JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS
Terms of Use

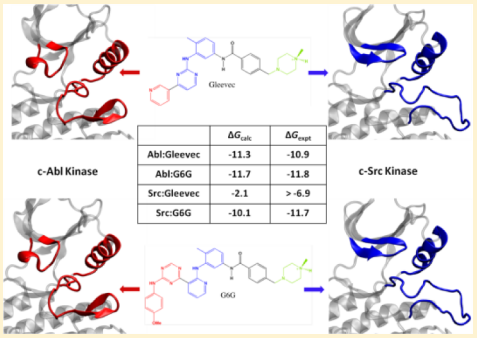
Computational Study of Gleevec and G6G Reveals Molecular Determinants of Kinase Inhibitor Selectivity

Yen-Lin Lin,[†] Yilin Meng,[†] Lei Huang, and Benoît Roux*

Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, The University of Chicago, 929 57th Street, Chicago, Illinois 60637, United States

Supporting Information

ABSTRACT: Gleevec is a potent inhibitor of Abl tyrosine kinase but not of the highly homologous c-Src kinase. Because the ligand binds to an inactive form of the protein in which an Asp-Phe-Gly structural motif along the activation loop adopts a so-called DFG-out conformation, it was suggested that binding specificity was controlled by a "conformational selection" mechanism. In this context, the binding affinity displayed by the kinase inhibitor G6G poses an intriguing challenge. Although it possesses a chemical core very similar to that of Gleevec, G6G is a potent inhibitor of both Abl and c-Src kinases. Both inhibitors bind to the DFG-out conformation of the kinases, which seems to be in contradiction with the conformational selection mechanism. To address this issue and display the hidden thermodynamic contributions affecting the binding selectivity, molecular dynamics free energy simulations with explicit solvent molecules were carried out. Relative to Gleevec, G6G forms highly favorable van der Waals dispersive interactions upon binding to the kinases via its triazine functional group, which is considerably larger than the corresponding pyridine moiety in Gleevec. Upon binding of G6G to c-Src, these interactions offset the unfavorable free energy cost of the DFG-out conformation. When binding to Abl, however, G6G experiences an unfavorable free energy penalty due to steric clashes with the phosphate-binding loop, yielding an overall binding affinity that is similar to that of Gleevec. Such steric clashes are absent when G6G binds to c-Src, due to the extended conformation of the phosphate-binding loop.



	ΔG_{calc}	ΔG_{expt}
Abl:Gleevec	-11.3	-10.9
Abl:G6G	-11.7	-11.8
Src:Gleevec	-2.1	> -6.9
Src:G6G	-10.1	-11.7

INTRODUCTION

Kinases are major therapeutic targets for a variety of diseases such as cancer, diabetes, and inflammation. In recent years, many small-molecule inhibitors of kinases have been developed as possible treatments of these diseases.[1,2] Gleevec (also known as STI-571 or Imatinib),[3−5] a novel drug against chronic myelogenous leukemia (CML) caused by constitutively activated Abl tyrosine kinase, offers a particularly interesting case. It exhibits highly inhibitory activity for Abl (K_i = 13 nM;[6] IC_{50} = 11 nM[7]) but has a much reduced affinity for most other closely related homologous kinase targets, such as c-Src (K_i = 31 100 nM;[6] IC_{50} > 10 000 nM[7]). The disparity in binding affinity is surprising, given that Abl and c-Src have a high degree of sequence identity[8,9] (47%) and the binding pocket is lined by similar amino acids, providing very similar molecular contacts.[6,10]

The mechanism underlying the binding specificity of Gleevec for different kinases is still being debated. Quantitatively, the specificity of the binding process is affected by two different factors. First, a short motif comprised of the conserved residues Asp-Phe-Gly (DFG) in the binding pocket of the kinase undergoes a conformational transition via a 180° rotation, referred to as DFG flip, switching the enzyme from an active (DFG-in) to an inactive (DFG-out) state.[11] Gleevec binds only to the inactive DFG-out state. Therefore, whether a given kinase can readily adopt this conformation has an important impact on the apparent binding affinity of Gleevec. The accessibility and relative stability of the DFG-in and DFG-out conformations give rise to a "conformational selection" mechanism for the binding of Gleevec. A second factor affecting the specificity is the affinity of Gleevec for the DFG-out conformation of a given kinase. This directly depends on the magnitude of the interactions of the ligand with the residues lining the binding pocket of a given kinase. Differences regarding both the conformational selection and the ligand−protein interactions have been proposed to explain and rationalize the binding specificity of Gleevec for Abl over c-Src. It is sometimes argued that the key to the binding specificity is that the DFG-out conformation (required for the drug binding) is allowed in Abl but forbidden in c-Src.[12] However, crystal structures of Abl and c-Src in complex with Gleevec[6,10] and with its derivatives[7,13] determined subsequently, as well as the observations of single-point mutagenesis,[6] raised the possibility that differences in the ligand−protein interactions could be at the origin of kinase inhibitor binding specificity rather than conformational selection. According to this view, c-Src does not incur a large energetic penalty for adopting the DFG-out conformation required for the ligand binding, and distinct protein−ligand interactions give

Received: April 30, 2014
Published: September 22, 2014

   dx.doi.org/10.1021/ja504146x | J. Am. Chem. Soc. 2014, 136, 14753−14762

rise to the observed binding specificity.[14] Such conflicting views about the mechanism of binding specificity cannot be easily resolved with the limited information currently available from experiments.

Additional insight about the mechanisms of specificity can be gleamed by considering additional ligands designed to probe the same binding site. The ligand G6G studied by Maly and collaborators[7] is especially interesting. It shares a similar central chemical scaffold with Gleevec (Figure 1) but displays equally

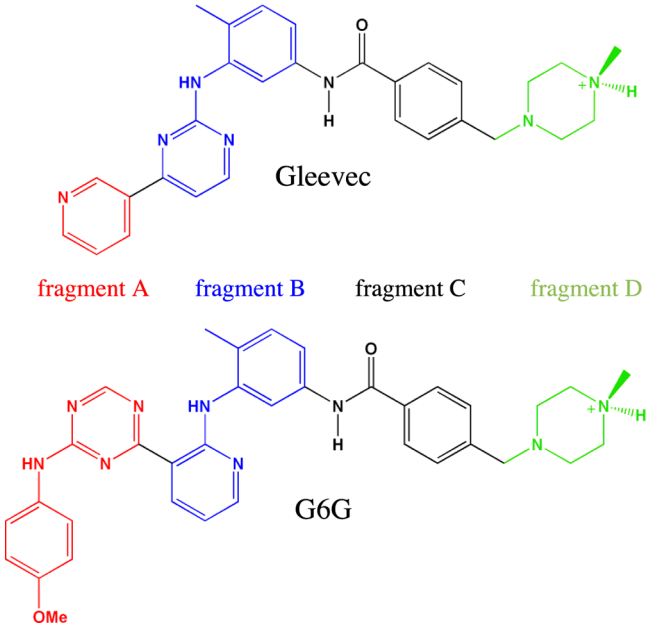


Figure 1. Two-dimensional structures of Gleevec (top: fragment A in red, pyridine group; fragment B in blue, 2-phenylaminopyrimidine group; fragment C in black, benzamide group; and fragment D in green, *N*-methylpiperazine group) and G6G (bottom: fragment A in red, 4-methoxyaniline-*N*-(1,3,5)-triazine group; fragment B in blue, 2-phenylaminopyridine group; fragment C in black, benzamide group; and fragment D in green, *N*-methylpiperazine group).

high inhibitory activity in Abl (IC_{50} = 2.7 nM) and c-Src (IC_{50} = 2.8 nM).[7] According to these results, both c-Src and Abl can access the DFG-out conformation, suggesting that conformational selection is perhaps not the predominant mechanism underlying binding specificity, as previously proposed.[7] However, it is difficult to explain, on the basis of a simple structural argument, why Gleevec exhibits substantial specificity for Abl over c-Src but one of its close chemical analogues, G6G, is an equally potent inhibitor of both kinases. A better understanding of the physical origin of the observed differential effects with G6G would provide additional insights about kinase inhibitor specificity.

As illustrated in Figure 1, G6G is derived from Gleevec by keeping the central scaffold while substituting a 4-methoxyaniline-*N*-(1,3,5)-triazine moiety for the pyridine group (fragment A) and a 2-phenylaminopyridine group for the 2-phenylaminopyrimidine group (fragment B). Crystallography has shown that Gleevec and G6G adopt similar binding modes in the binding pocket of c-Src (Figure 2). The benzamide and the *N*-methylpiperazine groups (fragments C and D, respectively) are common to both Gleevec and G6G (Figure 1); thus, the interaction of this part of the compounds with c-Src should be essentially equivalent. The NH group of fragment

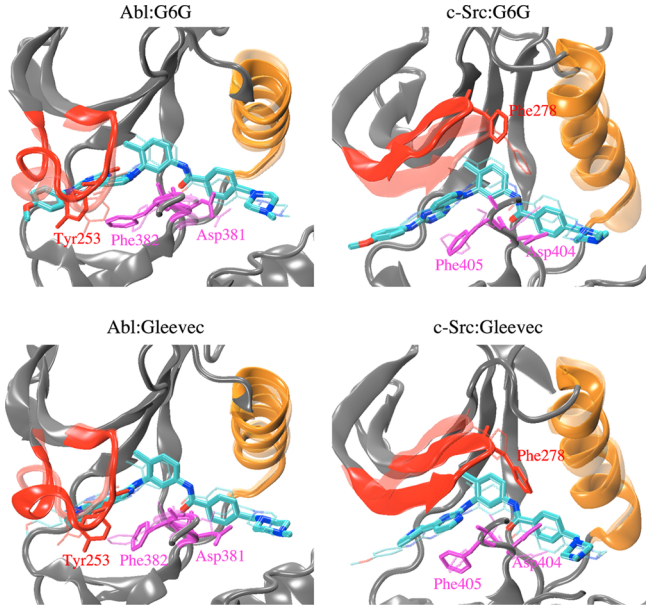


Figure 2. Superimposing the conformations of Abl:G6G, Abl:Gleevec, c-Src:G6G, and c-Src:Gleevec complexes. The P-loop, αC helix, and DFG motif are in red, orange, and magenta, respectively. G6G and Gleevec are represented by sticks, where carbon atoms are in cyan, nitrogen atoms are in blue, and oxygen atoms are in red.

B in Gleevec forms a hydrogen bond with the hydroxyl side chain of Thr338 (called the gatekeeper), but this interaction is broken between G6G and c-Src due to the rotation of the pyridine group (fragment B). This rotation results in a different orientation of the hydrophobic face of the 4-methoxyaniline-*N*-(1,3,5)-triazine group (fragment A) relative to the corresponding pyridine group (fragment A) in Gleevec, which could be key to the binding selectivity of the compounds to c-Src.

Gleevec binds to Abl and c-Src kinases using nearly identical binding site residues in similar conformations according to available crystal structures.[6,10] Interestingly, the phosphate-binding loop (P-loop) in these complexes displays distinct conformations (Figure 3). The 10-residue P-loop (residues 248−257), often referred to as the glycine-rich loop, is a part of

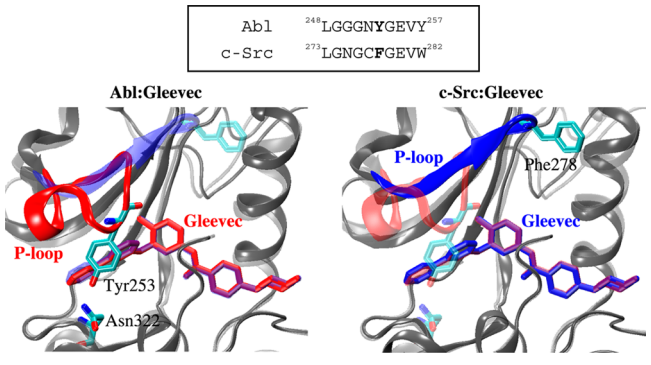


Figure 3. P-loop of Abl and c-Src kinases. Top: amino acid sequence of the P-loop region for Abl (residues 248−257) and c-Src (residues 273−282). Bottom: superimposing the crystal structures of Gleevec in Abl and c-Src complexes. The amino acid backbone of each protein structure is represented in gray. The P-loops of Abl and c-Src are shown in red and blue, respectively, in which the key residue (Tyr253 in Abl and Phe278 in c-Src) is shown by thick sticks. Gleevec bound in Abl and c-Src is illustrated with red and blue sticks, respectively.

dx.doi.org/10.1021/ja504146x | J. Am. Chem. Soc. 2014, 136, 14753−14762

the N-lobe that provides hydrophobic interactions with the adenine ring of ATP.[15] The sequence conservation in the P-loop region, displayed in Figure 3, shows six conserved residues in this region of Abl and c-Src kinases. Two conserved hydrophobic residues, L3 (Leu248 in Abl and Leu273 in c-Src) and V5 (Val256 in Abl and Val281 in c-Src), are almost always in contact with Gleevec in Abl and c-Src (any atom in L3 and V5 is within 5 Å of Gleevec). The P-loop of the Abl:Gleevec complex structure reveals an unusual folded and kinked conformation that allows the binding site of Abl to adopt a tunnel-like shape. The phenol side chain of Tyr253 further stabilizes the P-loop by forming a hydrogen bond with Asn322 in the C-lobe to provide a significant hydrophobic enclosure to the bound ligand. In the c-Src:Gleevec complex, however, the P-loop adopts an extended conformation, pointing outward from the binding pocket. Consequently, structural differences between the P-loop conformations of the tyrosine kinase domains of Abl and c-Src point to possible explanations for the observed binding specificity.[16]

Computational studies have provided a wealth of molecular detail on the energetic determinants of the binding affinity and specificity of Gleevec.[9,12,14] Alchemical free energy perturbation molecular dynamics (FEP/MD) simulations and the molecular mechanics Poisson−Boltzmann with surface area (MM/PBSA)[17,18] approximation have shown that the binding free energies for Gleevec to these DFG-out inactive kinases are similar, supporting the view of conformational selection mechanism.[14] All-atom MD simulations with enhanced sampling techniques have revealed the different stability of the DFG-out conformation in Abl and c-Src kinases, emphasizing that conformational selection mechanism contributes to the binding selectivity of Gleevec in the kinases.[12] A more recent study which investigated both the conformational transition and ligand−protein interactions led to the conclusion that the binding selectivity is controlled by both conformational changes and binding affinity mechanisms.[9] While such simulations rely on an atomistic force field that is approximate, they make it possible to undertake a fine thermodynamic dissection of the binding process.

Despite previous efforts, several critical issues about the molecular determinants underlying the binding selectivity of Gleevec and G6G to an inactive DFG-out conformation of Abl and c-Src kinase domains remain unresolved. The study presented herein aims to identify the physical/chemical determinants governing the differing selectivity of the chemical derivatives, Gleevec and G6G, binding to the homologous kinases, Abl and c-Src. The importance of structural variations in the ligands as well as small differences in amino acid identity in the kinases on selective kinase inhibition will also be examined. To address those issues, a rigorous step-by-step approach of calculating absolute binding free energy was employed to study the binding of Gleevec and G6G to Abl and c-Src kinase domains, in which solute and solvent atoms are treated explicitly with atomic force fields. Calculations considering different P-loop structures in Abl kinase were also carried out to address how different conformations of the P-loop affect G6G binding. Lastly, the sensitivity of computed binding free energy to potential energy function parameters was ascertained by comparing results obtained from different force fields. The study demonstrates that detailed analyses can provide valuable information to explain the observed target preference. Information derived from this study will provide useful principles to guide lead optimization studies aimed at increasing potency and selectivity of drugs.

MATERIALS AND METHODS

Atomic Model and Simulation Details. All simulations employed the CHARMM22 all-atom force field[19] with the CMAP correction for proteins.[20,21] The TIP3P[22] water model was used. The force field parameters for Gleevec and G6G to represent the potential functions of the ligands were obtained using the GAAMP[23] method (see Supporting Information (SI)). The initial structures of the Abl:Gleevec, c-Src:Gleevec, and c-Src:G6G complexes were taken from PDB entries 1IEP,[10] 2OIQ,[6] and 3G6G,[7] respectively. Because there is no crystal structure of Abl in complex with G6G, the system must be modeled by similarity to experimentally known structures. To model the kinase domain, we used the X-ray structure of Abl in complex with Gleevec (1IEP). To model the pose of the ligand in Abl, we used the G6G coordinates from the c-Src:G6G crystal structure (3G6G) after superposing the two kinase domains. In the model of the Abl:G6G complex, the P-loop is in a kinked conformation. Histidine residues in the ligand-bound complex systems were treated as neutral by protonated at $N^{\delta 1}$ or $N^{\epsilon 2}$ according to their local environment. The protonation state of the ionizable amino acid residues was assigned corresponding to the ionization states at pH 7. The *N*-methyl-piperazinyl nitrogen of the ligands, of which the pK_a = 7.7 in water,[24] was treated as protonated to make hydrogen bonds with the backbone carbonyl oxygen of amino acid residues in the binding pockets.[24,25] All crystal water molecules were retained, and hydrogen atoms were added to the initial protein−ligand complex using the HBUILD[26] module of CHARMM.[27] For each simulation, the initial structure of the ligand-bound complex was immersed in a truncated octahedral water cell of edge length 80 Å. Water molecules within 2.6 Å of any non-hydrogen atom of the protein or ligand were deleted. Sodium or chloride ions were added to reach the physiological salt concentration of 0.15 M and overall electric neutrality for each protein−ligand system. To remove any steric hindrance the resulting system was minimized in the presence of strong harmonic constraints (k = 25 kcal/(mol·Å^2)) on all non-hydrogen atoms of the protein−ligand with the steepest descent[28] method for 200 steps, followed by additional 200 steps of minimization using the adopted-basis Newton−Raphson[28] method. The solvated systems of Abl:Gleevec, Abl:G6G, c-Src:Gleevec, and c-Src:G6G complexes consist of 36 753, 36 756, 37 990, and 37 115 atoms, respectively. The solvated system of Abl:G6G with extended P-loop consists of 36 756 atoms.

The fully solvated protein−ligand complex systems were initially equilibrated for 2 ns with harmonic restraints of 25 kcal/(mol·Å^2) applied on the non-hydrogen atoms of the biopolymer complex to ensure that the protein and bound ligand remain near the crystal structure. The constraints were then released and the simulation was continued for 10 ns. The solvation free energy simulations of the ligands in bulk solution included one chloride counterion in a 45-Å cubic water box. The solvated system then was subjected to 600 ps MD simulation with harmonic restraints of 10 kcal/(mol·Å^2) applied to the ligand to keep it near its reference conformation in the bound state of each complex system. All MD simulations of equilibration were performed using the NAMD program.[29] All systems were simulated with periodic boundary conditions (PBC). Initial velocities were assigned according to the Maxwell−Boltzmann distribution at 300 K. The isobaric−isothermal (NPT) ensemble was employed for all MD calculations. The pressure and temperature were kept at 1 atm and 300 K by the Langevin piston method and Langevin dynamics, respectively. Long-range electrostatic interaction was treated by particle-mesh Ewald (PME) algorithm.[30] Short-range nonbonded interactions were truncated at a cutoff distance of 14 Å. A smooth switching function was applied to smoothly reduce the potential to zero at the cutoff distance, starting from 12 Å. The nonbonded interaction list was updated every MD step with a 16-Å cutoff. Covalent bonds involving a hydrogen atom (including the TIP3P water) were constrained to their equilibrium distances using the SHAKE[31] algorithm, and a 2 fs time step was used in all calculations.

dx.doi.org/10.1021/ja504146x | J. Am. Chem. Soc. 2014, 136, 14753−14762

Absolute Binding Free Energy Calculations. Alchemical FEP/MD simulations with restraining potentials based on the double-decoupling method[32] (DDM) provide a rigorous step-by-step reversible work staging procedure to compute the absolute binding free energy of a ligand to a protein.[33−36] Reasonable success has been demonstrated with previous application to T4 lysozyme,[34] FKBP12,[36] bacterial ribosome,[36−38] and several tyrosine kinases.[9,39] The step-by-step staging procedure naturally yields a separation of the total free energy for transferring a ligand from bulk solution to receptor into three contributions,

$$\Delta G_b^\circ = \Delta\Delta G_{int} + \Delta\Delta G_{conf} + \Delta\Delta G_{t+r} \quad (1)$$

where $\Delta\Delta G_{int}$ is the free energy difference for the noncovalent association of the ligand and the protein and dissociation of the ligand from bulk solvent, $\Delta\Delta G_{conf}$ is the conformational free energy cost required to transfer the ligand with its bound-state conformation in bulk solution to the same conformation in the binding site, and $\Delta\Delta G_{t+r}$ corresponds to the work while introducing the translational (t) and rotational (r) restraints on the ligand in bulk solution as required for binding and then removing the restraints from the ligand in the binding site.

The nonbonded interaction energy, $\Delta\Delta G_{int}$, is further decomposed through a step-by-step staging procedure using three thermodynamic coupling parameters, λ_{rep}, λ_{dis}, and λ_{elec}, into the repulsive (rep), dispersive (dis), and electrostatic (elec) components,[33] whereby the ligand is sequentially decoupled from bulk solution and then rematerialized in the binding pocket:

$$\Delta\Delta G_{int} = \Delta\Delta G_{rep} + \Delta\Delta G_{dis} + \Delta\Delta G_{elec} \quad (2)$$

The repulsive and dispersive components are obtained by means of the separation scheme for the Lennard-Jones 6-12 pair potential introduced in the Weeks−Chandler−Andersen[40] (WCA) theory. At this stage, restraining potentials are applied to control the translational and rotational movements as well as the conformation of the ligand while it is gradually decoupled from the bulk solution and then reinserted into the binding pocket of the receptor (site).

The free energy contribution for restricting the translational and rotational motions of the ligand is given by

$$\Delta\Delta G_{t+r} = [-k_B T \ln(F_t C^\circ) - \Delta G_t^{site}] + [-k_B T \ln(F_r) - \Delta G_r^{site}] \quad (3)$$

where F_t and F_r are the translational and rotational factors corresponding to simple numerical integrals over restraining quadratic potentials used to define the position and the orientation of the bound ligand,[34,36] and C° is the standard concentration to cancel the unit in F_t (1661 Å^3). Six internal coordinates of the receptor−ligand complex were randomly selected to describe the positions and orientations of the ligand in free and bound state. At this stage, a conformational restraint was applied on the ligand to keep it near its bound-state conformation.

The free energy contribution for losing the conformational degrees of freedom (DOF) of the ligand upon binding is given by

$$\Delta\Delta G_{conf} = \Delta G_{conf}^{bulk} - \Delta G_{conf}^{site} \quad (4)$$

where ΔG_{conf}^{bulk} and $-\Delta G_{conf}^{site}$ describe the addition of a biasing potential on the ligand to restrain its conformation near its bound-state conformation in bulk solution (bulk) and then release of the restraint in the binding site (site), respectively.

Analysis of the various free energy components extracted from the step-by-step reversible work described above provides insight into the microscopic forces driving ligand binding, with the caveat that the individual terms represent conditional free energy contributions associated with the set of intermediates states used to carry out the full process and that only the total absolute binding free energy is ultimately invariant and independent of the intermediate steps. It is worth noting that simulations of the system with the WCA-repulsive ligand (λ_{rep} = 1, λ_{dis} = 0, λ_{elec} = 0) typically provides an accurate representation of the mean protein and solvent structure around the ligand, such that the average van der Waals dispersive interaction, $\langle U_{dis}\rangle_{(\lambda_{rep}=1,\lambda_{dis},\lambda_{elec})}$, is not very sensitive to the value of λ_{dis} and λ_{elec}. For example, FEP/MD test simulations with Gleevec show that switching λ_{dis} after λ_{elec} (rather than before) merely changes the value of the dispersive component of the binding free energy $\Delta\Delta G_{dis}$ by about 5% (e.g., from −36 to −34 kcal/mol, see the theory section and the associated table in the SI). Such predominance of the repulsive part of the potential to yield the mean solvation structure is already embedded in the WCA[40] theory. This analysis implies that $\Delta\Delta G_{dis}$ is well approximated by the difference in van der Waals dispersive interactions energy of the ligand when it is in the binding site or in the bulk solvent. This observation is important to support the decomposition of $\Delta\Delta G_{int}$ as a sum of residue-based average van der Waals interactions used herein.

All free energies were calculated using the PERT module of the program CHARMM version c36a6.[27] To enhance conformational sampling and accelerate the convergence of the calculated free energies, the FEP/MD methodology was extended by a global replica-exchange molecular dynamics (REMD) scheme with respect to the thermodynamic coupling parameters "λ" implemented in the dual parallelization scheme of REPDSTR[41−44] module in CHARMM.[41,42] We refer to this method as the FEP/λ-REMD approach. Umbrella sampling (US)/REMD simulations were used to compute the potentials of mean force as a function of the root-mean-square deviation (RMSD-PMFs) of the ligand relative to its bound conformation in bulk solution as well as in the binding site to obtain ΔG_{conf}^{bulk} and $-\Delta G_{conf}^{site}$ in eq 4, respectively.[45] A series of 30 independent FEP/λ-REMD and US/REMD simulation runs (cycles) in bulk solution and in the binding pockets of receptors were performed consecutively, starting from the last configuration saved in the previous run for all corresponding windows (replicas), using an alternative random seed number. Exchange of replicas was attempted every 100 steps all the simulations. The new velocity-verlet (VV2) integration scheme was used in all REMD simulations. The nonbonded interaction list was updated at every integration step using a cutoff of 16 Å. The van der Waals energies were switched to zero between 12 and 14 Å with a shift function. For all simulation windows, each trajectory of the simulations started from various initial velocities and were collected every 1000 simulation steps. The reference values of the six randomly selected internal coordinates of each protein−ligand complex for the distance, angle, and dihedral restraints were calculated from the average of last 1 ns equilibration trajectory, as reported in SI Table S1. The force constants for all distance, angle, and dihedral restraints are 10 kcal/(mol·Å^2) and 200 kcal/(mol·rad^2), respectively. The reference structure of the bound-state ligand used for the RMSD restraint was calculated from the average conformation of last 1 ns equilibration MD trajectory of the ligand in the protein complex. The force constant for all RMSD restraints in simulations is 10 kcal/(mol·Å^2).

For the alchemical FEP/λ-REMD simulations, the thermodynamic coupling parameters, λ_{rep}, λ_{dis}, and λ_{elec} were used to control the coupling of the ligand to its environments with respect to the nonbonded interaction shown in eq 2. Nonlinear scaling was applied to describe the repulsion as a soft-core potential and linear scaling was used to compute the dispersive and electrostatic energies. The values for λ_{rep} ranged from 0 to 1 (0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1), and the values for λ_{dis} and λ_{elec} ranged from 0 to 1 in increments of 0.1, respectively. Thus, a total of 40 replicas were used to compute the nonbonded interaction energies for binding. The configuration of the last λ-staging replica of the repulsion term is allowed to exchange with that of the first λ-staging replica of the dispersion term, so as the last λ-staging replica of the dispersion term and the first λ-staging window of electrostatics term. Calculations using the coupling parameters described above were performed for both the binding site and bulk phase simulations. This ensured that the internal energy contributions in the gas phase were eliminated from the free energy both in the binding site and bulk solution. For the translational and rotational FEP/λ-REMD simulations, 15 replicas were used, each of which introduced a thermodynamic coupling parameter, λ_{t+r} (λ_{t+r} = 0, 0.0025, 0.005, 0.0075, 0.01, 0.015, 0.024375, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6,

0.8, and 1) to control the strength of restraining potential. The simulation length of each replica window in the alchemical FEP/λ-REMD calculations and in the orientational FEP/λ-REMD calculations consists of 220 and 100 ps for each run of sampling, per λ, respectively. The US/REMD simulations in bulk solution as well as in the binding site individually consist of 21 replicas, centered on RMSD offsets increasing from 0.0 to 5.0 Å in steps of 0.25 Å. Each replica window in the US/REMD simulations comprises 200 ps for data collection.

All data points were collected and employed in the weighted histogram analysis method (WHAM)[46,47] to obtain the free energies of repulsive, dispersive, electrostatic, translational, and rotational contributions as well as to compute the PMFs as a function of RMSDs for the unbiased systems upon ligand binding.

RESULTS AND DISCUSSION

Convergence and Validation with Experiment. The absolute binding free energies for the ligands to the kinases were computed for 30 consecutive FEP/λ-REMD cycles. This represents a total of 6.6, 3.0, and 6.0 ns sampling per replica in the three phases of FEP calculations, respectively, for a total aggregated simulation time of 435 ns for each system. Convergence of the calculations may be ascertained by monitoring the evolution of the results over several successive production cycles (SI Figure S1). The results show that the convergence is reached within ~10 cycles, after which the calculated free energy of binding begins to fluctuate around a mean value. The data points of the last 10 cycles were collected to compute the block average and standard error of the free energies of binding for Gleevec and G6G, yielding −12.7 ± 1.2 and −13.1 ± 1.4 kcal/mol in Abl, and −7.5 ± 1.3 and −15.9 ± 1.5 kcal/mol in c-Src, respectively (Table 1).

In a prior computational study, we showed that the free energy cost for the DFG flip is 1.4 and 5.4 kcal/mol in Abl and c-Src, respectively.[9] Thus, the total standard binding free energies of Gleevec and G6G, ΔG°, including the thermodynamic contributions associated with both the conformational change of the DFG-flip in the apo kinase and the free energy of binding for ligand to the DFG-out conformation, is estimated to be −11.3 and −11.7 kcal/mol in Abl kinase, and −2.1 and −10.5 kcal/mol in c-Src, respectively (Table 1). Notably, the present calculated standard binding free energies of the ligands to the kinases are in quantitative agreement with experimental affinities obtained by measuring the inhibitory potency of the ligands to unphosphorylated form of the proteins.[7] Thus, the present computational results are sufficiently accurate to support the conclusions about binding selectivity of the ligands to the kinases. The step-by-step FEP strategy naturally leads to a dissection of the absolute binding free energy of ligand into five components following eqs 1 and 2. Consequently, the detailed dissection of the absolute binding free energy summarized in Table 1 provides meaningful information to understand the molecular determinants responsible for the differential binding affinities among scaffold-similar ligands to the homologous kinase binding sites. In the following, we analyze each of these contributions to provide a physical basis for Gleevec and G6G binding to Abl and c-Src kinases.

Structure−activity relationships are often discussed and explained on the basis of fragment- or residue-based decompositions. Fragment decomposition of the nonbonded interaction energies between the bound ligand and binding site residues of a receptor helps to identify the contribution of each individual fragment of the ligand responsible for its binding affinity to potential receptors. In the case of G6G and Gleevec,

Table 1. Standard Binding Free Energies[a]

(A) Gleevec and G6G in Abl Kinase		
bulk → Abl	Gleevec	G6G[e]
$\Delta\Delta G_{rep}$	6.3	10.4
$\Delta\Delta G_{dis}$	−28.6	−35.9
$\Delta\Delta G_{elec}$	−3.4	−1.5
$\Delta\Delta G_{int}$	−25.7	−27.0
$\Delta\Delta G_{conf}$	7.7	8.6
$\Delta\Delta G_{t+r}$	5.4	5.3
total ligand binding	−12.7	−13.1
DFG flipping[b]	1.4	1.4
ΔG_b°	−11.3	−11.7
expt[c]	−10.9	−11.8
(B) Gleevec and G6G in c-Src Kinase		
bulk → c-Src	Gleevec	G6G
$\Delta\Delta G_{rep}$	10.6	7.5
$\Delta\Delta G_{dis}$	−26.7	−32.3
$\Delta\Delta G_{elec}$	−0.1	−2.2
$\Delta\Delta G_{int}$	−16.2	−27.0
$\Delta\Delta G_{conf}$	5.8	7.1
$\Delta\Delta G_{t+r}$	2.9	4.0
total ligand binding	−7.5	−15.9
DFG flipping[b]	5.4	5.4
ΔG_b°	−2.1	−10.1
expt[d]	>−6.9	−11.7

[a]Units: kcal/mol. [b]Data were taken from the previous study.[9] [c]The inhibitory potency of Gleevec and G6G for unphosphorylated Abl is 11 ± 3 and 2.7 ± 0.3 nM, respectively, when measured at pH 8.0.[7] The standard deviation of the ΔG_b° in the final 10 cycles is 1.2 and 1.4 kcal/mol for Abl:G6G, and Abl:Gleevec, respectively, and is used as the error estimation here. [d]The inhibitory potency of Gleevec and G6G for unphosphorylated c-Src is >10 000 and 2.8 ± 0.2 nM, respectively, when measured at pH 8.0.[7] The standard deviation of the ΔG_b° is 1.3 and 1.5 kcal/mol for c-Src:G6G and c-Src:Gleevec, respectively, and is used as the error estimation here. [e]Calculations on the Abl:G6G complex based a model with a kinked P-loop conformation. Each $\Delta\Delta G$ term indicates that it is a free energy difference of ligand association with the protein and dissociation from bulk solvent. The subscript of each $\Delta\Delta G$ term suggests the contribution of thermodynamic coupling. $\Delta\Delta G_{int}$ is the total noncovalent contribution which is the sum of $\Delta\Delta G_{rep}$, $\Delta\Delta G_{disp}$, and $\Delta\Delta G_{elec}$. A more detailed description of each free energy component can be found in eqs 1−4 in the Materials and Methods section.

the intermolecular interaction energies were decomposed into the contributions from the four fragments (A−D) shown in Figure 1. Residue-based decomposition of the nonbonded interaction energy in the complex systems helps to determine the key molecular determinants responsible for the binding selectivity of the kinase to the ligands (SI Figures S8 and S9). These decomposition analyses were performed to provide more insight into how receptor sequence and chemical scaffold of ligand lead to variation in computed ligand binding affinity.

G6G and Gleevec Are Equally Potent Inhibitors of Abl Kinase. The FEP calculations reported in Table 1A indicate that the binding affinities for G6G and Gleevec for Abl are quantitatively similar. Consistent with previous results,[9] the absolute binding affinities for both ligands in Abl kinase is predominantly driven by intermolecular dispersive forces (see the progression of the dispersive free energies of the ligands as a function of the coupling parameter λ_{dis} in SI Figure S2B). The slope of the linear progression of the dispersive interaction in the binding site is greater than that in bulk solution because the

dx.doi.org/10.1021/ja504146x | J. Am. Chem. Soc. 2014, 136, 14753−14762

protein, relative to bulk solvent, provides an environment with a higher density of van der Waals centers to stabilize the ligand in the binding pocket. This contribution is more favorable by −7.3 kcal/mol for G6G binding to Abl compared with Gleevec. This is consistent with the solvent-accessible surface area of G6G (1076 Å^2), which is larger than that of Gleevec (892 Å^2). Two opposing factors, however, approximately couterbalance the difference in the dispersive contribution. One is the unfavorable repulsion contribution due to the insertion of the large G6G molecule in the binding pocket; the latter is more costly by 4.1 kcal/mol for G6G than for Gleevec. Futhermore, the electrostatic contribution for G6G is less favorable by 1.9 kcal/mol than for Gleevec. Consequently, the free energy decomposition reveals that an internal cancellation of the nonbonded interactions affects ligand binding to Abl kinase, resulting in nearly equivalent nonbonded contributions to the molecule association. We return to the repulsive contributions below.

The binding specificity of the ligands to Abl kinase in terms of $\Delta\Delta G_{dis}$ is mainly governed by the fragment A; $\Delta\Delta G_{dis}$ = −23.8 kcal/mol for G6G relative to −11.8 kcal/mol for Gleevec (Table 2A). The reason is that the fragment A of G6G has a

Table 2. (A) Average van der Waals and (B) Electrostatic Interaction Energies[a] of the Binding Site Residues in Abl or c-Src Kinase with the Fragments of the Bound Gleevec or G6G

(A) Average van der Waals Interaction Energies, E_{vdW}				
kinase:ligand	fragment A	fragment B	fragment C	fragment D
Abl:Gleevec	−11.8	−26.0	−14.9	−11.8
Abl:G6G	−23.8	−27.0	−15.0	−11.7
c-Src:Gleevec	−10.6	−23.6	−14.4	−12.6
c-Src:G6G	−20.0	−22.4	−14.0	−12.5
(B) Electrostatic Interaction Energies, E_{elec}				
kinase:ligand	fragment A	fragment B	fragment C	fragment D
Abl:Gleevec	−5.4	−3.9	−3.3	−14.8
Abl:G6G	−8.8	−1.9	−9.1	−13.1
c-Src:Gleevec	−5.2	−3.4	−6.7	−18.3
c-Src:G6G	−9.6	−1.5	−8.9	−13.5

[a]Unit: kcal/mol.

better surface complementary to the hydrophobic cleft lined by Leu248, Tyr253, Phe317, and Gly321 than the corresponding group in Gleevec (SI Figure S3A). The results of residue-based decomposition between each binding site residue of Abl and the bound ligand highlight the contributions of Phe382 and Lys271 from the P-loop (Figure 3). In Abl kinase, the P-loop displays a folded and kinked conformation toward the binding site[10] (Figure 2, left panel) that is unusual compared to other kinases. In the ligand-bound Abl complexes, the P-loop is folded over fragments A and B of the ligands, enabling the kinase to form favorable van der Waals contacts with the ligands in the binding pocket.[10,39,48,49] The residue-based decomposition calculations show that the van der Waals interaction energies of the P-loop residues with the fragment A of G6G are more favorable than with the corresponding moiety in Gleevec (by summing up E_{vdW} of residues 248−257 in ligand-bound Abl in SI Table S2). This suggests that the P-loop plays a role in strengthening the interactions of Abl kinase with G6G relative to Gleevec. The conserved Phe382 in the DFG motif, which is involved in anchoring the fragment B of the ligand at the binding pocket of Abl, makes more favorable interactions with G6G than with Gleevec. The interaction of Lys271 with G6G is more favorable by 2.5 kcal/mol than with Gleevec. All the observations demonstrate that G6G has more favorable interactions with Abl relative to Gleevec.

As noted above, G6G experiences stronger repulsive interactions than Gleevec in the Abl binding cavity. The free energy to insert a purely repulsive molecular entity with no electrostatic and van der Waals dispersion in the Abl binding cavity is more costly for G6G than for Gleevec (SI Figure S2A). This contribution ΔG_{rep} is essentially the reversible work required for pushing protein or solvent atoms out of the way while the WCA-repulsive ligand is materialized in the system. The repulsive free energy associated with pushing water molecules is correlated with the average number of water molecules within the binding pockets of the tyrosine kinases as a function of λ_{rep} during the FEP/MD calculations; SI Figure S4 shows that the insertion of G6G expels two more water molecules from the binding pockets than the insertion of Gleevec. Alternatively, the repulsive free energy for pushing protein atoms is associated with the free energy cost for structurally rearranging the Abl binding pocket itself. The bulky fragment A (triazine moiety) of G6G is more likely to cause steric clashes with the kinked P-loop of Abl than the corresponding pyridinyl moiety in Gleevec (SI Figure S5). To support this presumption, one can compare the protein conformation for the unbound (apo) and bound (holo) forms, and monitor the displacement of P-loop residues as a function of the thermodynamic coupling parameter λ_{rep} (SI Figure S6). Analysis of the RMSD deviations of the P-loop as a function of the thermodynamic coupling parameter λ_{rep} shows that it deviates more from the apo-conformation to avoid undesired steric clash in the case of G6G than in the case of Gleevec (SI Figure S6).

The electrostatic free energy contribution of G6G is not significantly different from that of Gleevec (Table 1A and SI Figure S2C). A simple visual comparison of the Abl:G6G and Abl:Gleevec complexes shows that the pattern of their hydrogen-bonding network is nearly identical except for two regions (SI Figure S7). One is that Abl makes a hydrogen-bonding interaction directly with the NH group of the fragment B of Gleevec using the hydroxyl side chain of the gatekeeper residue Thr315. In contrast, the pyridinyl group in fragment B of G6G adopts a different orientation and does not form a hydrogen bond with Thr315. It has been suggested that loss of the critical hydrogen-bonding interaction with the gatekeeper residue is the reason for the observed loss of Gleevec affinity for the Abl T315I mutant.[50] The lack of Thr315 hydrogen-bonding interaction with G6G appears to confer some advantages to G6G because it can be simultaneously a potent Abl inhibitor of both wild-type and the T315I mutant, as evidenced by experiments of activity.[7] One additional difference in hydrogen-bonding patterns between the two ligands involves Met318. Met318 forms one hydrogen-bonding interaction with the pyridinyl group (fragment A) of Gleevec via its backbone NH group, whereas, it forms two hydrogen bonds with the triazine group of G6G through the backbone groups. The gain of the second hydrogen bond with Met318 is apparently sufficient to compensate for the loss of the one with Thr315 (SI Figure S8), which may explain the insignificant change in electrostatic binding affinity between the two inhibitors against wild-type Abl kinase.

Differential Sensitivity of c-Src to G6G and Gleevec. The absolute binding affinities reported in Table 1B of G6G and Gleevec for c-Src are −15.9 and −7.5 kcal/mol, respectively, closely tracking the experimental trend. The increased potency of G6G for c-Src can be explained mainly by the character of intermolecular dispersive van der Waals interactions. By summing up the dispersive interaction energy of each binding site residue of the kinases (SI Table S2), the total protein−ligand dispersive interaction energy is −74.8 and −61.3 kcal/mol for the c-Src:G6G and c-Src:Gleevec complex, respectively. c-Src kinase buries the triazine group of G6G into a hydrophobic pocket composed of the side chains of Leu273, Tyr340, and Gly344 (SI Figure S3B). In line with the fragment-based (Table 2B) and residue-based (SI Figure S9) decomposition analyses, these hydrophobic residues have stronger van der Waals interactions with the triazine group of G6G than with the corresponding pyridinyl moiety of Gleevec. Furthermore, in contrast to Abl, the WCA-repulsive component for c-Src binding to G6G is slightly less costly by ∼3 kcal/mol than that of Gleevec. This could be attributed to the different conformations of the P-loop in the Abl and c-Src complexes. The main structural differences between the c-Src and Abl complexes are due to the region of the P-loop, which is in an extended conformation in c-Src but is in a kinked conformation in Abl (Figure 2 and SI Figure S5). The RMSD deviations show that G6G displaces the P-loop of c-Src in an analogous manner as Gleevec does (SI Figure S6). This implies that the unfavorable repulsive interaction that is caused by the steric clash between the kinked P-loop and the fragment A of G6G is not observed in the c-Src:G6G complex due to different P-loop conformations. Therefore, substituting the pyridinyl group with a triazine moiety increases the intermolecular dispersive interactions with c-Src but does not introduce additional steric clashes, resulting in an increase in binding affinity.

Loss of Conformational, Translational, and Rotational Degrees of Freedom. It has been shown previously that the relative free energies, calculated from switching-on the various restraining potentials when the ligand is in bulk solution and switching-off the latter when the ligand is in the binding pocket, can be interpreted in terms of a loss of conformational, translational, and orientational DOF of the ligand.[35,45,51] For instance, it is expected that a ligand can access a wider variety of stable conformations in bulk solution than in the binding pocket, and that a considerable free energy cost must be associated with the loss of conformational freedom ($\Delta\Delta G_{conf}$). This free energy cost can be evaluated by computing the PMF (in bulk solution as well as in the binding sites) as a function of the RMSD of the ligand relative to the bound conformation (SI Figure S10).[45] The present strategy also accounts for the free energy associated with the loss of translational and rotational entropy of ligand accompanying binding ($\Delta\Delta G_{t+r}$). It is achieved by applying strong harmonic restraints to the translational and rotational DOF of the ligand in bulk solution and then releasing these restraints in the binding site. Both $\Delta\Delta G_{conf}$ and $\Delta\Delta G_{t+r}$ are reported in Table 1. One should note that these free energy penalties are commonly neglected or implicitly assumed to be constant in most end-point docking/scoring schemes such as MM/PBSA.

The restriction on the conformational, translational, and rotational motions of G6G and Gleevec as required for binding to Abl and c-Src are considerable, leading to a free energy cost on the order of ∼8−14 kcal/mol, strongly disfavoring ligand binding. In particular, the magnitude of these free energy penalties offsets about one-third of the gains from the favorable dispersive interactions. Additionally, the entropic penalty for the loss of conformational, translational, and rotational motions upon binding is larger for G6G than for Gleevec by ∼1−2 kcal/mol. The difference can be attributed to the size of the ligands and the structural complementarity to the binding pockets. G6G, having one additional soft dihedral angle compared to Gleevec, can explore more stable conformations in bulk solution than Gleevec. However, the restricted space in the binding pockets forces G6G to adopt a binding pose similar to that of Gleevec, resulting in a larger loss of conformational freedom upon binding. In general, the translational and rotation motions of a bound ligand depend on how tightly it is held in the binding pocket, which can vary from one molecule to another. Due to the bulky size of G6G compared to Gleevec, one can imagine that the binding cavity is more restricted upon binding G6G. These observations point to the importance of considering the restrictions of the translational, orientational and conformational motions of ligand while computing binding affinity.

Impact of the P-loop Conformation on Ligand Specificity. The P-loop adopts a different conformation in the crystal structure of c-Src[6] and Abl[10] in complex with Gleevec (Figure 3); it is extended in the case of c-Src, and it is kinked in the case of Abl. However, there is no experimental structure for the Abl:G6G complex. The Abl(*kink*) model used in the present computations is based on the crystal structure of the Abl:Gleevec complex and the P-loop is in a kinked conformation. The possible role of the P-loop conformation in the Abl:G6G complex has been noted previously,[7] although its impact on the binding affinity remains unknown. To address this issue, a model of the Abl:G6G complex was built based on the crystal structure of Abl in complex with the kinase inhibitor NVP-AEG082 (PDB entry 2HZ0[52]). According to FEP/MD calculations carried on the basis of this hypothetical Abl(*ext*) model (SI Table S4), the binding of G6G to Abl(*ext*) is more favorable than to Abl(*kink*), which is consistent with the absence of steric clash with the terminal triazine fragment discussed above. However, this analysis neglects the reversible work needed to transform the conformation of the P-loop from kinked to extended. The alchemical FEP/MD simulations are not expected to spontaneously yield a proper thermal averaging over these slowly exchanging P-loop conformations. To ensure a fair comparison of the binding G6G affinity with the two possible P-loop conformations, the free energy difference between Abl(*ext*) and Abl(*kink*) has to be also taken into account. The relative free energy between the kinked and extended P-loop conformations was calculated using umbrella sampling simulations based on the Abl:G6G complex (considering the *holo* rather than the *apo* form simplifies the sampling required for the PMF calculations because the protein fluctuations are smaller in the presence of the bound ligand). The one-dimensional PMF for the P-loop conformation shows that the free energy of the extended conformation is more unfavorable by ∼5 kcal/mol (see SI Figure S11 for the free energy profile). This PMF calculation indicates that the Abl:G6G complex with the P-loop in the extended conformation is less stable than the Abl:G6G complex with the P-loop in the kinked conformation. For this reason, the binding free energy value calculated with the P-loop in the kinked conformation is expected to be more representative of the actual binding free energy.

dx.doi.org/10.1021/ja504146x | J. Am. Chem. Soc. 2014, 136, 14753−14762

Sensitivity of the Calculated Ligand Binding Affinity to Force Field Parameters. An important concern with alchemical FEP/MD simulations in the field of drug discovery is the accuracy of the atomic molecular mechanics force field for the molecules of interest. Compounds such as Gleevec and G6G are not typically in standard biomolecular force fields and an accurate parametrization must be generated according to some objective computational protocol. In a previously study, the force field of Gleevec was developed to be consistent with the CHARMM force field for proteins and nucleic acids.[25] The binding affinity of Gleevec for Abl calculated with the CHARMM-like model was shown previously to be in excellent agreement with the experiment.[9] However, such a CHARMM-like model is not available for G6G. To avoid possible inconsistencies between force field models that could undermine a meaningful comparison of the Gleevec and G6G results, both compounds were parametrized using the general automated atomic model parametrization[23] (GAAMP) Web server. This new model of the Gleevec molecule provides a unique opportunity to examine the sensitivity of calculated binding free energy to force field parameters. Decomposition of the binding affinity in FEP simulation previously demonstrated that the dominant binding free energy contribution arises from the favorable nonbonded intermolecular interactions between the ligand and the receptor, i.e., dispersive van der Waals and electrostatic interactions. Another important factor was the free energy cost due to the loss of conformational DOF of the ligand upon binding. The magnitude of this free energy cost directly depends on the soft DOF of the ligand such as the dihedral torsions. To examine quantitatively the sensitivivity of the FEP results to the underlying force field parameters, two tests of the force field of Gleevec are considered here. The first test is aimed at evaluating the sensitivity of the $\Delta\Delta G_{conf}$ to the torsion potential for the ligand. The second test is aimed at evaluating the sensitivity of the intermolecular protein−ligand interactions to the nonbonded the point-charge and LJ parameters of the ligand. The results of the sensitivity tests are reported in SI Tables S5 and S6.

Because the CHARMM-like and GAAMP force fields share a similar functional form, comparison of the calculated free energy components to the binding using these two force fields is meaningful. Furthermore, the system setups as well as the simulation protocols and simulation parameters were under the same conditions to provide exact comparisons and to avoid any inconsistency resulted from different simulation schemes. According to the FEP calculations, the difference in $\Delta\Delta G_{conf}$ caused by changing from the CHARMM-like torsion parameters to the GAAMP parameters is estimated to be about −1 kcal/mol (SI Table S5). This computational experiment suggests that the effect of the ligand torsion parameters on the conformational free energy component is roughly unchanged. A straightforward test on the sensitivity of the nonbonded free energy component (dispersive van der Waals and electrostatic interactions) to the force field of the ligand is to progressively make small perturbations to the ligand charge and LJ parameters from one force field to the other and to determine the change in the overall binding free energy. The results indicate that the charge models in the two force fields perform similarly in the electrostatic interactions of binding. While the contribution from dispersive interactions remains very large, it appears slightly more sensitive to the LJ parameters of the ligand (SI Table S6). Overall, the change in the binding affinity caused by the perturbation in the nonbonded parameters of the ligand from the CHARMM-like force field to the GAAMP force field is estimated to be 2.8 kcal/mol. One may note, however, that both ligand force fields have the ability to reproduce the experimental binding affinity (see ref 9 and Table 1), due to compensatory cancelation of errors. Taken together, the calculated binding affinity based on GAAMP force field deviates from that using CHARMM-like force field by an average of 1.8 kcal/mol. While these variations with respect to force field parameters are not small, the analysis increases our confidence in the conclusions from our previous study.[9]

■ CONCLUSION

In this study, all-atom molecular dynamics (MD), free energy perturbation (FEP), umbrella sampling (US), and energy decomposition analyses were carried out for two kinase inhibitors, G6G and Gleevec, in complex with Abl and c-Src. Our primary goal was to understand how variations in receptor sequence and structure as well as chemical modifications of the ligand affect binding specificity and drug selectivity. The kinase inhibitor G6G was synthesized in part to see if binding specificity to c-Sr and Abl kinases was indeed controlled through the DFG motif by a conformational selection mechanism.[7] As G6G was shown to respectively bind to c-Src and Abl kinase in the DFG-out conformation with equal potency, this experimental study strongly suggested that conformational selection may not be the predominant mechanism underlying binding specificity. But that remained difficult to definitively conclude from the available information. From this point of view, detailed free energy calculations based on all-atom MD simulations with explicit solvent can add important clarification by revealing hidden thermodynamic factors. The results from the present and prior computational studies' free energy calculations show unambiguously that the binding preference of Gleevec and G6G to the two tyrosine kinases is controlled both by conformational changes and by differences in protein−ligand interactions. There is indeed a larger unfavorable free energy cost to c-Src compared to Abl for adopting the DFG-out conformation, but substituting the small pyridinyl group of Gleevec with a bulkier triazine group in G6G significantly enhances the favorable dispersive van der Waals interactions of the ligand with the protein, thereby yielding a high-affinity inhibitor of c-Src kinase. However, the situation is different when G6G binds to Abl kinase, where the triazine group leads to undesired steric clashes with the kinked-shape P-loop, canceling out the favorable van der Waals interactions. Therefore, substitution of the small pyridinyl group in Gleevec by the triazine group in G6G results in no change in overall binding affinity to Abl. Remarkably, the substitution leads to stronger dispersive interactions between G6G and c-Src while avoiding steric clashes with the extended-shape P-loop, resulting in an increase in binding affinity to c-Src. The good agreement with the experimental data is encouraging, suggesting an energetic explanation for the selectivity of G6G and Gleevec against Abl and c-Src kinases. Comparison of the computational results obtained with different force fields increases our confidence that robust conclusions about the mechanism of binding specificity can be achieved with current methodologies.

The conformational variation of the P-loop in Abl and c-Src clearly encodes important structural features that affect the binding mechanism of inhibitors.[53] The present calculations suggest that there is an energetic compromise required to

accommodate G6G or Gleevec in the binding site of the tyrosine kinase in the context of different P-loop conformations. Differences in flexibility and conformational plasticity of the P-loop could be exploited to design ligand selectivity for particular kinases.[54]

A major concern in developing drugs that target kinases is the insurgence of resistance after prolonged treatment.[26] Designing potent second-generation inhibitors from variations of the original ligand is an approach that is often used to overcome drug resistance. G6G provides one example of an inhibitor that requires less conformational distortion of the receptor to retain binding activity. An important observation from the present FEP study is that shape-adaptation of a compound to the kinase-specific P-loop conformation is a structural constraint that could be exploited in trying to improve specificity. Ultimately, the long-term goals of such computational analysis are to help fine-tune the inhibitory profile of specific compounds by optimizing the interactions of unique residues surrounding the drug-binding site, and to rationally improve the ability of an inhibitor to target a specific kinase by optimizing specific interactions. It is our hope that, in the near future, the physical insight gained from atomistic computations will facilitate the discovery and rational design of small-molecule inhibitors of kinases.

■ ASSOCIATED CONTENT

Supporting Information

More detailed introduction of the step-by-step free energy perturbation framework used here; Figure S1, plot of convergence of the absolute binding affinities of Gleevec and G6G with Abl and c-Src kinases; Figure S2, progression of the free energy components with respect to the coupling parameters of λ_{rep}, λ_{dis}, λ_{elec}, and λ_{t+r} for ligand in the binding sites or in bulk solution; Figure S3, cartoon representation and surface representation of key residues forming additional hydrophobic contacts with the triazine group of G6G in Abl:G6G and c-Src:G6G complexes; Figure S4, progression of the number of water molecules in the binding pockets of Abl and c-Src in response to G6G or Gleevec binding during the FEP calculations; Figure S5, RMSD for the binding site residues of the kinases as a function of the coupling parameter λ_{rep}; Figure S6, superimposed conformations of the kinase in apo state and in holo state with bound G6G or Gleevec in the binding pocket; Figure S7, snapshots of hydrogen-bonding interactions in the binding pockets of Abl and c-Src in complex with Gleevec and G6G; Figures S8 and S9, differential interaction energies between each kinase residue and G6G relative to the interaction energies between the corresponding residue and Gleevec; Figure S10, PMF profiles on the conformational restraints for Gleevec and G6G in bulk solution as well as in the binding pockets of Abl and c-Src kinases; Figure S11, free energy profile of the conformational change in P-loop and the distribution of the time series from umbrella sampling calculations; Table S1, definitions of the anchoring points used in the λ_{t+r}-staged FEP/λ-REMD calculations; Tables S2 and S3, averaged dispersive (E_{dis}) and electrostatic (E_{elec}) interaction energies between protein residues and ligand in the kinase binding pocket; Table S4, standard binding free energy of G6G with Abl Kinase, of which the P-loop is in kinked-shape (*kink*) or extended conformation; Table S5, sensitivity test of calculated free energy associated with conformational change upon Gleevec binding in Abl kinase to torsion parameters; Table S6, sensitivity tests of calculated dispersive and electrostatic free energies upon Gleevec binding in Abl kinase to Lennard-Jones parameters and point-charge models; Table S7, free energy components of solvating G6G in bulk solvent, with respect to the order in which the corresponding interactions are switched on in FEP/MD calculations; and topology and parameter files of G6G and Gleevec. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

roux@uchicago.edu

Author Contributions

[†]Y.-L.L. and Y.M. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Wei Jiang for helpful discussions. This research was supported by the National Cancer Institute of the National Institutes of Health (NIH) through Grant CAO93577. The computations were made possible by the Extreme Science and Engineering Discovery Environment (XSEDE) supported through NSF Grant OCI-1053575, and by additional resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory through NIH Grant S10 RR029030-01.

■ REFERENCES

(1) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. *Science* **2004**, *303*, 1800−1805.

(2) Sciabola, S.; Stanton, R. V.; Wittkopp, S.; Wildman, S.; Moshinsky, D.; Potluri, S.; Xi, H. L. *J. Chem. Inf. Model.* **2008**, *48*, 1851−1867.

(3) Buchdunger, E.; Zimmermann, J.; Mett, H.; Meyer, T.; Muller, M.; Druker, B. J.; Lydon, N. B. *Cancer Res.* **1996**, *56*, 100−104.

(4) Druker, B. J.; Lydon, N. B. *J. Clin. Invest.* **2000**, *105*, 3−7.

(5) Lydon, N. *Nature Med.* **2009**, *15*, 1153−1157.

(6) Seeliger, M. A.; Nagar, B.; Frank, F.; Cao, X.; Henderson, M. N.; Kuriyan, J. *Structure* **2007**, *15*, 299−311.

(7) Seeliger, M. A.; Ranjitkar, P.; Kasap, C.; Shan, Y. B.; Shaw, D. E.; Shah, N. P.; Kuriyan, J.; Maly, D. J. *Cancer Res.* **2009**, *69*, 2384−2392.

(8) Deininger, M.; Buchdunger, E.; Druker, B. J. *Blood* **2005**, *105*, 2640−2653.

(9) Lin, Y. L.; Meng, Y. L.; Jiang, W.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 1664−1669.

(10) Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. *Cancer Res.* **2002**, *62*, 4236−4243.

(11) Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J. *Science* **2000**, *289*, 1938−1942.

(12) Lovera, S.; Sutto, L.; Boubeva, R.; Scapozza, L.; Dolker, N.; Gervasio, F. L. *J. Am. Chem. Soc.* **2012**, *134*, 2496−2499.

(13) Dar, A. C.; Lopez, M. S.; Shokat, K. M. *Chem. Biol.* **2008**, *15*, 1015−1022.

(14) Aleksandrov, A.; Simonson, T. *J. Biol. Chem.* **2010**, *285*, 13807−13815.

(15) Liao, J. J. L. *J. Med. Chem.* **2007**, *50*, 409−424.

(16) Guimaraes, C. R. W.; Rai, B. K.; Munchhof, M. J.; Liu, S. P.; Wang, J.; Bhattacharya, S. K.; Buckbinder, L. *J. Chem. Inf. Model.* **2011**, *51*, 1199−1204.

(17) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tiradorives, J. *J. Chem. Phys.* **1988**, *89*, 3742−3746.

(18) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670−681.

(19) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(20) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(21) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(23) Huang, L.; Roux, B. *J. Chem. Theory Comput.* **2013**, *9*, 3543−3556.
(24) Szakacs, Z.; Beni, S.; Varga, Z.; Orfi, L.; Keri, G.; Noszal, B. *J. Med. Chem.* **2005**, *48*, 249−255.
(25) Aleksandrov, A.; Simonson, T. *J. Comput. Chem.* **2010**, *31*, 1550−1560.
(26) Brunger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148−156.
(27) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(29) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(30) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(31) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(32) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(33) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.
(34) Deng, Y. Q.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(35) Deng, Y. Q.; Roux, B. *J. Chem. Phys.* **2008**, *128*, 115103.
(36) Wang, J. Y.; Deng, Y. Q.; Roux, B. *Biophys. J.* **2006**, *91*, 2798−2814.
(37) Ge, X. X.; Roux, B. *J. Phys. Chem. B* **2010**, *114*, 9525−9539.
(38) Ge, X. X.; Roux, B. *J. Mol. Recognit.* **2010**, *23*, 128−141.
(39) Lin, Y. L.; Roux, B. *J. Am. Chem. Soc.* **2013**, *135*, 14741−14753.
(40) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237.
(41) Jiang, W.; Hodoscek, M.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 2583−2588.
(42) Jiang, W.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 2559−2565.
(43) Woodcock, H. L.; Hodoscek, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F.; Brooks, B. R. *J. Comput. Chem.* **2007**, *28*, 1485−1502.
(44) Woodcock, H. L.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F.; Brooks, B. R. *Theor. Chem. Acc.* **2003**, *109*, 140−148.
(45) Woo, H. J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.
(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(47) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(48) Lee, T. S.; Potts, S. J.; Albitar, M. *Recent Pat. Anti-Cancer* **2009**, *4*, 164−173.
(49) Lee, T. S.; Potts, S. J.; Kantarjian, H.; Cortes, J.; Giles, F.; Albitar, M. *Cancer* **2008**, *112*, 1744−1753.
(50) Milojkovic, D.; Apperley, J. *Clin. Cancer Res.* **2009**, *15*, 7519−7527.
(51) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(52) Cowan-Jacob, S. W.; Fendrich, G.; Floersheimer, A.; Furet, P.; Liebetanz, J.; Rummel, G.; Rheinberger, P.; Centeleghe, M.; Fabbro, D.; Manley, P. W. *Acta Crystallogr. D* **2007**, *63*, 80−93.
(53) Garcia-Bassets, I.; Kwon, Y. S.; Telese, F.; Prefontaine, G. G.; Hutt, K. R.; Cheng, C. S.; Ju, B. G.; Ohgi, K. A.; Wang, J.; Escoubet-Lozach, L.; Rose, D. W.; Glass, C. K.; Fu, X. D.; Rosenfeld, M. G. *Cell* **2007**, *128*, 505−518.
(54) Patel, R. Y.; Doerksen, R. J. *J. Proteome Res.* **2010**, *9*, 4433−4442.