Biophysical Journal Volume 103 July 2012 342–351

Insights from Free-Energy Calculations: Protein Conformational Equilibrium, Driving Forces, and Ligand-Binding Modes

Yu-ming M. Huang,[†] Wei Chen,[‡] Michael J. Potter,[‡] and Chia-en A. Chang[†*]
[†]Department of Chemistry, University of California, Riverside, California; and [‡]Verachem, Germantown, Maryland

ABSTRACT Accurate free-energy calculations provide mechanistic insights into molecular recognition and conformational equilibrium. In this work, we performed free-energy calculations to study the thermodynamic properties of different states of molecular systems in their equilibrium basin, and obtained accurate absolute binding free-energy calculations for protein-ligand binding using a newly developed M2 algorithm. We used a range of Asp-Phe-Gly (DFG)-in/out p38α mitogen-activated protein kinase inhibitors as our test cases. We also focused on the flexible DFG motif, which is closely connected to kinase activation and inhibitor binding. Our calculations explain the coexistence of DFG-in and DFG-out states of the loop and reveal different components (e.g., configurational entropy and enthalpy) that stabilize the apo p38α conformations. To study novel ligand-binding modes and the key driving forces behind them, we computed the absolute binding free energies of 30 p38α inhibitors, including analogs with unavailable experimental structures. The calculations revealed multiple stable, complex conformations and changes in p38α and inhibitor conformations, as well as balance in several energetic terms and configurational entropy loss. The results provide relevant physics that can aid in designing inhibitors and understanding protein conformational equilibrium. Our approach is fast for use with proteins that contain flexible regions for structure-based drug design.

INTRODUCTION

Molecular recognition (i.e., the binding of specific molecules by noncovalent interactions) is fundamentally important in chemistry. Recognition of chemical drugs and proteins is also important in medicine, and proteins frequently exist in multiple states in their equilibrium basin when a chemical drug associates to its protein target. Although experiments provide three-dimensional structures, they are limited to one static conformation. In addition, the mechanisms underlying ligand-protein recognition and dynamical equilibrium can only be addressed by further theoretical studies. In this study, we used a computational technique to calculate free energy for an attractive kinase system to study ligand-protein binding and protein conformational changes. Protein kinases regulate diverse cellular functions and play a key role in coordinating signal transduction cascades (1). All known protein kinases share the same tertiary structure, with an N-terminal lobe (N-lobe) and a C-terminal lobe (C-lobe). These two lobes and parts of the activation and glycine-rich loop form a binding site for ATP and kinase inhibitors (Fig. 1 *a*) (2). Nearly every eukaryotic protein kinase domain has a conserved Asp-Phe-Gly (DFG) motif. Flipping of the DFG motif connects catalytically active and inactive kinases (Fig. 1 *b*), but the mechanism of this flip is not precisely detailed. The Asp from the DFG motif is crucial for catalysis and contacts with the phosphates of ATP, but the function of the conserved Phe and Gly is not entirely clear. The movement of the DFG motif from the in to the out conformation opens up a hydrophobic pocket between the N-lobe and C-lobe. However, the movement and population of different states in the free kinases are not fully understood. In addition, it is of great interest to determine what drives different classes of inhibitors to bind to different DFG conformations. Discovering kinase inhibitors that target a particular kinase has been an important goal in structure-based drug design (3).

A protein kinase of interest as a drug target is p38α mitogen-activated protein kinase (MAPK), a serine-threonine kinase. The p38 MAPKs play a crucial role in regulating the production of proinflammatory cytokines. The enzymes have four isoforms and are drug targets for treating various inflammatory diseases, including rheumatoid arthritis, asthma, and cardiovascular disease (4). Several compounds can inhibit p38α MAPK in the nanomolar range, and rich structure data have revealed a range of various inhibitor binding modes with different DFG motif conformations (5). NMR studies have also used p38α as a model to study the DFG-in and DFG-out equilibrium.

To speed the discovery of new drugs, we need methods that will enable us to accurately compute thermodynamic differences between different states of a molecular system. Computational methods can help shed light on these issues. Molecular-dynamics and Monte Carlo methods with explicit waters are alternative methods to study protein dynamics (6–10). Such methods can also be used to study free-energy pathways, such as free-energy perturbation (FEP) and thermodynamic integration (TI), to compute the absolute or relative work of binding ligands to a protein (11–18). However, such approaches tend to be too computationally demanding for routine use in drug-design projects, and may be impractical for studying molecular systems that undergo considerable conformational changes. At the other

Submitted March 8, 2012, and accepted for publication May 29, 2012.
*Correspondence: chiaenc@ucr.edu
Editor: Bert de Groot.

0006-3495/12/07/0342/10 $2.00
doi: 10.1016/j.bpj.2012.05.046

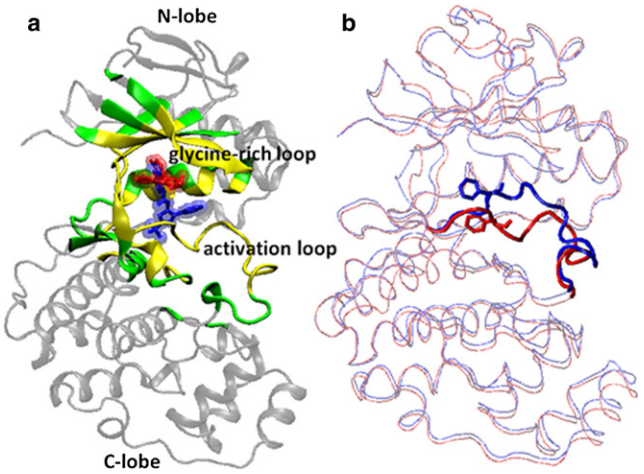

FIGURE 1 Overall view of the p38α structure. (*a*) Two inhibitors, SB218655 (*red*) and urea 16 (*blue*), bind to p38α. The flexible and rigid sets are shown in yellow and green, respectively. (*b*) The activation loop can adopt both DFG-in (*red*, PDB 1a9u) and DFG-out (*blue*, PDB 1w82) states for ligand binding.

end of the spectrum are the simplest and most efficient docking methods, which search for the single most stable ligand-binding mode in a protein pocket and estimate binding affinity with a scoring function (19–22). These fast methods can rapidly screen large compound databases but typically oversimplify binding. Endpoint free-energy methods, such as molecular mechanics Poisson-Boltzmann/surface area (MM/PBSA) and linear interaction energy (LIE) (23–29), are middle-of-the-road, relatively fast methods as compared with primer extension FEP or TI, but are more accurate and provide greater physical details than molecular docking methods.

In this study, we used the M2 algorithm to characterize the DFG motif equilibrium of free p38α, and changes in ligand and protein conformations with protein-inhibitor binding. In addition, we aimed to reveal energetic and entropic components that are important in ligand binding and compute the overall binding free energies by analyzing the association of various p38α inhibitors. M2 is an endpoint method that one can use to approximate the overall free energy of a molecular system by identifying a manageable set of conformations (local energy minima) and summing the computed configuration integral of each energy minimum (30,31). The binding free energies computed with M2 have shown encouraging agreement with experimental data in chemical host-guest systems (32,33). Here, we computed the conformational free energy of apo p38α and binding free energy of various type I and II p38α inhibitors using an optimal version of M2 for protein-ligand affinity calculation (34). The 30 inhibitors included a few sets of analogs with unavailable cocrystal structures (see Fig. S1 in the Supporting Material). The results accurately estimate the absolute binding free energy and capture many details about the physics that may be useful for understanding molecular recognition and designing inhibitors to enhance binding or specificity. Of note, our calculations also revealed the DFG-in and DFG-out conformations of free p38α, forces that stabilize different states, and the role of the conserved DFG glycine.

METHODS

Calculations with mining minima

As previously described (33,34), the second-generation mining minima method, M2, computes the standard free energy of a molecular system, such as a protein-ligand complex (*PL*), a free protein (*P*), or a free ligand (*L*). The standard binding free energy can be calculated as follows:

$$\Delta G^o = G^o_{PL} - G^o_P - G^0_L \qquad [1]$$

The standard free energy of each molecule ($G_X{}^o$) is calculated by a sum of contributions from *N* local wells *i*:

$$G^o_X = -RT \ln\left(\sum_{i=1}^{N} e^{-G^o_i/RT}\right)$$
$$G^o_i = -RT \ln\left(\frac{8\pi^2}{C^o} z_i\right) \qquad [2]$$
$$z_i = \int_i e^{-G_i/RT} dr$$

where G^o_i and z_i are respectively the standard free energy and the local configuration integral from distinct energy wells. Other details are provided in the Supporting Material Text1.

M2 is thus an endpoint method, because we only consider the free and final bound states of a molecular system without including the intermediate states during binding processes in Eq. 1. In brief, M2 consists of two parts: 1), an aggressive conformational search for distinct low-energy wells; and 2), an enhanced harmonic approximation for computing the configuration integral z_i of each well *i*. Both parts involve the Hessian matrix with respect to bond-angle-torsion coordinates, and our harmonic approximation accounts for anharmonicity of eigenvectors of the Hessian matrix with eigenvalues $<$ 2 kcal/mol/Å or 2 kcal/mol/rad. The correlation between different degrees of freedom (e.g., multiple dihedrals may rotate in concert or move with ligand translation/rotation) is captured in the Hessian matrix. The free energy, $G_X{}^\circ$, of a molecule or complex, $X = P, L$, or PL, is estimated by summing the local configuration integral z_i from distinct energy wells (conformations), as shown in Eq. 2. The probability of well *i* can be approximated on the basis of z_i, and thus the mean potential energy $<$U$>$ or solvation energy $<$W$>$ can be obtained. The configurational entropy at standard concentration can be computed as $-TS_{config}{}^\circ = G^\circ - <U+W>$. Note that the configurational entropy includes both a conformational part, which reflects the number of energy wells (conformations), and a vibrational part, which reflects the average width of the energy wells.

We used the VM2 package for the calculation, and performed two iterations for each ligand and four to 29 iterations for the free p38α and the complexes until the accumulated free energy converged (see Table S1 and Fig. S2 for details) (32,34). For a typical ligand-protein complex, it may take 12–14 h to finish one iteration using four cores of an Intel Xeon 2.4 GHz CPU. We removed any duplicated conformations before computing the final free energy. To reduce the computational cost, the program allows a user-defined region of the protein to be held rigid with flexibility of only a predefined binding-site region. All ligands are also flexible and can freely translate and rotate within the binding site without restraint. The rigid and flexible parts of p38α are the same for all systems, so we chose residues within 7 Å for all ligands and atoms on the activation

and glycine-rich loop as the flexible set (903 atoms included). The rigid set contained the residues within 5 Å of the flexible set (1666 atoms included; Fig. 1 *a* and Table S2), and the structure was the same in each calculation for the complexes and the free protein. Other atoms not included in these two sets were not considered during the M2 calculations.

Structure preparation and parameters

The initial structure of our p38α template was drawn from the crystal structure with PDB code 1a9u because it is a wild-type human p38α with no missing residues (35). The template shows a DFG-in conformation and a cocrystal structure with a ligand SB203580. A couple of DFG-out protein conformations were also provided as an initial structure, and the coordinates of the activation loop were taken from complex structures with a DFG-out protein conformation (PDB 1w82, 1w83, and 1wbv) to replace that in the template (36). After substituting the activation loop with a DFG-out conformation, we carried out a quick 100-step energy minimization with the flexible region of p38α by using the Amber10 package (37). The VM2 package was then used for further setup and free-energy calculations. The initial structures of the ligands were obtained by their cocrystal structures with p38α or were constructed with the use of Macromodel if crystal structures were not available. The Amber 99SB force field was used for proteins. The general amber force field (GAFF) for ligands was used for all systems. The final free energy considers the PBSA model, as described previously (34,38). Other details regarding the computation setup are provided in the Supporting Material Text2. All figure images were prepared with the use of VMD (39).

RESULTS AND DISCUSSION

Sampling of DFG-in and DFG-out conformations in free p38α protein

The M2 calculations revealed that the DFG-in and DFG-out conformations have similar computed free energies in apo p38α MAPK. The conformational free energies suggest that the two DFG positions have equal population, which is in good agreement with the dynamic equilibrium model suggested by NMR (5). The calculations include both energetic and entropic components. We focused on the residues around the ligand-binding site, glycine-rich loop, and activation loop to speed up the calculations (see Table S2 for details). Although the DFG-in conformations show ~3.7 kcal/mol stronger attraction energies than the DFG-out conformations, the attraction energy is compensated for by the less-favorable configuration entropy, yielding the same conformational free energies. Of note, the number of distinct conformations was similar in both states, and thus vibrational entropy plays a key role in governing the fluctuation of the free protein. The activation loop stayed in one major DFG-in conformation, except for small fluctuations between residues 172 and 174 (Fig. 2 *A*1), where coordinates are missing in human p38α x-ray structures (40,41). The glycine-rich loop shows moderate flexibility in both DFG-in and -out conformations, which is consistent with the structure alignment results with multiple ligand-p38α cocrystal structures. Interestingly, our calculations revealed several remarkably different DFG-out conformations, but comparatively more similar DFG-in conformations (Fig. 2, *A*1 and *A*2). Because Thr-180 is unphosphorylated in crystal structures in both DFG states, we kept it unphosphorylated in our calculations, and possible changes in protein fluctuation induced by phospho-Thr are not discussed in this study. Additional conformational searches showed that the Thr-180 positions of the most stable DFG-in and DFG-out conformations are the same as those in the crystal structures. Because our DFG-out loop conformations are diverse, and no existing experimental structures can be used for comparison, we computed the ϕ and ψ angles of Asp-168, and compared the angles with ligand-p38α cocrystal structures to ensure that our DFG-out conformations satisfied the consensus definition of the DFG-out structures (Fig. S4).

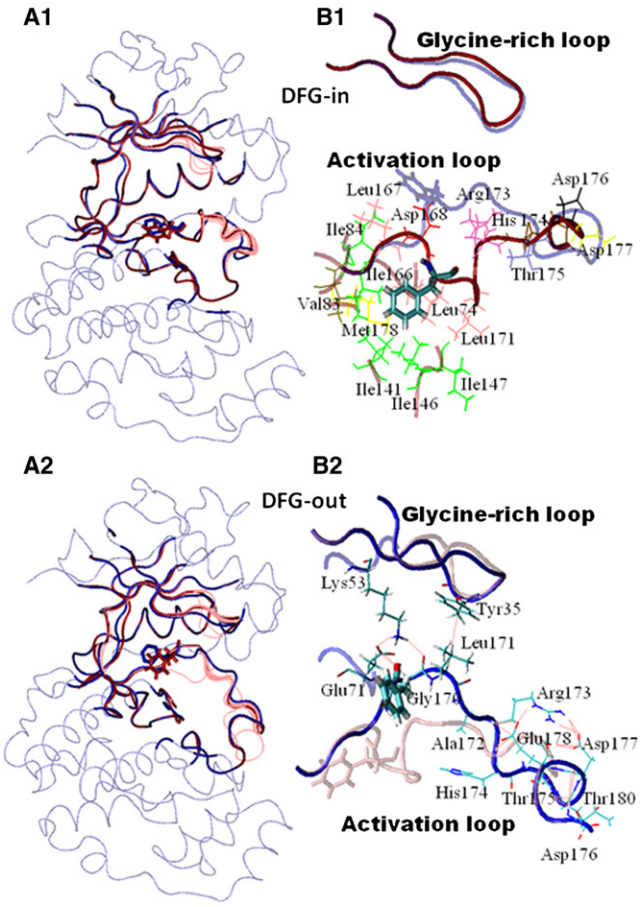

FIGURE 2 Sampled conformations of free DFG-in and DFG-out p38α. (*A1* and *A2*) Overview of DFG-in and DFG-out conformations, respectively. The crystal structure is in blue (see Fig. 1 for PDB codes), the sampled conformation with the lowest free energy is in red, and conformations with energies within 10RT above the global energy minimum are in pink. (*B1*) Phe-169 (bond form) is buried in the hydrophobic cluster in a DFG-in complex and the activation loop is exposed to the solvent (*colored thin line*). The red and light blue tubes represent DFG-in and DFG-out conformations, respectively. (*B2*) Side-chain arrangements in the DFG-out state. Key interactions between important atom pairs are shown by the pink dashed line. The light red and blue tubes represent DFG-in and DFG-out conformations, respectively.

Although both DFG-in and DFG-out conformations are equally stable, the major attraction forces that stabilize them differ (Table 1). The DFG-in states show stronger van der Waal attractions, and the DFG-out states have more favorable charge attractions. These two terms balanced well, so <U+W> has a similar value in both states. However, the two major attractions may not compensate perfectly during the conformational change processes, which results in energy barriers between the two states. This situation may explain why the NMR spectra showed a slow exchange between the DFG-in and DFG-out conformations (5). With the DFG-in conformation, the Phe-169 is buried by a hydrophobic pocket, which forms stronger nonpolar attractions than it does with the DFG-out conformation. The nonpolar spine residues Leu-86, Leu-75, Phe-169, and His-148 also form a rigid and hydrophobic stack (Fig. S5) (42). Polar residues in the activation loops, such as Arg-173, Thr-175, Asp-176, and Asp-177, expose the side chains to the solvent (Fig. 2 *B1*). However, Lys-53, the polar residue that is important for catalysis, forms a salt bridge with Glu-71 instead of exposing itself to the solvent, which also creates a preorganized pocket for ligand binding. In similarity to the DFG-in conformations, Lys-53 and Glu-71 also formed a stable salt bridge in the DFG-out conformations we found. However, this was a case of the residues in the activation loop being attracted to each other rather than the side chains being exposed to the solvent (Fig. 2 *B2*). Therefore, compared with the DFG-in conformations, the DFG-out conformations yield stronger Columbic attractions.

Because of the structural changes of the activation loop, a 180° flip of Phe-169 breaks the assembled nonpolar spine residues. Nevertheless, Phe-169 interacts with a nearby nonpolar residue, Leu-171, and Leu-171 further recruits Tyr-35 of the glycine-rich loop to form another nonpolar cluster to stabilize both activation and glycine-rich loops (Fig. 2 *B2*). The two nonpolar clusters may play a crucial role in stabilizing the DFG-out states. In addition, the flipped DFG-out motif forms a stable hydrogen (H)-bond between Asp-168 and Gly-170, but the H-bond cannot be seen in all of the DFG-in conformations (Fig. 3). Of note, the same pair of H-bonds was found in the ligand and DFG-out cocrystal structures, but not in the ligand-bound DFG-in p38α. The lack of side chain in glycine avoids potential side-chain interactions with other residues,

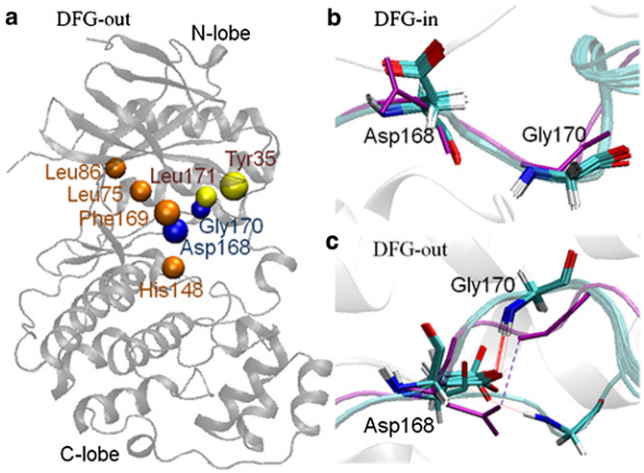

FIGURE 3 Positions and motions of key residues in DFG-out p38α. (*a*) Orange spheres: nonpolar spine residues; yellow spheres: nonpolar clusters (Tyr-35 and Leu-171); blue spheres: Asp-168–Gly-170 H-bond pair. (*b* and *c*) Low-energy conformations within 10RT of global energy minimum from computed DFG-in (*b*) and DFG-out (*c*) free p38α. Note that no H-bonds formed in the DFG-in state shown in *b*, and dashed lines in *c* indicate the H-bond between Asp-168 and Gly-170. Tubes and bonds in purple represent structures taken from PBD 1a9u (*b*) and 1w82 (*c*).

allowing an efficient bipositional switch and formation of an H-bond with Asp-168 to stabilize the DFG motif. Our findings contrast with those obtained for PKA and other serine/threonine kinases that form a Asp-168–Gly-170 H-bond in known active DFG-in structures (42). However, the Asp-168–Gly-170 H-bond in p38α further supports the importance of the conserved Gly-170, which may provide local stability for the DFG motif in DFG-in or DFG-out conformations in various kinases.

Binding free-energy calculations

The computed binding free energies, ΔG_{cal}, agree well with experimental binding affinities, ΔG_{exp}, for various types of inhibitors that bind to p38α with different conformations of the DFG motif (Table 2 and Fig. 4). This agreement validates the energy model used in this study and demonstrates that our search algorithm found all important conformations in the free and bound states of each ligand-p38α system. Here, we report the absolute binding affinities, not their related values. Binding modes similar to crystal structures

TABLE 1 Computed free energy, average potential, and solvation free energy (kcal/mol) of free p38α

DFG	PDB	G_{cal}	$<U+W>$	$-TS$	$<U_{Val}>$	$<U_{VDW}>$	$<W_{NP}>$	$<E_{np}>$	$<U_{Coul}>$	$<W_{PB}>$	$<E_{polar}>$
in	1a9u	−1349.40	−2712.99	1363.60	364.54	−358.22	45.55	−312.67	−2126.98	−1445.39	−3572.37
out	1w82	−1349.38	−2709.30	1359.91	380.43	−343.49	47.21	−296.28	−2244.37	−1356.58	−3600.95
combined		−1349.80	−2711.16	1361.36	372.40	−350.93	46.37	−304.56	−2185.06	−1401.45	−3586.51

Decomposed free energy, G_{cal}, from our calculations includes the average potential energy $<U+W>$, configurational entropy $-TS$, bonded terms $<U_{val}>$, Lennard-Jones energy $<U_{VDW}>$, nonpolar solvation free energy $<W_{NP}>$, Coulombic energy $<U_{Coul}>$, and PB solvation free energy $<W_{PB}>$. $<E_{np}>$ represents the sum of $<U_{VDW}>$ and $<W_{NP}>$; $<E_{ploar}>$ represents the sum of $<U_{Coul}>$ and $<W_{PB}>$.

TABLE 2 Binding free energy, average binding potential energy, and solvation free energy (kcal/mol) of p38α inhibitors

Ligand name	DFG	PDB	ΔG_{exp}	ΔG_{cal}	$\Delta(U+W)$	$-T\Delta S$	ΔU_{Val}	ΔU_{VDW}	ΔW_{NP}	ΔE_{np}	ΔU_{Coul}	ΔW_{PB}	ΔE_{polar}
Quinolinone 3	in	1ove	−12.6	−14.00	−34.25	20.25	11.44	−51.46	−5.46	−56.92	−56.31	65.76	9.44
Quinazolinone 2-(a)	in	1m7q_ang	−12.23	−14.28	−37.06	22.78	9.79	−44.40	−5.34	−49.74	−49.34	50.44	1.10
Quinazolinone 2	in	1m7q	−11.86	−14.29	−37.95	23.67	12.02	−38.09	−5.27	−43.36	−93.30	84.90	−8.39
Pyridol-pyrimidine 4	in	1ouy	−11.56	−11.98	−30.66	18.68	5.90	−50.08	−5.12	−55.20	−25.16	42.00	16.84
SB218655	in	1bmk	−10.50	−9.79	−32.46	22.68	7.71	−35.20	−4.26	−39.47	−76.10	73.61	−2.49
SB203580	in	1a9u	−10.11	−10.17	−31.23	21.06	10.72	−33.85	−4.35	−38.20	7.34	−12.88	−5.53
SB216995	in	1bl6	−9.39	−9.21	−28.60	19.39	−2.26	−36.34	−4.02	−40.36	−37.12	49.35	12.23
CHEBI:139669	in	1di9	−7.32	−6.25	−18.89	12.65	−4.06	−37.65	−4.15	−41.80	−9.43	34.61	25.18
CHEBI:139669-(a)	in	1di9_ang	−7.18	−6.06	−16.38	10.32	−6.78	−35.59	−3.57	−39.15	12.33	15.44	27.77
Indole 2	in	1w84	−6.16	−5.35	−24.66	19.30	3.39	−30.59	−3.59	−34.18	−26.91	31.26	4.35
Pyridine 5	in	1wbw	−6.02	−6.12	−22.57	16.45	−2.21	−36.25	−3.75	−40.00	−18.21	36.07	17.86
Pyridine 1-(a)	in	1w7h_ang	−5.47	−5.97	−20.30	14.34	0.46	−33.05	−3.43	−36.47	−20.98	34.91	13.93
Pyridine 1	in	1w7h	−3.99	−3.67	−16.23	12.57	5.22	−32.33	−3.18	−35.50	−35.69	47.95	12.26
BIRB-796	out	1kv2	−10.70	−10.57	−36.67	26.08	5.50	−61.09	−6.84	−67.93	−13.80	37.76	23.96
SB203580	out	3gcp	−10.11	−10.43	−27.83	17.38	−0.26	−45.66	−5.16	−50.82	−19.19	40.66	21.47
Pyridine 9	out	1w83	−9.70	−11.00	−34.03	23.02	1.68	−50.21	−4.83	−55.05	15.53	2.01	17.54
Pyrazolourea 1	out	3hv7	−9.35	−10.27	−35.71	25.42	9.35	−48.90	−5.34	−54.24	−41.26	48.65	7.39
Urea 16	out	1w82	−9.27	−8.71	−27.98	19.25	2.84	−47.95	−5.06	−53.01	20.82	−0.42	20.40
CHEBI:679695-(a)	out	3hv4_ang	−9.17	−9.18	−34.26	25.07	15.75	−56.71	−6.13	−62.84	−8.93	19.96	11.03
CHEBI:679695	out	3hv4	−8.99	−8.40	−32.75	24.34	10.58	−57.95	−6.21	−64.16	−0.79	19.82	19.04
Indole 24	out	1wbt	−8.94	−9.36	−30.11	20.74	1.39	−59.55	−5.72	−65.26	9.57	22.40	31.97
Pyridine 10	out	1wbn	−8.92	−7.02	−32.74	25.71	17.35	−47.83	−5.45	−53.28	−5.36	6.76	1.40
Indole 23	out	1wbs	−8.57	−8.59	−26.45	17.84	−1.68	−52.69	−4.36	−57.04	26.17	4.32	30.49
Pyrazolourea 1-(a)	out	3hv7_ang	−8.52	−9.38	−31.99	22.60	9.74	−47.31	−5.16	−52.46	3.81	5.13	8.94
Pyrazolourea 1-(b)	out	3hv7_ang	−8.13	−9.15	−28.71	19.55	5.99	−48.19	−5.24	−53.44	12.35	4.60	16.95
Pyrazolourea 1-(c)*	out	3hv6	−7.79	−7.27	−30.44	23.16	8.03	−60.91	−6.67	−67.59	−31.89	59.21	27.32
Diaryl urea 1	out	1kv1	−7.22	−8.23	−25.47	17.22	6.62	−41.74	−4.29	−46.03	−6.78	18.94	12.16
Pyrazolourea 1-(d)	out	3hv7_ang	−6.60	−6.57	−20.94	14.34	15.98	−44.74	−4.57	−49.31	−17.81	28.41	10.61
indole 22	out	1wbv	−5.24	−5.36	−21.58	16.20	−3.77	−44.92	−3.46	−48.38	18.02	10.76	28.78
Pyrazolourea 1-(e)	out	3hv7_ang	nb	−1.53	−28.39	26.84	8.17	−42.82	−5.58	−48.40	−7.51	17.56	10.05

Notations as in Table 1. The binding free energy was computed by $\Delta G_{cal}=G_{complex}-G_{free\ protein\ combine}-G_{free\ ligand}$. Each decomposed energy is obtained by $\Delta E_{DFG\text{-}in}=E_{DFG\text{-}in\ complex}-E_{DFG\text{-}in\ free\ protein}-E_{DFG\text{-}in\ free\ ligand}$ and $\Delta E_{DFG\text{-}out}=E_{DFG\text{-}out\ complex}-E_{DFG\text{-}out\ free\ protein}-E_{DFG\text{-}out\ free\ ligand}$. The binding free energy from experimental results is shown as ΔG_{exp}. The PDB labeled as XXX_ang is the analog from XXX.
*The ligand, pyrazolourea 1-(c), is also called CHEBI:266984.

can also be found in our calculations. In addition to a single ligand-binding mode available with experiments, our calculations provide details concerning all possible stable ligand-binding modes in the protein-binding site, ligand and protein conformational changes in their free and bound states, and driving forces, including energetic and entropic components (see examples in the Supporting Material Text3).

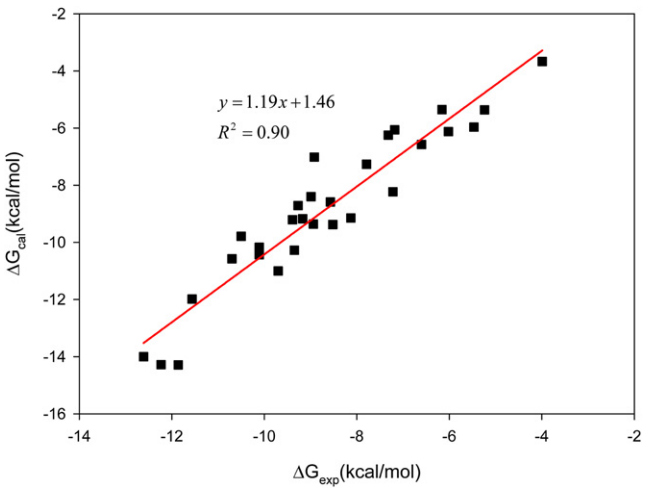

FIGURE 4 Calculated versus experimental binding free energies (kcal/mol) for p38α inhibitors.

Because our free-energy calculations suggest that both DFG states are equally stable, and experiments showed slow equilibrium (kHz) versus faster association rate constants ($>10^4$/Ms), we propose that p38α inhibitors commonly bind to a preorganized DFG-in or DFG-out conformation (5). This assumption is further supported by previous surface plasmon resonance measurements that showed that the kinetics of binding were largely unaffected by the bound kinase conformation (43). Therefore, induced conformational changes of the DFG loop by inhibitors are not discussed here. To more appropriately reveal the driving forces for ligand binding to a preorganized DFG motif, we separated the two DFG forms (Table 2). For possible induced DFG flipping during the binding processes, we computed the energy changes (Table S5) but did not address the driving forces in these cases. Determining whether any term is significantly different to distinguish DFG-in and DFG-out inhibitors is of interest. Because the configuration

entropy is larger in DFG-out than DFG-in conformations in apo p38α, if binding an inhibitor brings both complexes to similar flexibility, are more notable entropy loss might be expected in DFG-out binders. However, our calculations revealed no general trend of configuration entropy loss for DFG-in versus DFG-out inhibitors, as previously indicated by experimental thermodynamic data (5). For both types of inhibitors, in general, tight binders show a larger entropy loss that is compensated for by stronger attraction forces (Table 2). Of note, the values of our entropy changes cannot be directly compared with experimental data because this term reports conformational and vibrational entropy of molecules but no solvent entropy. The M2 calculations generated more distinct conformations of most free DFG-out ligands within 1.8 kcal/mol (3RT) of the most stable free conformer, which may due to the larger DFG-out ligands selected (Table S4). Although dropping ligand conformers in the bound state plays a role in the entropy loss, the main contribution is from dihedral vibration. Interestingly, ligands of similar size and with the same number of rotatable bonds can generate quite different numbers of stable low conformations in their free forms. For example, both quinolinone 3 and quinazolinone 2 ligands have three rotatable bonds and similar size but 17 and 57 low-energy conformations, respectively (Table S5 and Fig. S1). In addition, inhibitors show considerably different conformations in their free and bound states. Therefore, assuming that the most stable conformation in the inhibitor's free form is also the most stable in the bound state may produce errors in modeling.

In general, nonpolar interactions with the protein are stronger with DFG-out than with DFG-in ligands because the DFG-out inhibitors can contact with residues such as Leu-74, Leu-75, Ile-84, Leu-86, Leu-104, and Leu-167 in the hydrophobic cluster, and the compounds are usually larger. Although they are in a hydrophobic environment, the side chains of nonpolar residues may form multiple alternative contacts with neighboring nonpolar residues and provide a fluid-like interface (44), and the side chains in the hydrophobic cluster show only a few dominant conformers in the low-energy structures (Fig. S7). The side-chain conformations remain in their major rotameric states in the complex conformations, with few induced side-chain conformations in the hydrophobic cluster. Therefore, one could select the most populated side-chain conformations when performing in silico compound screening with molecular docking. The DFG-in inhibitors demonstrate more favorable Columbic attractions. Although this term is mostly compensated for by the desolvation penalty, the DFG-in ligands still show stronger polar interactions than the DFG-out ligands in most cases. The Asp-168 in the DFG motif plays a role in interacting with the DFG-in inhibitors, but such interactions are completely diminished in the DFG-out ligand-protein complexes.

SB203580 binding to both DFG-in and -out conformations

Although most inhibitors bind to DFG-in or DFG-out conformations, and the ligand-p38α interactions lock the DFG-in/DFG-out interconversion, a special inhibitor, SB203580, can bind to both conformations. Our binding free-energy calculations illustrated that the SB203580 compound binds equally well to both DFG-in and DFG-out states with a standard free energy of -10.17 and -10.43 kcal/mol, respectively. NMR revealed an equal population of both binding positions (5). The computed ΔG that considers SB203580 binding to both DFG conformations is ~ -10.32 kcal/mol, which also agrees well with experimental measurements (-10.11 kcal/mol (35)). In addition, the calculations reveal that the positions of SB203580 in the binding pocket do not hinder the in/out exchange of the DFG motif (Fig. 5 *a*). Our calculations easily reveal the space organization that allows the interconversion of the DFG motif, which is less clear in the crystal structure (Fig. 5 *c*). Of note, the SB203580 and DFG-out cocrystal structure (PDB 3gcp) has an A172C mutated residue (45). The position of Cys-172 in the crystal structure differs from that of other structures with wild-type Ala-172, and

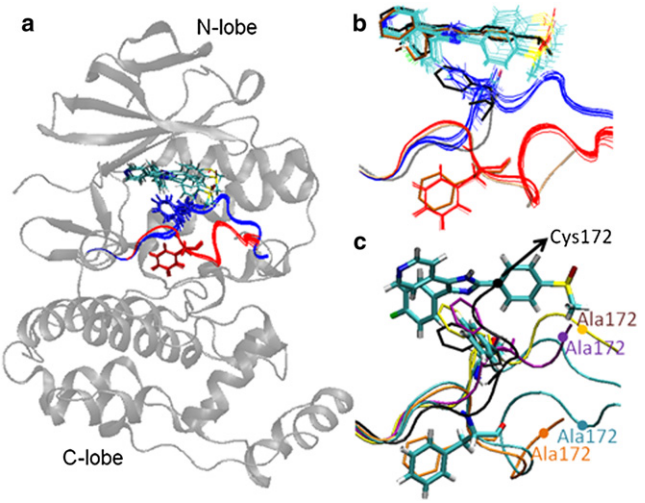

FIGURE 5 Complex conformations of SB203580 and p38α with both DFG-in and DFG-out conformations. (*a*) Sampled ligand-binding modes (*colored bond*), DFG-in (*red*), and DFG-out (*blue*) conformations, which illustrate space for interconverting the activation loop. (*b*) Low-energy conformations within 10RT of the global energy minimum from the calculations, showing the fluctuation of Phe-169 in the DFG-out state and the highly flexible methylsulfinylphenyl group. Sampled DFG-in and DFG-out conformations are red and blue, respectively, and structures from PDB IDs 1a9u and 3gcp are orange and black, respectively. (*c*) The sampled activation loop and Phe-169 in both DFG-in and DFG-out conformations are in cyan; crystal structures with PDB IDs, 3hv7, 1w82, 3gcp, and 1a9u are in purple, yellow, black, and orange, respectively. Note that the structure in complex with SB203580 from PDB 3gcp is the Cys-172 mutant (*black arrow*) and shows different conformation (*black*) from our samples and other experimental structures.

the subsequent loop structure is missing. Our calculations involved the wild-type and full-length activation loop, and the sampled conformations showed the same loop arrangement as for other DFG-out cocrystal structures with tight binders, such as compounds pyrazolourea 1 (PDB 3hv7) and urea 16 (PDB 1w82; Fig. 5 *c*) (36,46).

Despite the motions of the DFG motif, SB203580 stays in the same position in the binding pocket (Fig. 5). By analyzing the ligand-binding modes, we can see that inhibitors that can target both conformations of the DFG motif have the potential to enhance the binding affinities. The pyridine and fluorophenyl rings contact nicely with p38α, and the solvent-exposed methylsulfinylphenyl group is the only functional group that significantly adjusts with different DFG conformations (Fig. 5 *b*). Similarly to other DFG-out ligands, Asp-168 moves toward the solvent and breaks the charge interactions with the ligand, and Phe-169, together with Tyr-35, generates nice nonpolar attractions with the ligand. These findings are reflected in Table 2, and SB203580 shows more negative van der Waal attractions for binding to the DFG-out than DFG-in conformation, although the ligand does not have extensive contacts with the hydrophobic cluster. In addition, Tyr-35 and Leu-171 form nice nonpolar contacts in free DFG-out conformations that occlude the binding site. Therefore, significant rearrangements in Leu-171 and the glycine-rich loop open the gate for SB203580 binding. This binding position reveals the cross talk between the DFG motif and glycine-rich loop, and the ligand may play a role in mediating the interplay of the two loops (45). In contrast, with free DFG-in conformations, the two loops are open for the ligand, and the easy-access binding site may contribute to a fast binding of SB203580.

Binding-affinities prediction for analogs

Ligand discovery processes involve the synthesis of many analogs based on a lead compound to assess structure-activity relationships. A series of analogs provide details of molecular interactions to guide lead optimization, but most analogs do not have cocrystal structures with the protein. In this study we computed the binding free energies of a few series of analogs with unavailable experimental structures. All of the analogs studied here are also very similar to their template compounds (e.g., quinazolinone 2, CHEBI:139669, Pyridine 1, pyrazolourea 1, and CHEBI:679695 and their analogs (Fig. S1)). Because those analogs are merely different from a functional group, the predicted binding modes are similar to each other, as expected. For example, the unique naphthyl moiety of the pyrazolourea 1 compound locates in the hydrophobic cluster, and the nitrogen and oxygen atoms of the urea group show electrostatic attractions with Glu-71 and Asp-168. Similarly, the benzene ring of the analog stays in the hydrophobic cluster, and the charge interactions between the urea group and p38α also hold (Fig. 6 *b*). However, because the functional groups are not identical, the protein side chains and ligand adjust to each other to optimize their interactions. Of note, as compared with the free p38α with DFG-out conformations, the loop of Ile-84 and Phe-169 rotate farther to create a space for ligand binding. The positions of Phe-169 in the complex conformations found with M2 are varied and differ from those shown in crystal structures, which suggests that the activation loop has high flexibility during ligand binding.

The calculations can successfully distinguish small differences in the analogs and reflect them in the free-energy calculations (e.g., see the pyrazolourea analogs in Table S1). Interestingly, the entropic term may play a crucial role in successfully ranking the computed binding affinities, such as in pyrazolourea 1 analogs. The worst binder in this series, pyrazolourea 1-(e), retains attractive interaction energies as compared with other analogs. However, the significantly unfavorable entropy loss results in an extremely weak binding free energy, which agrees with experiments (46). Of note, the large entropic penalty cannot be easily observed from the ligand binding modes (Fig. 6 *c*), which suggests that detailed free-energy calculations are necessary for accurate binding prediction. In addition, the individual energetic terms offer more information about the driving forces to further improve binding. For example, replacing the thiomethyl group ($S\text{-}CH_3$) group by bromine in the CHEBI:139669 analog reduces both polar and nonpolar attractions. Because the data in Table 2 include changes in

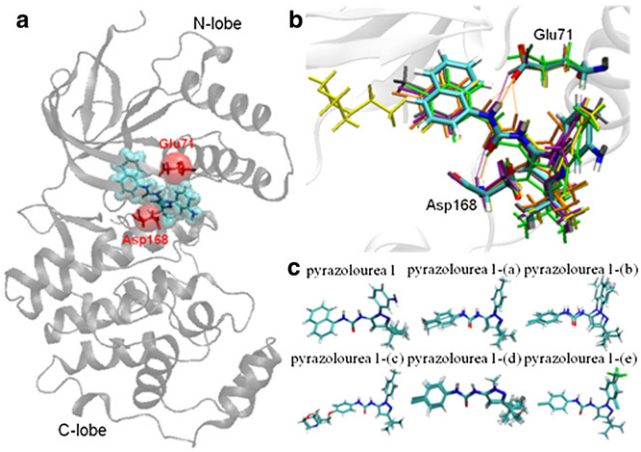

FIGURE 6 Binding modes of pyrazolourea 1 and its analogs. (*a*) Stable charge interactions among Glu-71 and Asp-168 (*red*) and the ligand (*cyan*). (*b*) Global energy minimum found in the complex. The template pyrazolourea 1 is shown as a colored bond. Gray, purple, yellow, orange, and green represent the pyrazolourea 1 analogs (a), (b), (c), (d), and (e), respectively. The charge interactions among Glu-71, Asp-168, and the inhibitors are represented as dashed lines. (*c*) Low-energy conformations within 10RT of the global energy minimum of bound inhibitors. Each analog shows flexibility in different function groups; however, changes in configurational entropy upon binding cannot be easily predicted from looking at the ligand-binding modes.

intra- and intermolecular interactions, binding to different ligands can sometimes yield different side-chain rearrangements that noticeably change the intramolecular interactions. For instance, quinazolinone 2 and its analog can both form nice interactions with Thr-106, Met-109, and Gly-110, and the difluorophenyl/chlorophenyl ring is oriented into a hydrophobic environment. Nevertheless, the small difference in the difluorophenyl and chlorophenyl ring reorganizes the protein side chains nearby. As a result, Asp-168 moves closer to the difluorophenyl ring and forms an H-bond with Lys-53, which is not seen in either the free p38α conformations or the complex with the chlorophenyl ring analog (Fig. S8). The charge interactions yield considerably more stable polar attractions in the complex form in the quinazolinone 2 complex. Without considering the energy changes from p38α, the changes in intermolecular interactions are similar in both ligands.

General comments on the method

Overall, the accuracy of our calculations is promising and suggests that the method presented here will be broadly useful for interpreting experimental studies of protein conformations and ligand-protein binding. However, the rigorous calculations are inevitably highly system-dependent for the setup. Although our free-energy calculation employs standard preparation steps as described in Materials and Methods, careful assessments and manual modifications are required to prepare the initial conformations. For example, to speed up the calculations, only a predefined binding-site region is flexible; thus, we need to be sure that the rigid part has minimal changes in its free and bound states. In addition, the computed Hessian matrix is very memory-demanding and we used computers with 4GB RAM; thus, the flexible region is set to be $<$1330 atoms. Proper selection of flexible residues is very important because the conformational flexibility can significantly affect the accuracy of computing configuration entropy. In this study, we analyzed $>$100 crystal structures of this system to select the flexible region and our template structures. For other systems with only a handful of experimental structures, other simulation tools, such as molecular-dynamics simulations, may be performed to select the flexible parts. To sample low-energy conformations for the ligand-protein complexes more efficiently and thoroughly, we did not rely solely on crystal structures. We selected multiple apo p38α conformations found by our conformational free-energy calculations for the protein, and docked inhibitors in the binding site to generate several initial complex conformations.

Water molecules may play an important role in ligand binding. All of the inhibitors chosen in this study do not have bridging waters when they are forming a ligand-protein complex. However, for systems with key bridging waters, a hybrid implicit-explicit model should be used. In this study we used the GAFF force field for ligands, and the force-field parameters can reproduce the experimental structures of p38α inhibitors in general. However, for a few ligands with unique functional groups, we manually set the equilibrium values of bond angles to be the same as those measured in crystal structures, as detailed in the Supporting Material Text2 and Fig. S2. For uncommon ligands without any experimental reference structure, it may be necessary to test different force fields, such as the OPLS and CHARMM force fields.

CONCLUSIONS

Using p38α as a model system, we computed the conformational and binding free energies of the free protein of p38α and various p38α inhibitors using the M2 algorithm. Because of the large-scale motions and flexibility of the loop regions, it is experimentally challenging to obtain the structures of free kinases. Our computational work successfully sampled multiple conformations of both DFG-in and DFG-out states of free wild-type human p38α and revealed the conformational fluctuations of the activation loop, glycine-rich loop, and DFG motif. The free-energy calculations showed that both DFG-in and -out states are equally populated in free p38α, which agrees with NMR results. In addition, we suggest that potential interactions and the conserved Phe-169 and Gly-170 have a role in stabilizing both states. The calculations also indicate that the vibrational entropy, not just the number of distinct conformations, plays an important role in the conformational ensemble. Further use of the M2 method may identify intermediate states during the DFG flip, and provide a more complete picture of pathways and free-energy barriers for switching kinase activation and inactivation.

Our results also demonstrate the high predictive power of binding affinities for analogs that lack experimental structures. As noted in previous theoretical studies for other systems, the net free-energy changes are balances of notably larger energy and entropy contributions (32,33,44). In addition to providing overall binding free energies, our calculations illustrate key differences in energetic and configurational entropic terms when a protein and inhibitor are in their free and bound states, which should be useful for designing structure-based inhibitors. Finally, the computational work can help elucidate experimental observations. For example, examination of the bound complex of the inhibitor SB203580 revealed how ligand binding does not affect flipping of the DFG motif.

SUPPORTING MATERIAL

Additional details and examples, eight figures, five tables, and references are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00626-1.

We thank Myungshim Kang and Chris Roberts for helpful discussions, and Simon P. Webb for preparing input files.

This work was supported by grants from the National Science Foundation (MCB-0919586); the University of California, Riverside (to C.A.C.); and the National Institutes of Health (5R44GM075350 to M.J.P.).

REFERENCES

1. Taylor, S. S., and A. P. Kornev. 2011. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* 36:65–77.
2. Huse, M., and J. Kuriyan. 2002. The conformational plasticity of protein kinases. *Cell.* 109:275–282.
3. Bert, K., M. Gerhard, and H. Michael, editors. 2011. Protein Kinases as Drug Targets. Wiley-VCH, Weinheim.
4. White, A., C. A. Pargellis, …, B. T. Farmer, 2nd. 2007. Molecular basis of MAPK-activated protein kinase 2:p38 assembly. *Proc. Natl. Acad. Sci. USA.* 104:6353–6358.
5. Vogtherr, M., K. Saxena, …, H. Schwalbe. 2006. NMR characterization of kinase p38 dynamics in free and ligand-bound forms. *Angew. Chem. Int. Ed. Engl.* 45:993–997.
6. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
7. Frembgen-Kesner, T., and A. H. Elcock. 2006. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. *J. Mol. Biol.* 359:202–214.
8. Shan, Y., M. A. Seeliger, …, D. E. Shaw. 2009. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. USA.* 106:139–144.
9. Filomia, F., F. De Rienzo, and M. C. Menziani. 2010. Insights into MAPK p38α DFG flip mechanism by accelerated molecular dynamics. *Bioorg. Med. Chem.* 18:6805–6812.
10. Fatmi, M. Q., and C. E. A. Chang. 2010. The role of oligomerization and cooperative regulation in protein function: the case of tryptophan synthase. *PLOS Comput. Biol.* 6:e1000994.
11. Deng, Y. Q., and B. Roux. 2009. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B.* 113:2234–2246.
12. Christ, C. D., A. E. Mark, and W. F. van Gunsteren. 2010. Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* 31:1569–1582.
13. Zuckerman, D. M. 2010. Statistical Physics of Biomolecules: An Introduction. CRC Press, Boca Raton, FL.
14. Gilson, M. K., and H. X. Zhou. 2007. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* 36:21–42.
15. Mobley, D. L., A. P. Graves, …, K. A. Dill. 2007. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* 371:1118–1134.
16. Chodera, J. D., D. L. Mobley, …, V. S. Pande. 2011. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* 21:150–160.
17. Guo, Z., C. L. Brooks, and X. Kong. 1998. Efficient and flexible algorithm for free energy calculations using the lambda-dynamics approach. *J. Phys. Chem. B.* 102:2032–2036.
18. Pohorille, A., C. Jarzynski, and C. Chipot. 2010. Good practices in free-energy calculations. *J. Phys. Chem. B.* 114:10235–10253.
19. Wong, C. F. 2008. Flexible ligand-flexible protein docking in protein kinase systems. *Biochim. Biophys. Acta.* 1784:244–251.
20. Kufareva, I., and R. Abagyan. 2008. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *J. Med. Chem.* 51:7921–7932.
21. Andricopulo, A. D., L. B. Salum, and D. J. Abraham. 2009. Structure-based drug design strategies in medicinal chemistry. *Curr. Top. Med. Chem.* 9:771–790.
22. Clark, M., S. Meshkat, …, J. S. Wiseman. 2009. Fragment-based computation of binding free energies by systematic sampling. *J. Chem. Inf. Model.* 49:1901–1913.
23. Swanson, J. M. J., R. H. Henchman, and J. A. McCammon. 2004. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* 86:67–74.
24. Gallicchio, E., M. Lapelosa, and R. M. Levy. 2010. The binding energy distribution analysis method (BEDAM) for the estimation of protein-ligand binding affinities. *J. Chem. Theory Comput.* 6:2961–2977.
25. Aqvist, J., C. Medina, and J. E. Samuelsson. 1994. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 7:385–391.
26. Genheden, S., and U. Ryde. 2011. Comparison of the efficiency of the LIE and MM/GBSA methods to calculate ligand-binding energies. *J. Chem. Theory Comput.* 7:3768–3778.
27. Lyne, P. D., M. L. Lamb, and J. C. Saeh. 2006. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* 49:4805–4808.
28. Joshi, M., J. O. Ebalunode, and J. M. Briggs. 2009. Computational insights into the interaction of the anthrax lethal factor with the N-terminal region of its substrates. *Proteins.* 75:323–335.
29. Kollman, P. A., I. Massova, …, T. E. Cheatham, 3rd. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.
30. Chang, C. E., and M. K. Gilson. 2003. Tork: conformational analysis method for molecules and complexes. *J. Comput. Chem.* 24:1987–1998.
31. Chang, C. E., M. J. Potter, and M. K. Gilson. 2003. Calculation of molecular configuration integrals. *J. Phys. Chem. B.* 107:1048–1055.
32. Chang, C. E., and M. K. Gilson. 2004. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* 126:13156–13164.
33. Chen, W., C. E. Chang, and M. K. Gilson. 2004. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys. J.* 87:3035–3049.
34. Chen, W., M. K. Gilson, …, M. J. Potter. 2010. Modeling protein-ligand binding by mining minima. *J. Chem. Theory Comput.* 6:3540–3557.
35. Wang, Z. L., B. J. Canagarajah, …, E. J. Goldsmith. 1998. Structural basis of inhibitor selectivity in MAP kinases. *Structure.* 6:1117–1128.
36. Gill, A. L., M. Frederickson, …, H. Jhoti. 2005. Identification of novel p38α MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* 48:414–426.
37. Case, D. A., T. E. Cheatham, …, R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
38. Cai, Q., M.-J. Hsieh, …, R. Luo. 2010. Performance of nonlinear finite-difference Poisson-Boltzmann solvers. *J. Chem. Theory Comput.* 6:203–211.
39. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
40. Wilson, K. P., M. J. Fitzgibbon, …, M. S. Su. 1996. Crystal structure of p38 mitogen-activated protein kinase. *J. Biol. Chem.* 271:27696–27700.
41. Patel, S. B., P. M. Cameron, …, G. Scapin. 2004. Lattice stabilization and enhanced diffraction in human p38 α crystals by protein engineering. *Biochim. Biophys. Acta.* 1696:67–73.
42. Kornev, A. P., N. M. Haste, …, L. F. Eyck. 2006. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. USA.* 103:17783–17788.

43. Angell, R. M., T. D. Angell, …, A. L. Walker. 2008. Biphenyl amide p38 kinase inhibitors 4: DFG-in and DFG-out binding modes. *Bioorg. Med. Chem. Lett.* 18:4433–4437.

44. Chang, C. E. A., W. A. McLaughlin, …, J. A. McCammon. 2008. Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association. *Proc. Natl. Acad. Sci. USA.* 105:7456–7461.

45. Simard, J. R., M. Getlik, …, D. Rauh. 2009. Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *J. Am. Chem. Soc.* 131:13286–13296.

46. Klüter, S., C. Grütter, …, D. Rauh. 2010. Displacement assay for the detection of stabilizers of inactive kinase conformations. *J. Med. Chem.* 53:357–367.