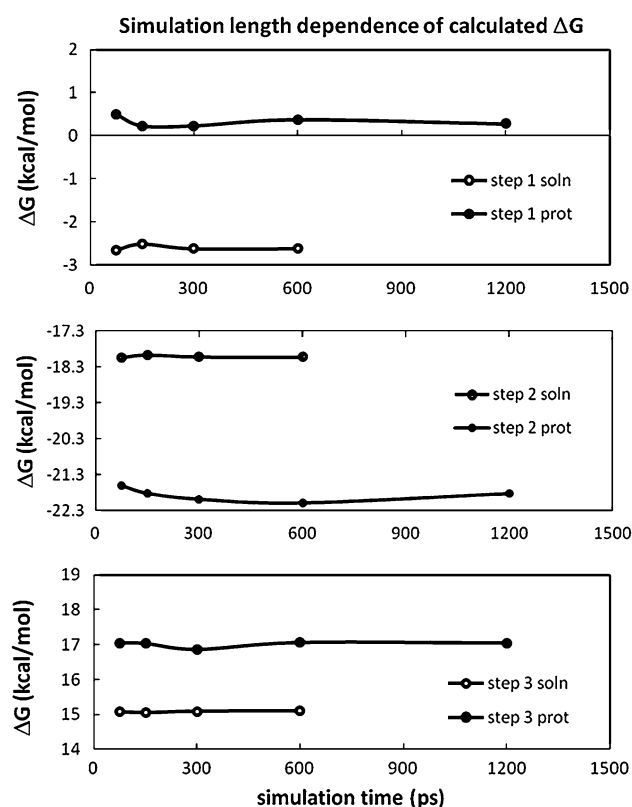

Fig. 4 Simulation length dependence of the calculated values ΔG of the three mutational steps (*soln* solution state, *prot* protein state). Note that the results of the second mutational step in the protein state were obtained from averaging six independent trajectories (see text for detailed description)

standard deviation at λ points. A straightforward evaluation gave a relative binding free energy of −1.1 kcal/mol, in good agreement with the experimental value of −2.3 kcal/mol, based on values of K_i of 2 and 86 nM for Przp and its analog, respectively [38]. The calculated value is better than the value of −5.6 kcal/mol estimated using PBSA method in a previous study [41]. As mentioned above, we

estimated the error of the calculated results to be 0.5 kcal/mol; the error of the experimental results is not available. Figure 4 presents the effect of the simulation length on the calculated values of ΔG_{mut} and the estimated errors. Among the three mutation steps, the second (disappearing/appearing vdw interactions) had the largest errors. The smaller errors for the protein state resulted from the averaging of six independent trajectories; an individual trajectory did not always result in convergence after simulation for up to 1,200 ps for the protein state, whereas more simulation trajectories (up to six) of 1,200 ps each gave good convergence. The mutational steps for the solution state, and the first and third mutational steps for the protein state, gave converged calculated values of ΔG_{mut} after simulation for hundreds of picoseconds at each λ point, but the second mutational step in the protein state required approximately 5 ns (ca. 10 times longer than the other mutational steps). These results suggest how long the MD simulation lengths should be for the first and third mutational steps (switching charges on/off) and for the second mutational step (disappearing/appearing vdw interactions) in order to obtain converged results using the present TI computational protocol for these kinds of ligand–protein complexes.

Residue-based contributions

The key features regarding the residue-based contributions to the free energy difference are summarized as follows. Table S2.A lists the residues providing significant contributions, with absolute values greater than 0.3 kcal/mol; Fig. 6 highlights the residues in the largest group. The largest three contributions (>1.0 kcal/mol) were 1.7, 1.5, and −1.1 kcal/mol for Ser151, Arg13, and Lys52, respectively. Together with the hydrogen bonding analysis described in the next paragraph, the interactions of Przp with these three residues were all polar interactions. This result, along with many other charged residues providing higher contributions in Table S2.A, reveals that charged residues interact strongly with the dipoles at the mutational sites of Przp. Our findings are consistent with the previous results reported using the MM-PBSA approach, where
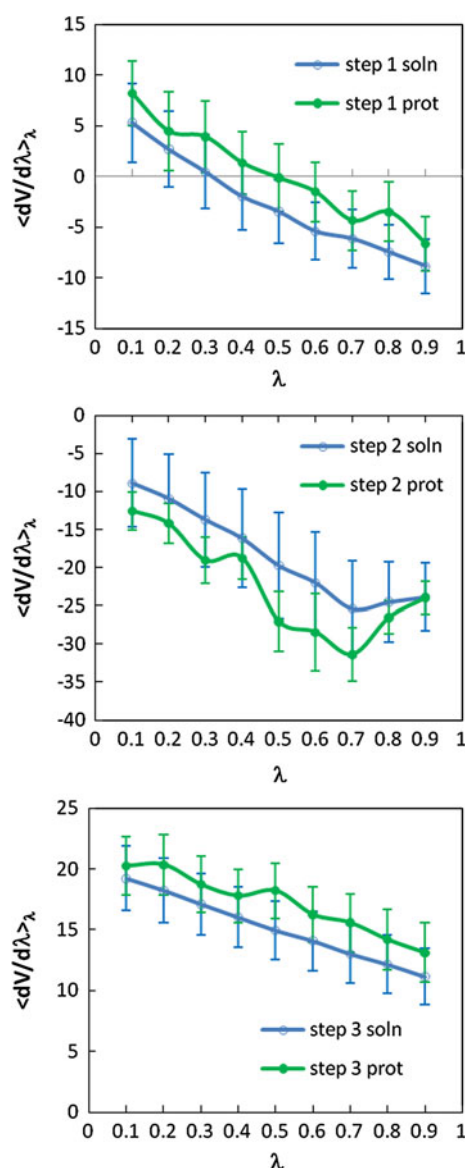

Fig. 5 Calculated $<dV/d\lambda>_\lambda$ and its standard deviation at λ points for three mutational steps in solution and protein states. Their values are given in the Table S1 of supportive information

electrostatic interactions were determined to be the major interactions stabilizing the Przp–Erk kinase complex [40]. Notably, however, our present calculated residue-based

Table 1 Calculated free energy changes (kcal/mol) using BCC-AM1 charge

Mutational step	In solution	Bound in protein	$\Delta\Delta G$	Exp.[a]
1	−2.6	0.3		
2	−18.0	−21.8		
3	15.1	17.1		
Total	−5.5	−4.4	−1.1	−2.3

[a] Converted from experimental values of K_i according to $\Delta G = -kT \ln K$. The values of K_i for Przp and its analog are 2 and 86 nM, respectively [38]

contributions at the active site agree with the previously reported results (MM-PBSA method) for only some of the residues [40]. Furthermore, our present computation identified residues Arg13 and Res109, located at intermediate distances from the active site, as also contributing significantly to the difference in binding free energy. Moreover, we examined the effects of vdw contributions; Table S2.B lists the residues for which contributions were greater than 0.2 kcal/mol. The largest contributions were those of the residues Tyr34 and Gly35, located at the glycine-rich loop close to the F and Cl atoms of Przp at the mutational site. This result supports the notion, determined from the experimental study, that the F and Cl atoms at this location stabilize the interaction between Przp and Erk kinase, mainly through vdw contacts [38]. When interpreting molecular interactions between ligands and proteins, such calculated values of free energy decomposition can be complementary to structural information, and might be useful for designing inhibitors, understanding how the mutation of a residue around the binding site can alter the binding free energy, determining inhibition effects, and increasing the efficacy of drugs toward protein mutants.

Hydrogen bondings

We investigated the hydrogen bonding between the compounds and the Erk kinase in the MD simulations. Note that simulations at λ points are intermediate states during mutations and are not physical molecules; the λ points of 0.1 in the first mutational step and 0.9 in the third mutational step are closest to the structures of the Przp–Erk and

Fig. 6 Residue-based free energy decomposition. Residues providing the largest contributions—Arg13, Tyr34, Lys52, Asp109, Ser151, and Asp165 (0.8–1.7 kcal/mol) are highlighted in *orange-red*. See the supporting information for the smaller contributions of the other residues

analog–Erk complexes, respectively. Straightforward calculation revealed that the OE1 atom of Gln103, the O atom of Met106, and the O atom of Ser151 form hydrogen bonds with the NH, NH, and O1-H units of Przp, respectively, in populations of 98, 72, and 99 %, respectively, at the λ point of 0.1 in the first mutational step. The first two hydrogen bonds, so-called hallmark hydrogen bonds of kinase inhibitors at the hinge segment of kinases, remained at high population in all nine of the λ point simulations; in contrast, the third hydrogen bond disappeared gradually upon increasing the λ point up to 0.9. This hydrogen bond differs from that in the crystal structure [38], where the OH group of the CH$_2$OH unit of Przp forms hydrogen bonds with residues Asn152 and Asp165. Notably, the initial structure from the pdb and the energy-minimized structure did contain a hydrogen bond with Asn152. After MD simulation, however, the OH group shifted and formed a hydrogen bond with the backbone CO group of the Ser151 residue. This feature was also observed in a previous MD simulation using a similar force field [40]. The difference from the experimental structure might arise from the simulations being at intermediate states, the modeling of a solution environment rather than a crystal environment, and/or inaccuracy in the force field. Moreover, interestingly, the F atom did not form any hydrogen bonds with Erk kinase, because the C–F bond was oriented toward the space within the β-hairpin of the glycine-rich loop and did not align to form a hydrogen bond with the NH units on the backbone or other hydrogen donors on side chains. In the structure of the third mutational step, which corresponds to the derivative inhibitor, the O atom of the derivative formed hydrogen bonds not only with three NH units on the side chain of the Lys52 residue but also with the NH unit on the side chain of the Gln103 residue (with a lower population of 10 %)—interactions that we did not observe in the Przp simulation. These hydrogen bonds slightly enhance the interaction between the derivative and the hinge segment. Such studies are useful in rationalizing the stabilizing interactions between the examined ligands and Erk.
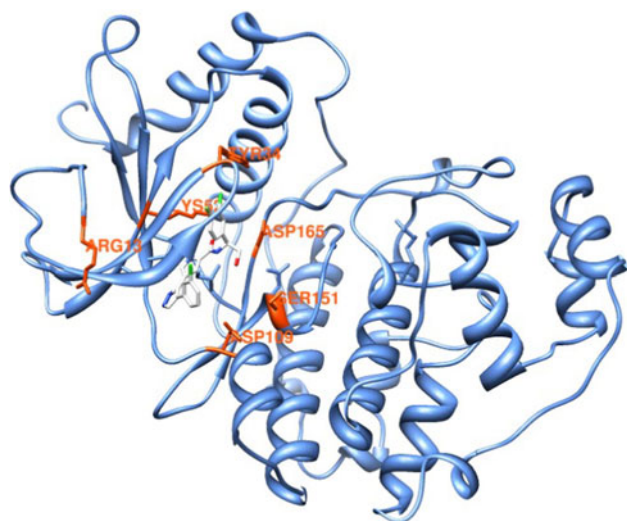
Table 2 BCC-AM1 and RESP charges of the selected atoms differing between Przp and its analogous inhibitors shown in the Fig. 1

Atom[a]	BCC-AM1	RESP
F31	−0.1259	−0.21352
CL33	−0.0684	−0.08227
C2	0.1244	0.074994
O1	−0.6008	−0.65439
H21A	0.0407	0.054045
H22A	0.0777	0.054045
HO1	0.4050	0.436719

Unit: charge of a proton

[a] Atom names in the original pdb file (pdb code: 2ojj)

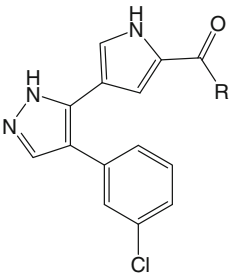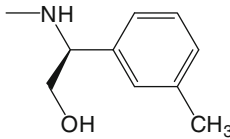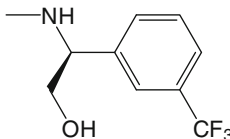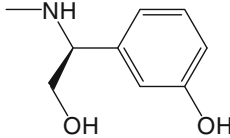**Table 3** Calculated free energy changes (kcal/mol) using RESP charge
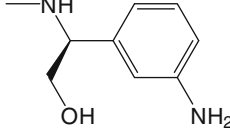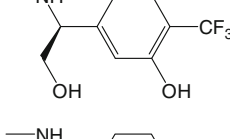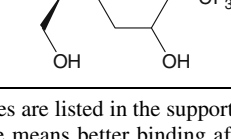
Mutational step	In solution	Bound in protein	$\Delta\Delta G$	Exp.
1	4.1	9.4		
2	−26.1	−31.5		
3	21.4	23.1		
Total	−0.6	1.0	−1.6	−2.3

Computation with RESP charge

To investigate further, we carried out simulation with RESP charge for ligands. Gaussian 03 [46] was used to carry out quantum chemistry calculation for two ligands at the level of HF/6-31G* and the outputs were input into Amber package to derive atomic charges of ligands following the Amber manual [43] Partial charges of the selected atoms in difference between two ligands are listed in Table 2. In general, the charges of more electronegative O, F, Cl atoms were underestimated in BCC-AM1 charges by charge amount of 0.05–0.1. With RESP charges in hand, TI MD simulations were carried out in the same protocol with the simulations with BCC-AM1 charge described above except the simulation length in the protein state of mutational steps 1 and 3 were shortened from 1,200 to 600 ps because these simulations converged well. The computed results are listed in Table 3. The computation gave $\Delta\Delta G$ of −1.6 kcal/mol. Convergence of calculated results, hydrogen bonding, and residue-based contributions

Table 4 Predicted $\Delta\Delta G$ (kcal/mol) for new analogous ligands, taking the Przp compound as reference

Ligand scaffold	R group	$\Delta\Delta G$ (kcal/mol)
		−0.1
		−0.6
		2.0
		1.3
		1.0
		−0.8

Values of calculated ΔG for each step in solution and protein states are listed in the supportive information. Here, the $\Delta\Delta G$ was calculated taking the original Przp ligand as reference ligand. Positive $\Delta\Delta G$ value means better binding affinity than the original Przp ligand
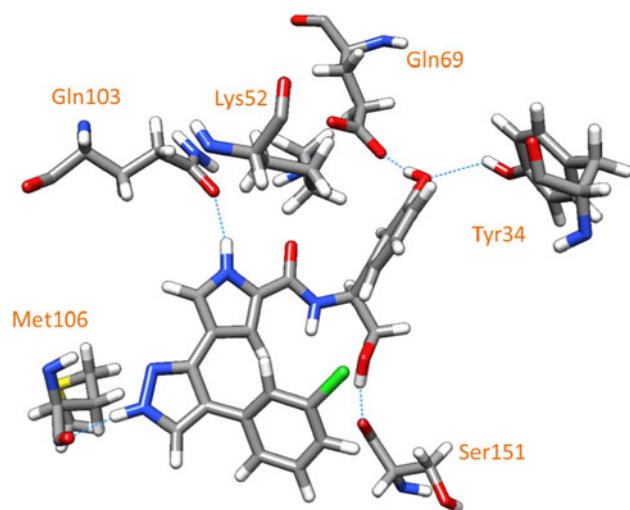
Fig. 7 Hydrogen bondings of the predicted better inhibitor interacting with the Erk kinase at the binding site. Hydrogen bondings with Gln69 and Tyr34 are additional hydrogen bonds due to presence of substituted –OH group on the benzene ring in addition to the three hydrogen bonds existing in the reference Przp ligand and its first analog interacting with Erk kinase. Hydrogen bonding with Gln69 is stronger than with Tyr34 and their percentages of presence in MD simulation are 89 and 48 %, respectively

are similar to the results of BCC-AM1 charge. Overall, the computed results in the present simulations are encouraging in obtaining relative binding free energy for two analogous inhibitors binding with a kinase using the present computational protocol.

Prediction for new analogous ligands

The above results of Przp and its analog showed the computational protocol with RESP charge gave result in better agreement with experimental value. Computations with this protocol were thus carried out to predict binding free energy of several new analogous ligands listed in the Table 4. To our knowledge, binding affinity data of these chosen ligands were not available, and we chose to place –CH$_3$, –CF$_3$, –OH, and –NH$_2$ at meta site of the benzene ring shown in the Table 4 to see if these substitutions enhance binding affinity compared with Przp ligand. The computed $\Delta\Delta G$ are listed in Table 4. The $\Delta\Delta G$ values were calculated taking the Przp ligand as the reference ligand. In case that $\Delta\Delta G$ is a positive value, this means the examined ligand has better binding affinity than Przp ligand, thus, better inhibitor. Clearly, the analog with polar group –OH and –NH$_2$ has better affinity than Przp by 2.0 and 1.3 kcal/ mol, respectively, but affinity of the analogs with –CH$_3$ and –CF$_3$ nonpolar groups are lower than or similar to Przp compound's affinity. The fifth ligand with –CF$_3$ at para site also lowers down affinity compared with the third ligand. In addition, we also examined if a cyclohexane group has
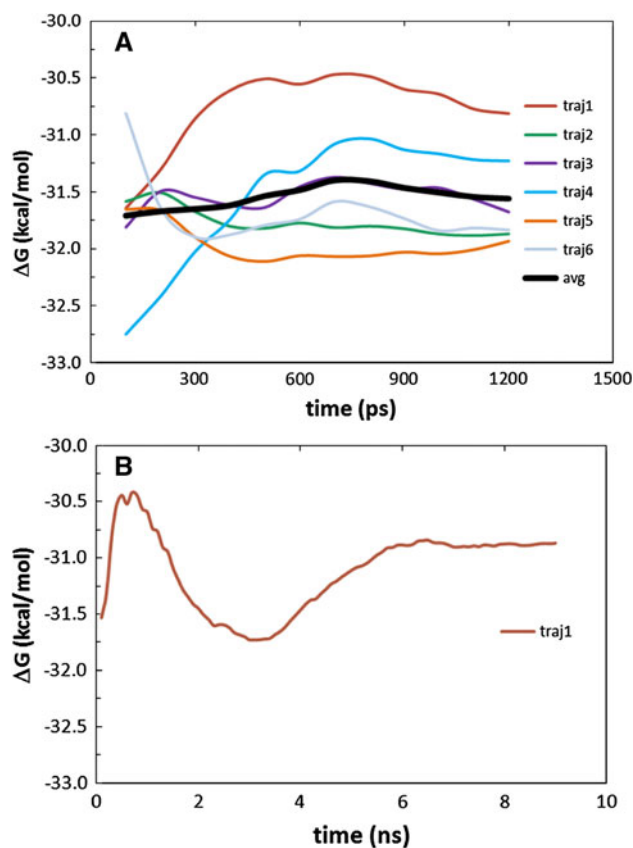

Fig. 8 **a** Simulation length dependence of calculated ΔG of the second mutational step in protein state (RESP charge). traj1 denotes the 1st trajectory etc. avg denotes the average of these trajectories. **b** Simulation length dependence of calculated ΔG of the second mutational step in protein state (long trajectory of traj1. see text for detailed description)

better affinity than benzene group as shown in the last raw of the table. The results show the planar benzene ring is favored compared with the cyclohexane ring by 1.8 kcal/ mol. To investigate which interactions enhance affinity, a standard MD simulation of complex of Erk kinase and the analog with –OH group was carried out for 2.0 ns. Analysis of hydrogen bonds showed that, this –OH group formed a hydrogen bond with Gln69 (89 % presence of hydrogen bond in trajectory) and a minor hydrogen bond with Tyr34 (48 % presence of hydrogen bond in trajectory) enhancing the binding affinity (see Fig. 7). The present results show that presence of a polar group at the meta site of the benzene ring enhances affinity for this kind of compounds. It is noted that previously examined analogs are mostly non-polar functional groups on the benzene ring shown in the Fig. 1 [38]. The present computed results suggest that placing a polar group on the meta site of this benzene ring enhances binding affinity with the Erk kinase. This kind of compounds should be interesting to experimentalists for further enzyme and/or cell assay studies.

Multiple short trajectories versus a single long trajectory

Previous studies showed that, in general, relative and absolute binding free energy calculations for ligand–protein complexes of this kind need simulation length in the order of 10 ns to obtain binding free energy with precision level of desire (see Ref. [47] and references therein). MD simulation in the order of nanosecond length is expected to be too short and may not give result of high precision for many cases. However, in the relative binding free energy calculation, this problem may be avoided if conformational change affects the binding free energy for each ligand in the same way [47]. To investigate this issue of long trajectory, we chose one of the cases above and extended simulation trajectory. Among all run ligand mutation cases, we chose the case of the first two ligands, which have both experimental complex structure and binding affinity data available, to examine, and chose the RESP charge case, which gave closer value to the experimental value, instead of AM1-BCC charge. Among the 6 run trajectories, we chose the first trajectory which has the largest deviation from the averaged value to have a critical evaluation. A long trajectory of $\sim$9 ns was obtained and the result is shown in Fig. 8b together with the results of 6 short trajectories shown in Fig. 8a for comparison. Although in the first 6 ns, the calculated values have a spread (the difference between the maximum and the minimum) of $\sim$1.6 kcal/mol. However, in the remaining $\sim$3 ns, the curve shows converging to −30.9 kcal/mol, which differs from the averaged value of −31.5 kcal/mol obtained from 6 short trajectories by 0.6 kcal/mol. This number suggests that multiple short trajectories give result comparable with the long simulation run in the present case. Notably, the spread of the average curve (think black line in Fig. 8a) is as small as $\sim$0.6 kcal/mol as seen in the AM1-BCC charge case shown in the Fig. 3. Use of multiple short trajectories is also supported by a number of previous computations in literature comparing multiple short runs and a single long trajectory (see, for example, Ref. [48]).

Concluding remarks

In summary, we carried out TI MD simulation with fully solvated model and without restraints for a pair of similar inhibitors of pharmaceutical interest binding with Erk kinase. The computations with BCC-AM1 and RESP charge gave $\Delta\Delta G$ of −1.1 and −1.6 kcal/mol, respectively, in good level of accuracy, compared with experimental value of −2.3 kcal/mol. We estimated the error of computed value to be 0.5 kcal/mol. Among the mutational steps, that of the disappearing/appearing vdw interaction in

the protein state had the largest uncertainty and required approximately 10 times more CPU time to obtain convergence, relative to the steps of switching charge on and off. In addition, residue-based free energy decomposition and the hydrogen bonding interactions between the ligands and this protein were analyzed and discussed. In addition, computations for five new analog ligands were carried out to predict $\Delta\Delta G$ as well and an analog ligand with polar –OH group at meta position of the benzene ring shown in the Table 4 was found to have better binding affinity than the Przp ligand examined previously by 2.0 kcal/mol. This result suggests that ligands of this kind are better inhibitors and should interest experimentalists for future enzyme and/or cell assays.

References

1. Cohen P (2002) Protein kinases—the major drug targets of the twenty-first century? Nat Rev Drug Discov 1:309–315
2. Mobley DL, Dill KA (2009) Binding of small-molecule ligands to proteins: "What You See" is not always "What You Get". Structure 17:489–498
3. Steinbrecher T, Labahn A (2010) Towards accurate free energy calculations in ligand protein-binding studies. Curr Med Chem 17:767–785
4. Deng YQ, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113:2234–2246
5. Jorgensen WL, Thomas LL (2008) Perspective on free-energy perturbation calculations for chemical equilibria. J Chem Theory Comput 4:869–876
6. Christ CD, Mark AE, van Gunsteren WF (2010) Feature article basic ingredients of free energy calculations: a review. J Comput Chem 31:1569–1582
7. Pearlman DA, Charifson PS (2001) Are free energy calculations useful in practice? A comparison with rapid scoring functions for the P38 map kinase protein system. J Med Chem 44:3417–3423
8. Yang WC, Pan YM, Fang L, Gao DQ, Zheng F, Zhan CG (2010) Free energy perturbation simulation on transition states and high-activity mutants of human butyrylcholinesterase for (−)-Cocaine hydrolysis. J Phys Chem B 114:10889–10896
9. Satpati P, Clavaguera C, Ohanessian G, Simonson T (2011) Free energy simulations of a Gtpase: Gtp and Gdp binding to archaeal initiation factor 2. J Phys Chem B 115:6749–6763
10. Pearlman DA, Charifson PS (2001) Improved scoring of ligand-protein interactions using Owfeg free energy grids. J Med Chem 44:502–511

11. Beierlein FR, Kneale GG, Clark T (2011) Predicting the effects of basepair mutations in DNA-protein complexes by thermodynamic integration. Biophys J 101:1130–1138

12. Fidelak J, Juraszek J, Branduardi D, Bianciotto M, Gervasio FL (2010) Free-energy-based methods for binding profile determination in a congeneric series of Cdk2 inhibitors. J Phys Chem B 114:9516–9524

13. Deng N-J, Zhang P, Cieplak P, Lai L (2011) Elucidating the energetics of entropically driven protein–ligand association: calculations of absolute binding free energy and entropy. J Phys Chem B 115:11902–11910

14. Kolář M, Ji Fanfrlík, Hobza P (2011) Ligand conformational and solvation/desolvation free energy in protein–ligand complex formation. J Phys Chem B 115:4718–4724

15. General IJ, Dragomirova R, Meirovitch H (2010) New method for calculating the absolute free energy of binding: the effect of a mobile loop on the avidin/biotin complex. J Phys Chem B 115:168–175

16. Elenewski JE, Hackett JC (2010) Free energy landscape of the retinol/serum retinol binding protein complex: a biological host–guest system. J Phys Chem B 114:11315–11322

17. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. J Phys Chem B 114:10235–10253

18. Ge XX, Roux B (2010) Absolute binding free energy calculations of sparsomycin analogs to the bacterial ribosome. J Phys Chem B 114:9525–9539

19. Shirts MR, Mobley DL, Chodera JD (2007) Alchemical free energy calculations: ready for prime time? Ann Rep Comput Chem 3:41–59

20. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA (2007) Predicting absolute ligand binding free energies to a simple model site. J Mol Biol 371:1118–1134

21. Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow CD, Shirts MR, Sorin EJ, Pande VS (2005) Direct calculation of the binding free energies of Fkbp ligands. J Chem Phys 123:084108

22. Wang JY, Deng YQ, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. Biophys J 91:2798–2814

23. Lee MS, Olson MA (2006) Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. Biophys J 90:864–877

24. Jayachandran G, Shirts MR, Park S, Pande VS (2006) Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. J Chem Phys 125:084901

25. Steinbrecher T, Case DA, Labahn A (2006) A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. J Med Chem 49:1837–1844

26. Steinbrecher T, Mobley DL, Case DA (2007) Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. J Chem Phys 127:214108

27. Krapf S, Koslowski T, Steinbrecher T (2010) The thermodynamics of charge transfer in DNA photolyase: using thermodynamic integration calculations to analyse the kinetics of electron transfer reactions. Phys Chem Chem Phys 12:9516–9525

28. Steinbrecher T, Hrenn A, Dormann KL, Merfort I, Labahn A (2008) Bornyl (3,4,5-Trihydroxy)-Cinnamate—an optimized human neutrophil elastase inhibitor designed by free energy calculations. Bioorg Med Chem 16:2385–2390

29. Pitera JW, van Gunsteren WF (2002) A comparison of nonbonded scaling approaches for free energy calculations. Mol Simul 28:45–65

30. Shirts MR, Pande VS (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. J Chem Phys 122:134508

31. Beutler TC, Mark AE, Vanschaik RC, Gerber PR, van Gunsteren WF (1994) Avoiding singularities and numerical instabilities in

32. Zacharias M, Straatsma TP, McCammon JA (1994) Separation-shifted scaling, a new scaling method for lennard-jones interactions in thermodynamic integration. J Chem Phys 100:9025–9031

33. Jiao D, Golubkov PA, Darden TA, Ren P (2008) Calculation of protein-ligand binding free energy by using a polarizable potential. Proc Natl Acad Sci USA 105:6290–6295

34. Michel J, Verdon ML, Essex JW (2007) Protein-ligand complexes: computation of the relative free energy of different scaffolds and binding modes. J Chem Theory Comput 3:1645–1655

35. Michel J, Essex JW (2010) Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. J Comput Aided Mol Des 24:639–658

36. Michel J, Foloppe N, Essex JW (2010) Rigorous free energy calculations in structure-based drug design. Mol Inf 29:570–578

37. Genheden S, Nilsson I, Ryde U (2011) Binding affinities of factor Xa inhibitors estimated by thermodynamic integration and MM/GBSA. J Chem Inf Model 51:947–958

38. Aronov AM, Baker C, Bemis GW, Cao JR, Chen GJ, Ford PJ, Germann UA, Green J, Hale MR, Jacobs M, Janetka JW, Maltais F, Martinez-Botella G, Namchuk MN, Straub J, Tang Q, Xie XL (2007) Flipped out: structure-guided design of selective pyrazolylpyrrole Erk inhibitors. J Med Chem 50:1280–1287

39. Aronov AM, Tang Q, Martinez-Botella G, Bemis GW, Cao JR, Chen GJ, Ewing NP, Ford PJ, Germann UA, Green J, Hale MR, Jacobs M, Janetka JW, Maltais F, Markland W, Namchuk MN, Nanthakumar S, Poondru S, Straub J, ter Haar E, Xie XL (2009) Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (Erk) using conformational control. J Med Chem 52:6362–6368

40. Zhan JH, Zhao X, Huang XR, Sun CC (2009) Molecular dynamics and free energy analyses of Erk2-Pyrazolylpyrrole inhibitors interactions: insight into structure-based ligand design. J Theor Comput Chem 8:887–908

41. Del Rio A, Baldi BF, Rastelli G (2009) Activity prediction and structural insights of extracellular signal-regulated kinase 2 inhibitors with molecular dynamics simulations. Chem Biol Drug Des 74:630–635

42. Kollman P (1993) Free-energy calculations—applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417

43. Case DA, Darden T, Cheatham TE, Simmerling C, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Wang B, Hayik S, Roitberg A, Seabra G, Kolossvary I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Vusarov S, Kovalenko A, Kollman PA (2010) Amber 11. University of California, San Francisco

44. Wang JM, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph 25:247–260

45. Kreyszig E (2005) Advanced engineering mathematics. Wiley, Hoboken, USA

46. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD,

Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03. Gaussian Inc., Wallingford, CT

47. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS (2011) Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol 21:150–160

48. Sadiq SK, Wright DW, Kenway OA, Coveney PV (2010) Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistant Hiv-1 proteases. J Chem Inf Model 50:890–905