Alchemical FEP Calculations of Ligand Conformer Focusing in Explicit Solvent

Alexey A. Zeifman,[†,‡] Victor V. Stroylov,[†,§] Fedor N. Novikov,[†,§] Oleg V. Stroganov,[†,§] Val Kulkov,[‖] and Ghermes G. Chilov[†,§,*]

[†]N. D. Zelinsky Institute Of Organic Chemistry (ZIOC RAS), Leninsky avenue, 47, 119991 Moscow, Russian Federation
[‡]National Research Centre "Kurchatov Instititute", Akademika Kurchatova pl., 1, 123182 Moscow, Russia
[§]Molecular Technologies, Ltd., Leninskie gory, 1/75G, 119992, Moscow, Russian Federation
[‖]BioMolTech Corp., 226 York Mills Road, Toronto, Ontario M2L 1L1, Canada

Ⓢ *Supporting Information*

ABSTRACT: Slow rotational degrees of freedom in ligands can make alchemical FEP simulations unreliable due to inadequate sampling. We addressed this problem by introducing a FEP-based protocol of ligand conformer focusing in explicit solvent. Our method involves FEP transformations between conformers using equilibrium dihedral angle as a reaction coordinate and provides the cost of "focusing" on one specific conformational state that binds to a protein. The calculated conformer focusing term made a considerable difference of 5−10 kJ/mol in computed relative binding free energies of studied Syk inhibitors and significantly improved the resulting accuracy of predictions.
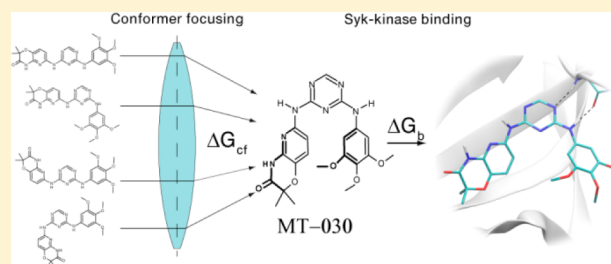
INTRODUCTION

Fast and accurate prediction of binding affinities is a core goal of computational medicinal chemistry. There is usually a trade-off between speed and accuracy. Empirical force field molecular docking[1] methods allow efficient high-throughput computation but lead to significant errors (~2 kCal/mol). These errors seem to be caused by the wrong implicit accounting for several explicit effects, such as protein flexibility and solvation.[2] Significantly more accurate are the statistical mechanics-based methods such as free energy perturbation (FEP)[3] and thermodynamic integration (TI).[4] Both methods are theoretically straightforward, and they should provide precise free energy values when perfect force field is employed and adequate sampling is achieved.[5] The increased accuracy of the statistical mechanics-based methods comes at higher computational cost, and therefore, these methods are typically used during the lead optimization stage of structure based drug design.

The conventional FEP strategy to calculate relative binding affinities of ligands L and L′ includes simulation of L to L′ transition in aqueous ($\Delta G_{aq}^{L \rightarrow L'}$) and protein ($\Delta G_{protein}^{L \rightarrow L'}$) environment. One can then obtain the resulting difference in free energy of binding by thermodynamic cycle closure: $\Delta\Delta G_{binding}^{L \rightarrow L'} = \Delta G_{binding}^{L'} - \Delta G_{binding}^{L} = \Delta G_{protein}^{L \rightarrow L'} - \Delta G_{aq}^{L \rightarrow L'}$.[6] This approach was successfully applied in the design of HIV-1 reverse transcriptase,[7] secreted phospholipase A2[8] and p38α MAP kinase inhibitors,[9] and theoretical studies of cathechol-O-methyltransferase inhibitors,[10] N1 influenza neuraminidase inhibitors,[11] glucosidase inhibitors,[12] etc., and is currently recognized as a standard FEP protocol.

The conformation of bound ligand is usually not the lowest-energy conformation of the unbound ligand. The free energy that is required to convert all conformers existing in the solution into the binding one is termed "conformer focusing" (CF).[13] The CF energy may be significant enough to introduce systemic errors in energy estimates that ignore this energy term. For example, in a study of 150 protein−ligand complexes, the bound conformer was shown to be 4−5 kcal/mol higher in energy on average than the lowest-energy conformer in solution.[14] The ligand's conformational space was directly addressed by Tirado-Rives and Jorgensen who employed QM and MM calculations in implicit solvent to estimate the CF energy. The authors established that the absence of explicit accounting for the CF energy term in docking or scoring methods is expected to introduce random noise of 0−10 kcal/mol[13] in the predicted values of binding energy.

The evaluation of CF energy by FEP in explicit solvent was the main objective of our work. Computation of ΔG_{cf} requires knowledge of the population of conformational states for both the unbound and bound ligand. However, a problem arises when L or/and L′ occupy different conformation states that are separated by high free energy barriers and thus are slowly interconverting. Such slow transitions are typical for large protein molecules; however, they can also occur in small

molecules. Several organic moieties (e.g., diarylamines, amides, biaryls, etc.) are known to exist in two or more slowly interchangeable conformations, and thus, molecules containing these groups cannot be adequately sampled under normal conditions. Sterical hindrance caused by bulky groups can also prevent adequate sampling in some cases.[13]

Genuine approaches to accelerate local conformational sampling in FEP calculations were suggested by Yang et al.,[15,16] in which dihedral potentials of interest (which govern the desired conformational transitions) were considered as a perturbation parameter and appropriate schemes of λ sampling were developed. However, further practical application of these ideas would need appropriate modification of the most commonly used molecular mechanics programs in order to give a user enough freedom to modify energy potentials and λ sampling schemes. Recently, several groups attempted to rectify this pitfall of the conventional FEP protocol. In 2010, Roux et al. proposed the replica exchange molecular dynamics protocol to account for the existence of kinetically trapped conformations within the protein receptor during calculations of ligand's absolute binding free energy.[17] In that protocol, λ and temperature are incremented independently, thus providing extensive sampling at each λ value but resulting in large computational costs. A more computationally efficient implementation of this idea was proposed by Friesner et al.[18] In their FEP/REST method, λ and temperature are modified simultaneously: λ steadily increases during simulation while temperature rises when $0 < \lambda < 0.5$ and falls when $0.5 < \lambda < 1$, thereby resulting in similar sampling performance at smaller computational cost. Both methods require no prior knowledge of the exact location of slow degrees of freedom; however, the latter method involves an assumption that such slow degrees of freedom are located near the bound ligand.

Relative binding affinities of different conformations of the same ligand molecule can be estimated using dual-topology FEP.[19] Such a method does not require dihedrals rotation, but due to large changes between A and B states, frequent λ spacing ($\Delta\lambda = 0.001$) is needed to achieve adequate phase space overlap. A different approach was proposed by Mobley et al. in an effort to account for the slow protein transition that occurs upon ligand binding to T4 lysozyme.[20] His 'confine-and-release' method utilizes a closed thermodynamic cycle, which includes transformation of *holo* protein into *apo* conformation by umbrella sampling of rotated dihedral ('confine'), FEP simulation of ligand binding to prepared *apo* protein conformation, and removal of applied potential from protein dihedral ('release'). Such a method provides a deeper insight into the contribution of conformational energy to the whole free energy of binding, but it requires a prior knowledge of protein's slow degrees of freedom. A similar approach based on umbrella sampling was used for the conformer's contribution to the free energy of hydration.[21]

The same problem arises when the ligand has different noninterconverting binding poses (which still could be sampled in the solution). Estimation of absolute binding free energy for each pose allows one to accurately handle this situation,[22] however, at a high computational cost.

It should be noted that the reviewed FEP-based methods do not allow one to directly estimate the population of ligand's conformational states and their contribution to the final free energy of binding. Knowledge of the population of ligand's conformational states can also be very useful for a medicinal chemist in guiding lead optimization.

We have developed a method of CF that rigorously accounts for the slow ligand conformer transitions in FEP simulations while treating solvent explicitly. The method produces free energy values that are consistent with alchemical FEP energies (as opposed to the empirical estimates of CF) and gives useful additional information on the population of each ligand conformer. We applied this method to solve a conformation sampling problem that arouses the design of novel Syk kinase diarylamine inhibitors (ref 23; Figure 1), which are of great
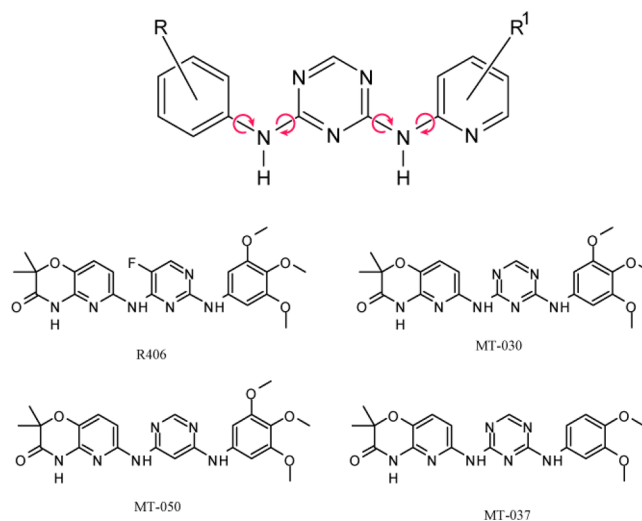
Figure 1. Diarylamino-triazine based SYK inhibitor possesses four slow degrees of freedom. Rotation around each Aryl-NH bond is impaired, and thus, $2^4 = 16$ separated conformers may exist in the solution.

potential utility in the treatment of autoimmune and oncologic diseases.[24] These inhibitors have different conformations in water, but only one conformer binds to the protein according to X-ray data.[25] Significant potential barriers between these conformers lead to very slow transition rates. In consequence, an alchemical transformation in water produces transition energy between ligands in only one conformation. Given 16 non-interconverting conformers, accounting for only one enzyme-bound conformer is expected to produce considerable error.

To systematically sample ligand's slow degrees of freedom, we calculate all conformational transitions for each ligand (or even protein) by means of FEP. The equilibrium dihedral angle corresponding to potential energy minimum is used as a transition coordinate; its initial and final values of 180° and −180° correspond to $\lambda = 0$ and $\lambda = 1$. Such boundary conditions are represented by the same potential functions for both conformers, but since the intermediate equilibrium dihedral at each λ step is obtained as linear composition,[26]

$$V(\varphi) = k_d(1 + \cos[n_\varphi \varphi - (1-\lambda)\varphi_s^A - \lambda \varphi_s^B]) \quad (1)$$

Thus, monotonically changing from 180° to −180°, FEP results in a conversion of one conformer to another (Figure 2).

The conformer transition energy $\Delta G_{B\rightarrow A}$ is obtained as the free energy of FEP transformation (derivation is similar to refs 27 and 28). When only two ligand L conformers (as in N-methylformamide case), namely A and B, are present, the resulting solution concentration of conformer A is equal to
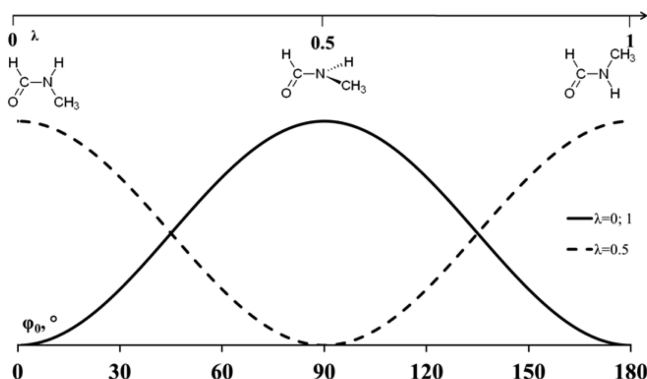
Figure 2. N-methylformamide conformer transition scheme. Solid and dashed lines represent the dihedral potential of H−C−N−C for $\lambda = 0$, 1, and 0.5, respectively. As λ increases, the nearest potential minimum moves to the right; $\lambda = 0$ corresponds to the *cis*-amide conformation, $\lambda = 1$ to the *trans*-amide conformation, when $\lambda = 0.5$ amide substituents are in perpendicular planes. One should mention that the potential functions for $\lambda = 0$ and $\lambda = 1$ are identical.

$$[A]_{eq} = \frac{1}{1 + e^{-\Delta G_{B\to A}/RT}} C_{tot} \tag{2}$$

where $\Delta G_{B\to A}$ is the energy of B to A conversion, and C_{tot} is the total concentration of the ligand. Assuming that only one of the conformers (A) tightly binds to the protein, one can obtain the conformation focusing energy:

$$\Delta G_{CF} = RT \ln(1 + e^{-\Delta G_{B\to A}/RT}) \tag{3}$$

When ligand L has more than one slow degree of freedom and thus possesses several conformers in solution (L_1, L_2, ..., L_n), among which only L_1 binds to the protein, the focusing energy can be obtained in quite the same manner:

$$\Delta G_{CF}^{L} = RT \ln(1 + \sum_{n}^{i=2} e^{-\Delta G_{L_i \to L_1}/RT}) \tag{4}$$

This term for both L and L′ is finally included into the resulting free energy expression (Figure 3):

$$\Delta G_{binding}^{L'} - \Delta G_{binding}^{L} = \Delta G_{protein}^{L\to L'} - \Delta G_{aq}^{L\to L'} + \Delta G_{CF}^{L'} - \Delta G_{CF}^{L} \tag{5}$$

One can anticipate that at least one ligand (L or L′) binds to the protein in more than one conformation (namely L'_1, L'_2, ..., L'_m where $m \leq n$). Additional steps are then required to calculate the final binding free energy. By definition,
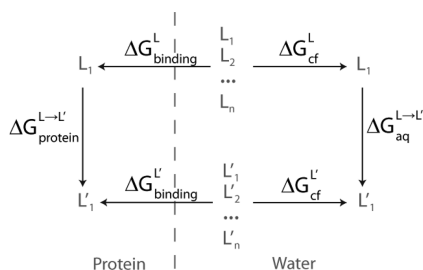

Figure 3. Thermodynamic cycle for a conformer accounting in free energy of binding calculations (only L_1 and L'_1 conformers bind to the protein).

$$\Delta G_{binding}^{L'} = -RT \ln \frac{[EL']_{tot}}{[E][L']_{tot}} = -RT \ln \frac{\sum_{j=1}^{m} [EL'_j]}{[E][L']_{tot}} \tag{6}$$

Given the difference in free energy of binding between each L'_j and L_1, which can be obtained for example through conformer focusing in protein (Figure 4)

$$\Delta G_{binding}^{L'_j} - \Delta G_{binding}^{L_1}$$
$$= \Delta G_{protein}^{L'_1 \to L_1} + \Delta G_{protein}^{L'_j \to L'_1} - \Delta G_{water}^{L'_1 \to L_1} - \Delta G_{water}^{L'_j \to L'_1} \tag{7}$$
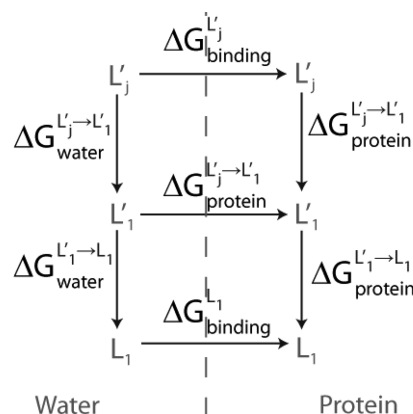
Figure 4. Additional FEP transitions are required when ligand binds to the protein in more than one conformation. Conformer transition of L'_j to L'_1 in water and protein followed by alchemical transformation of L'_1 to L_1 in water and protein finally gives $\Delta G_{binding}^{L'_j} - \Delta G_{binding}^{L_1}$ free energy difference.

, one can obtain the concentration of EL'_j species:

$$[EL'_j] = \exp\left(-\frac{\Delta G_{binding}^{L'_j} - \Delta G_{binding}^{L_1}}{RT}\right)[E][L'_j]$$
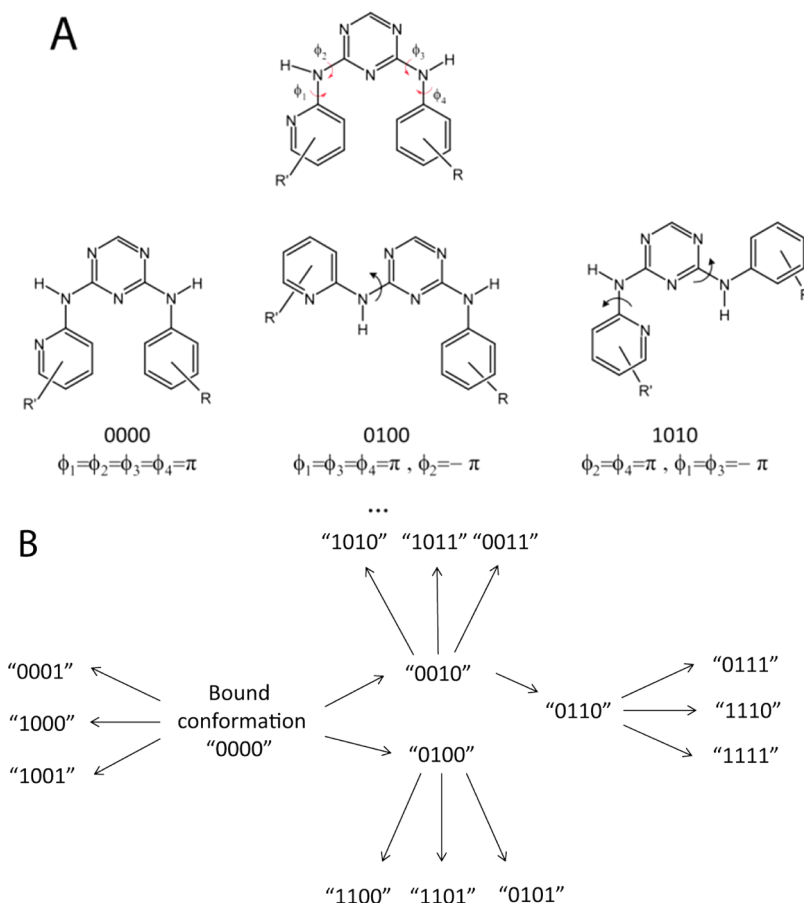$$\exp\left(\frac{\Delta G_{binding}^{L_1}}{RT}\right) \tag{8}$$

With the conformer focusing data one can calculate L'_j population:

$$[L'_j] = \exp\left(-\frac{\Delta G_{cf}^{L' \to L'_j}}{RT}\right)[L']_{tot} \tag{9}$$

Given eqs 8 and 9, one can modify eq 6 to

$$\Delta G_{binding}^{L'} = -RT \ln(e^{\Delta G_{binding}^{L_1}/RT})$$
$$\sum_{m}^{j=1} \exp\left(-\frac{\Delta G_{binding}^{L'_j} - \Delta G_{binding}^{L_1} + \Delta G_{cf}^{L' \to L'_j}}{RT}\right) \tag{10}$$

The resulting difference in final free energy of binding is obtained in the following form:

Figure 5. (A) Conformer designation. "0" and "1" in conformer designation denote conformations on each of the $\varphi_1-\varphi_4$ dihedrals with "0" referring to the conformation of bound ligand. (B) Proposed conformer transition scheme for SYK kinase inhibitors.

$$\Delta G_{\text{binding}}^{L'} - \Delta G_{\text{binding}}^{L}$$
$$= -RT \ln \sum_{m}^{j=1} \exp\left(-\frac{\Delta G_{\text{binding}}^{L'_j} - \Delta G_{\text{binding}}^{L_1} + \Delta G_{\text{cf}}^{L' \to L'_j}}{RT}\right) \tag{11}$$

which degenerates to eq 5 if only one binding conformer is present.

Such an approach is theoretically straightforward, and it delivers several advantages. First, it improves FEP accuracy. Second, it makes possible the analysis of the free energy of each conformation transition, which can be useful during lead optimization as the population of undesired conformer can be reduced and the population of binding conformer can be increased. To our knowledge, this is the first protocol for direct conformer focusing accounting in FEP calculations.

The proposed conformer focusing protocol was applied to study conformer population of the novel Syk kinase inhibitors. The obtained conformer focusing energy significantly differed for different inhibitors, and its accounting improved the overall accuracy of relative free energy of binding predictions. Interestingly, the lowest conformer focusing energy was observed for the most potent compound in R406[29] series, pointing to the potential utility of the proposed method in lead optimization.

■ METHODS

General Points. The molecular dynamics simulations were performed using the GROMACS simulation package[24] with the OPLS/AA force field parameter set.[30] The topologies for SYK inhibitors were created using ACPYPE[31] in conjunction with Antechamber.[32] The resulting free energy differences were calculated using BAR.[33] Free energy errors in each λ window of FEP simulations were estimated using block averaging over 5 blocks ($\pm1\sigma$); the errors of the free energy of completed FEP transitions were calculated using conventional statistic rules for arithmetic operations. The error of the free energy of CF, which is a composite of individual interconformer transitions, was estimated as follows. First, the minimum, the maximum, and the average values of relative occupancy of the active conformer (which binds to the protein) were estimated according to the formula:

$$[L_1^{\max}] = 1 + \sum_{n}^{i=2} e^{-\Delta G_{L_i \to L_1} + \sigma/RT} \tag{12}$$

$$[L_1^{\min}] = 1 + \sum_{n}^{i=2} e^{-\Delta G_{L_i \to L_1} - \sigma/RT} \tag{13}$$

$$[L_1^{\text{av}}] = 1 + \sum_{n}^{i=2} e^{-\Delta G_{L_i \to L_1}/RT} \tag{14}$$

0000 $\frac{1}{2}00\frac{1}{2}$ **1000**

$\lambda=0$ $\lambda=0.5$ $\lambda=1$
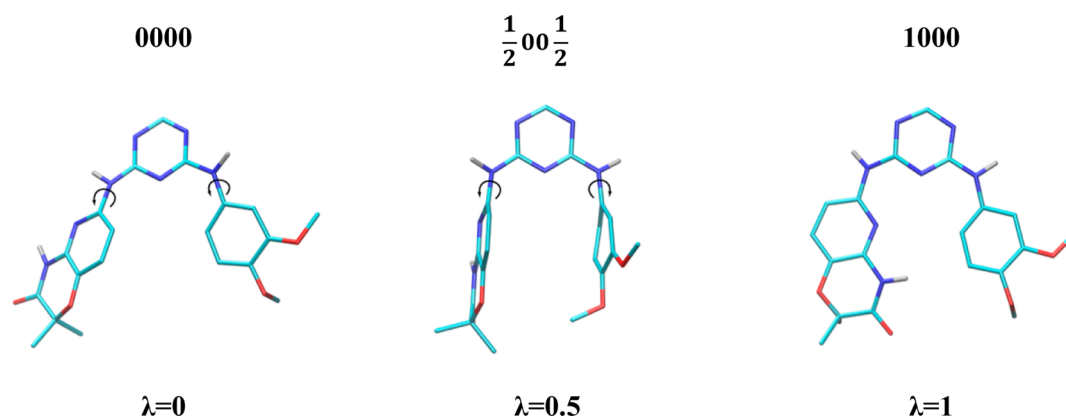
Figure 6. Conformational transition from "0000" to "1000" in MT-037 SYK inhibitor. At $\lambda = 0$, both side cycles are in the same plane, and NH and pyrimidine N are at the same side; at $\lambda = 0.5$, lateral cycles are perpendicular to central triazine cycle; at $\lambda = 1$, all cycles are again in one plane but NH and pyrimidine N are on the different sides of C−N bond.

Then, the maximum free energy difference between $[L_1^{max}]$ and $[L_1^{av}]$ and $[L_1^{min}]$ and $[L_1^{av}]$ is taken as a precision of CF calculation:

$$\delta_{\Delta G_{CF}} = \max\left(RT \ln\left(\frac{[L_1^{max}]}{[L_1^{av}]} \right); \; RT \ln\left(\frac{[L_1^{av}]}{[L_1^{min}]} \right) \right) \quad (15)$$

The phase space overlap between neighboring λ was obtained by integrating overlap in free energy histograms.

The initial coordinates of R406 bound to the SYK kinase were obtained from the crystal structure with PDB entry code 3FQS.[25] Crystallographic water molecules were removed, and the full-atom protein model was prepared using conventional GROMACS tools. The protein bound conformations of the new SYK inhibitors (designated below as "0000", Figure 5A) were constructed from R406 bound X-ray conformation by atom replacement followed by 5000 steps of steepest descent and 5000 steps of l-bfgs MD vacuum minimization in vacuum to a tolerance of 10 kJ·mol^{-1}·nm^{-1}.

Alchemical FEP Transformations. For alchemical FEP simulations in solution ligands were solvated in octahedral box with approximately 1400 TIP4P[34] water molecules. Simulations of the ligands bound to Syk kinase were performed in rectangular periodic boxes containing 12480 TIP4P water molecules and three chloride ions to neutralize the total system charge. Transformations were performed in 11 separate charge ($\Delta\lambda = 0.1$) and 29 VdW ($\Delta\lambda = 0.01$ for $\lambda = 0-0.05$ and $\lambda = 0.95-1$, $\Delta\lambda = 0.05$ for $\lambda = 0.05-0.95$) steps. Simulation at each distinct λ value included 5000 steps of steep and l-bfgs energy minimization, 100 ps NVT equilibration, 500 (water) or 1000 (protein) ps NPT equilibration, and 2 ns (ligand in solution) or 5 ns (ligand bound to protein) NPT runs with $dH/d\lambda$ collection each 10 steps. The protein side-chain during NVT and NPT simulations was set fully flexible while the backbone was fixed by position restraints of 1000 kJ·mol^{-1}·nm^{-2} to improve convergence.

Ligand Conformer Focusing. The conventional alchemical FEP protocol was slightly modified to convert conformer A to B in order to obtain their relative population in solution (starting from the conformer A geometry). In the topology file all nonbonded atoms', bonds', and angles' parameters remained the same for both A and B states while several dihedrals were perturbed. The rotation around a single bond involves simultaneous change of four adjacent dihedral angles (for example, in N-methylformamide (Figure 2) such dihedrals are

H−C−N−H, O−C−N−H, H−C−N−C, and O−C−N-C). A modification of these dihedrals' parameters in the FEP topology leads to the conversion of one conformer to another (force constant k_d and phase multiplier n_φ remains the same for both states, while φ_s changes from 180° to −180°; see formula 1).

Serial minimization in vacuum starting from conformer A structure was employed to generate initial ligand geometries for each of the 29 λ windows ($\Delta\lambda = 0.01$ for $\lambda = 0-0.05$ and $\lambda = 0.95-1$, $\Delta\lambda = 0.05$ for $\lambda = 0.05-0.95$). The resulting ligand structures at every λ were solvated, minimized, and equilibrated for 100 ps NVT and 100 ps NPT dynamics and simulated for 2 ns NPT production dynamics. This scheme was used "as is" for N-methylformamide (the model system); however, several further modifications were introduced for Syk inhibitors due to their complex structure, as described below.

The studied Syk inhibitors possessed 4 dihedrals and $2^4 = 16$ conformers corresponding to the minimum of the dihedral energy. To traverse over those conformers with minimal structure alteration during each FEP round, we developed a special transition scheme (Figure 5B). Binary numbers from "0000" to "1111" were assigned to each ligand conformer respective to the conformation around φ_1, φ_2, φ_3, and φ_4 bonds. The basic "0000" conformation depicted in Figure 5A corresponds to the protein-bound form. Switching the i-th digit from 0 to 1 corresponds to the flip over i-th bond in the ligand molecule. For example, "1000" (Figure 6, right) means that the conformations on all bonds except φ_1 remain the same as in the bound structure; "1111" denotes a conformer in which all dihedral angles are flipped relative to the bound structure.

The internal geometry of studied compounds necessitated some special treatment. The rotation around peripheral bonds φ_1 and φ_4 is hindered by the bulky side rings: at $\lambda = 0.5$, one ring does not allow another to continue its rotation due to steric clash. Such a problem can be solved by performing absolute binding free energy calculations with dual-topology FEP[35] but at a considerable computational cost, especially for large ligands. We observed that in the case of SYK inhibitors the problem could be avoided when both rings are rotated simultaneously as in "0xx0" to "1xx1" transition because the impossible perpendicular conformation is not reached at any λ. Thus, in order to perform a single side ring rotation ("0xx0" to "1xx0" or "0xx1") both side rings are rotated. In the first part of FEP simulation (λ changes from 0 to 0.5), both rings are rotated in the same direction and "0xx0" conformation is

transformed to the unstable "$(^1/_2)xx(^1/_2)$" state, in which the side rings are perpendicular to the central one (Figure 6). During the second FEP run ($0.5 < \lambda < 1$), one ring continues rotation while the second ring rotates back to its initial state thus resulting into "1xx0" conformation. This trick is theoretically correct because the free energy is a function of state and it does not depend on the system path. Each run consisted of 15 λ windows that included energy minimization, NVT, and NPT equilibration and production run, as described above; a total of 30 runs were required for one transition.

Experimental Methods. All experimental Syk kinase inhibitors were chemically synthesized, as described previously.[23]
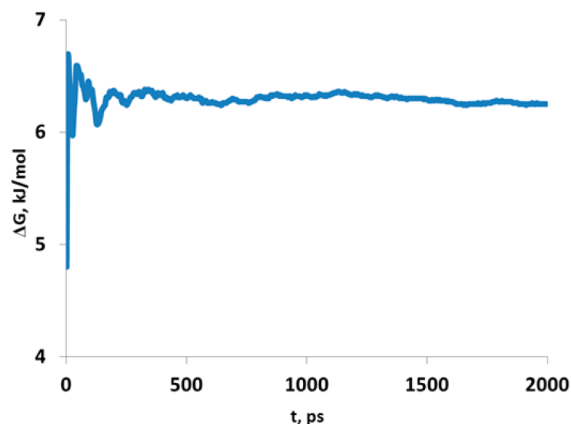
Syk kinase inhibition assay was conducted in Reaction Biology Corp.[36]

■ RESULTS AND DISCUSSION

We evaluated the proposed CF method on a simple model system (*cis–trans* isomerization of NMF in water) and then
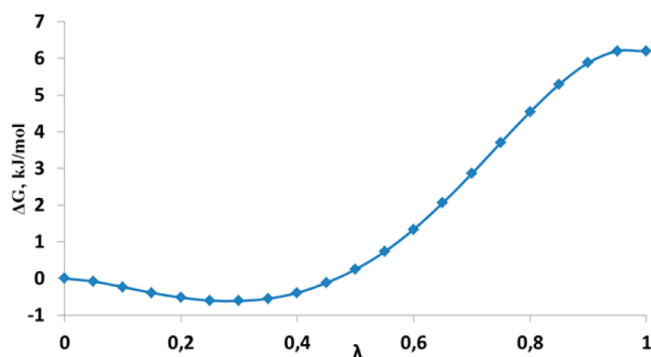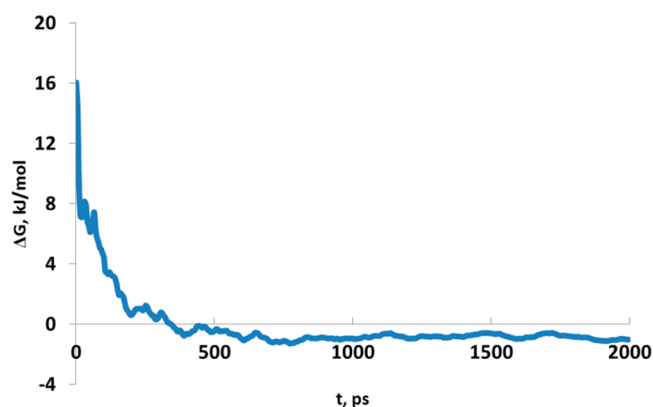
Figure 7. Cumulative free energy as a function of λ for NMF *trans* to *cis* transition.
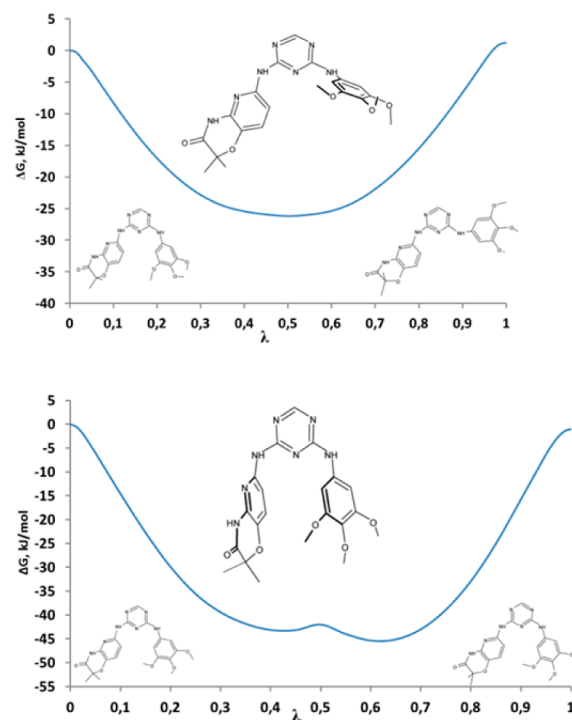
Figure 8. Cumulative free energy as a function of simulation time for NMF *trans* to *cis* transition.

applied it to estimate the binding affinity of novel Syk kinase inhibitors. The CF calculations were validated by the estimation of phase space overlap and convergence and accurate reproduction of known experimental population data for the model system. The population of individual Syk ligands' conformers was considerably influenced by their chemical structure and accounting of explicit solvation. Introduction of the CF free energy term into the final relative free energy of binding did improve the accuracy of predictions.

Figure 9. Cumulative free energy as a function of simulation time for the 0000→0001 MT-030 conformation transition.

Figure 10. Cumulative free energy curve as a function of λ for 0000→ 0010 and 0000→0001 transitions of MT-030. First transition was achieved by rotation of one side ring while in the second transition second side ring was additionally reversibly rotated to avoid sterical clashes.

N-methylformamide (NMF) Model Compound Study. The stepwise free energy change as a function of λ shown in Figure 7 demonstrated that ΔG differences between the neighboring λ windows were rather small ($<kT$) and that indicated a considerable phase space overlap. The total free energy of transition as a function of simulation time is given in Figure 8. Free energy stabilized at 500 ps, which indicated adequacy of the selected trajectory length and convergence of the simulations. The resulting free energy difference calculated at 2 ns $\Delta G_{trans\rightarrow cis}$ = 6.17 ± 0.04 kJ/mol was in a perfect agreement with the published experimental value of 6.05 kJ/mol (*cis–trans* ratio 8:92[37] in water).

The obtained free energy curve for NMF (Figure 7) did not show a sharp maximum at λ = 0.5, which would correspond to the unfavorable twisted conformation. Such smoothening of the

Table 1. Relative Population of Syk Ligands' Conformers in Water[a]

	relative population			
conformer	R406	MT-030	MT-050	MT-037
0000	1.00	1.00	1.00	1.00
0001	1.31	1.50	2.75	0.54
0010	0.73	0.61	5.97	12.4
0011	0.95	0.63	11.67	3.64
0100	**0.01**	**4.14**	**0.72**	**12.55**
0101	**0.00**	**3.26**	**0.41**	**2.49**
0110	**0.01**	**3.48**	**1.39**	**79.93**
0111	**0.00**	**4.75**	**1.11**	**34.12**
1000	*0.05*	*0.00*	*0.53*	*0.00*
1001	*0.05*	*0.00*	*0.46*	*0.00*
1010	*0.02*	*0.00*	*6.17*	*0.01*
1011	*0.02*	*0.00*	*7.46*	*0.00*
1100	0.00	0.00	0.00	0.01
1101	0.00	0.00	0.00	0.00
1110	0.00	0.00	0.00	0.05
1111	0.00	0.00	0.00	0.01

[a]**Bold** corresponds to "01xx" conformations, which are avoided by R406; *italic* denotes "10xx" conformations, which are occupied only for MT-050.
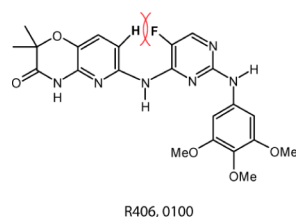

Figure 11. Sterical clashes between H and F makes 0100 conformer unfavorable for R406 but not for other compounds.
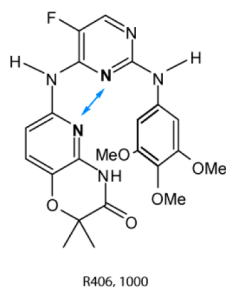

Figure 12. Repulsion of electronegative nitrogen atoms makes "1000" conformation unfavorable for R406, MT0-030, and MT-037.

conventional potential was due to the adjustment of equilibrium dihedral angle for each λ, resulting in small dihedral energy component at each λ window. The intermediate conformations obtained in this way are physically irrelevant (as it is usual in FEP simulations); however, the total free energy of the transition remains correct since ΔG is a function of state.

Syk Inhibitors Conformational Study. Conformer transitions of Syk ligands involved rotations around one or two dihedrals and were performed as described in Methods and Figure 5B. To validate simulations, we examined their phase space overlap and sampling efficacy. The phase space overlap between neighboring λ windows ranged from 0.05 to 0.9, which, taken alone, could not be used as a criterion of

Table 2. Conformer Population of MT-050 in Vacuum and in Water[a]

	MT-050	
conformer	vacuum	water
0000	1.00	1.00
0001	0.99	2.75
0010	**0.00**	**5.97**
0011	**0.00**	**11.67**
0100	1.18	0.72
0101	1.13	0.41
0110	**0.00**	**1.39**
0111	**0.00**	**1.11**
1000	0.69	0.53
1001	0.94	0.46
1010	**0.00**	**6.17**
1011	**0.00**	**7.46**
1100	0.00	0.00
1101	0.00	0.00
1110	0.00	0.00
1111	0.00	0.00

[a]Conformations, which population differs significantly between water and vacuum, are set **bold**.

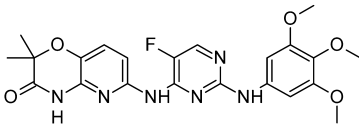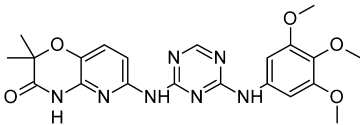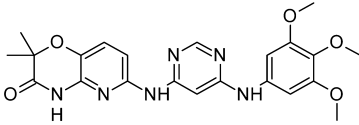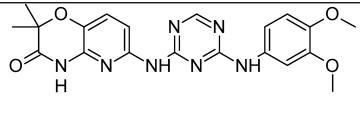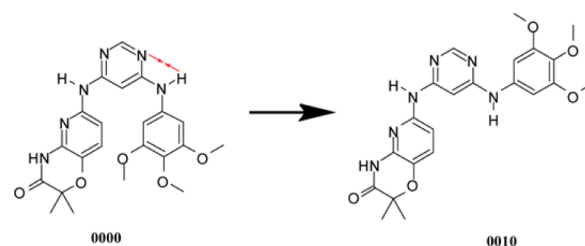**Figure 13.** Intramolecular electrostatic interaction between ring N and amine NH are disrupted in 0010 conformation of MT-050. Solvation significantly weakens these interaction in both conformations making 0010 state populated.

simulation quality. Hence, we studied if the introduction of additional intermediate λ windows would change the resulting free energy. No statistically significant difference was observed for several transformations (data not shown), indicating that the adequate acceptance ratio had already been achieved. The cumulative free energy calculated at different simulations time stabilized at ~500 ps (Figure 9), indicating good convergence of the simulations and adequate length of the trajectory.

Having assessed the validity of undertaken simulations, we studied the energy profile they produced. Conformer transitions in Syk kinase inhibitors are necessarily accompanied by the displacement of large molecule fragments, and thus, a rather complex energy pattern may arise. We explored the cumulative free energy as a function of λ for transformations involving the rotation of one and two dihedral angles (Figure 10). Both transition paths demonstrated the same general form with one distinct minimum at $\lambda = 0.5$, which seemed to be caused by the weakening of intramolecular repulsive interactions due to the perpendicular orientation of aromatic rings. When two dihedrals were rotated instead of one, more unfavorable contacts were weakened, and that resulted in approximately 2-fold deeper free energy minimum.

The proposed conformation focusing protocol implies the direct calculation of each conformer's relative free energy and, consequently, its population. Average concentrations of con-

Table 3. Conformer Focusing Energy and Experimental IC$_{50}$ for Studied Compounds

R406

$\Delta G_{cf}=1.8\pm0.5$ kJ/mol

IC$_{50}$ = 32 nM

MT-030

$\Delta G_{cf}=5.6\pm0.5$ kJ/mol

IC$_{50}$ = 635 nM

MT-050

$\Delta G_{cf}=7.4\pm1.1$ kJ/mol

IC$_{50}$>10000 nM

MT-037

$\Delta G_{cf}^{0000}=12.4\pm0.6$ kJ/mol

$\Delta G_{cf}^{0001}=14.0\pm0.7$ kJ/mol

IC$_{50}$>10000 nM
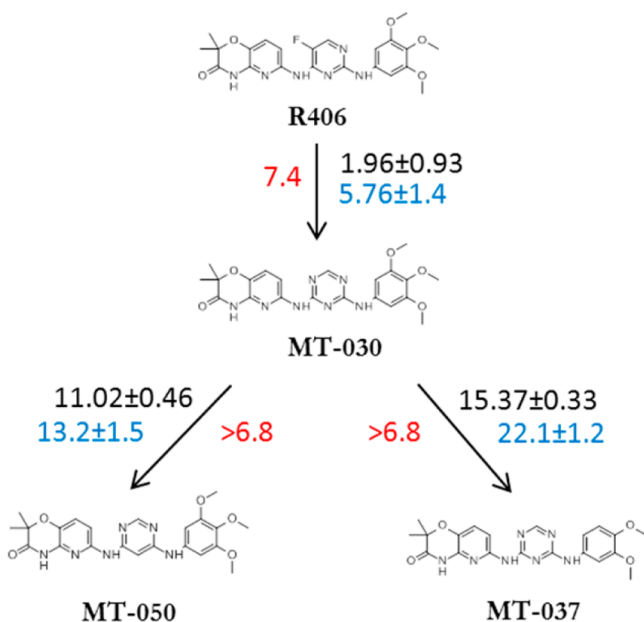
Figure 14. Graphical representation of predicted differences of free energy of binding in kJ/mol. $\Delta\Delta G_{binding}$ are colored black, $\Delta\Delta G_{binding+cf}$ are blue, and $\Delta\Delta G_{exp}$ are red.

formers of studied compounds in water were estimated and given in Table 1, relative to the concentration of protein-binding conformer "0000" which was set to be 1. The additional data (conformers' free energy with confidence interval and corresponding minimal and maximal population) are provided in the Supporting Information.

The first notable point concerns the symmetry of right aromatic ring of R406, MT-030, and MT-050 that makes "xxx0" and "xxx1" conformers structurally identical. The predicted average populations of those states were quite close to each other (in fact, equal within the error range), and that again points to the validity of the method.

To further assess the validity of the generated populations, we attempted to qualitatively interpret them. Some conformers were populated in a substantially different way than the others, and that represented a good departure point for interpretation. First, "01xx" conformations (Table 1, **bold**) were strongly avoided by R406 but not by other compounds. We suppose that such disfavor was caused by the sterical clashes between H and F in "01xx" conformers of R406, which were absent in other compounds (Figure 11).

Another notable difference was observed for "10xx" conformations (Table 1, *italic*) which were avoided by all compounds except MT-050. The high free energy of "10xx" conformer seemed to be caused by the intramolecular electrostatic repulsion (Figure 12), which was absent in MT-050.

The proposed interpretations do not directly consider the influence of solvent and thus to study the impact of solvent we performed similar calculations in vacuum. The most pronounced difference between populations in water and vacuum was observed for MT-050 (the results for other compounds are provided in the Supporting Information).

Several conformers of MT-050 (Table 2, **bold**) that had inversion around the third dihedral angle were not populated at all in vacuum but were considerably populated in solution, except "1110" and "1111" conformers that were not populated in water either. We suppose that the rotation around the third

Table 4. Relative Free Energies of Binding of Syk Ligands

	$\Delta\Delta G_{water}$, kJ/mol	$\Delta\Delta G_{protein}$, kJ/mol	$\Delta\Delta G_{cf}$, kJ/mol[b]	$\Delta\Delta G_{binding}$, kJ/mol	$\Delta\Delta G_{binding+cf}$, kJ/mol	$\Delta\Delta G_{exp}$, kJ/mol
R406→MT-030	−2.13 ± 0.28	−0.17 ± 0.89	3.8 ± 1.0	1.96 ± 0.93	5.76 ± 1.4	7.4
MT-050→MT-030	11.97 ± 0.45	0.95 ± 0.11	−1.8 ± 1.3	−11.02 ± 0.46	−13.2 ± 1.5	←6.8
MT-037→MT-030	2.90 ± 0.32	−12.47 ± 0.08	−6.8 ± 0.9	−15.37 ± 0.33	−22.1 ± 1.2	←6.8
		−21.36 ± 0.19[a]	−8.4 ± 1.0[a]	−24.26 ± 0.37[a]		

[a]For the "0001" conformer. [b]Error estimates were obtained using formula 15.
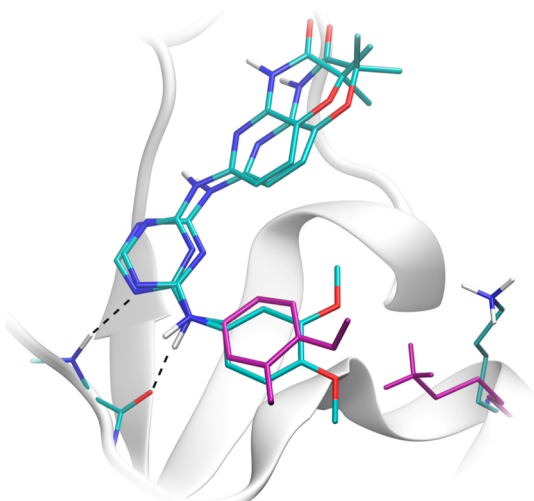
Figure 15. MT-037 possesses two distinct binding modes, which differ by the orientation of dimethoxybenzene ring. Other compounds have only one binding mode due to the symmetry.

dihedral angle disrupted the attractive intramolecular electrostatic interaction (Figure 13) and made the produced conformer less favorable. In water, that interaction was considerably weaker, however, and that resulted in smaller free energy change during rotation and comparable populations of such conformers. The presence of an additional nitrogen atom in other compounds conserved the same favorable interaction in "xx1x" conformer, making such conformer states plausible in vacuum.

Syk Binding Affinity Calculations. The final free energy of conformer focusing of Syk inhibitors together with IC_{50} values obtained in vitro are given in Table 3.

The predicted binding affinities of Syk ligands are shown in Figure 14 and Table 4.

Accounting for the CF energy dramatically improved the accuracy of free energy prediction for R406→MT-030 transition: the conventional $\Delta\Delta G_{binding}$ significantly differed from the experimental value while $\Delta\Delta G_{binding+cf}$ matched $\Delta\Delta G_{exp}$ within the error range. For the other two transitions (MT-050→MT-030 and MT-037→MT-030), both $\Delta\Delta G_{binding}$ and $\Delta\Delta G_{binding+cf}$ were considerably lower than the experimental threshold of −6.8 kJ/mol. We must note, however, that owing to the fact that MT-037 has two distinct protein-binding conformers, conformer focusing is the only way to accurately account for both binding conformers (Figure 15), and therefore, the proposed methodology (eq 10 in the Introduction) remains essential for the accurate handling of the MT-037→MT-030 transition.

Since the computation of the free energy of binding of Syk ligands involved the summation of FEP transformations for 16 conformers, one may expect significant error propagation from the individual experiments. However, the observed accuracy of

$\Delta\Delta G_{cf}$ prediction was not significant at approximately 1 kJ/mol which was quite close to the accuracy of alchemical perturbations. Interestingly, the lowest CF energy was observed with R406, the most potent and probably the most optimized inhibitor in the series. Its focusing energy indicated that approximately half of the R406 molecules in solution could efficiently bind to Syk kinase. On the other hand, only 1% of MT-037 molecules possessed active conformation indicating that MT-037 is not optimal in terms of conformer focusing. The moderately potent MT-030 demonstrated intermediate (∼5%) population of the protein-binding conformation.

The discussed results demonstrate relevance and utility of the proposed conformer focusing method. From the practical point of view, one might want to compare its computational cost with that of FEP/REST[18] method, which also potentially accounts for the different ligand conformers. Let us consider a simple system containing one amide bond possessing two possible conformations. The amide rotation kinetics is a first-order reaction with the pre-exponential factor A ∼ 10^{12} s^{-1} and $\Delta G^{\ddagger}$ ∼ 20 kCal/mol.[38] Adequate sampling of this degree of freedom in a 5 ns simulation can be achieved at 2000 K (about 70 transitions would occur at this temperature). With FEP/REST protocol at ΔT = 50 K, the total of 136 runs of 5 ns dynamics is required (assuming separate VdW and charge transitions). The proposed conformer focusing protocol requires 60 steps for the conformational energy estimation (30 for each ligand conformation) and 33 steps for the ligand interconversion (21 for VdW and 11 for charges perturbation is usually enough), thus requiring the total of 93 runs. However, if a ligand possesses more than one slow degree of freedom, the complexity of both methods grows exponentially.

■ CONCLUSION

The method of conformer focusing presented here provides a rigorous way to correct the problem of inadequate ligand sampling in FEP caused by the slow rotational degrees of freedom. The method delivers means to identify the most populous ligand conformer and determine if that conformer is the one that binds to the protein, thus providing valuable insight to the lead optimization scientists. The presented method is not principally restricted to the ligand simulation in water. It can potentially be used with other systems such as macromolecules.

■ ASSOCIATED CONTENT

ⓈSupporting Information

Structure for each of the 16 conformers of MT-030, conformational transition energies for R406, MT-030, MT-050, and MT-037, as well as relative population of its conformers. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*Phone: +7(495)135-53-13. Fax: +7(495)135-53-13. E-mail: ghermes.chilov@gmail.com.

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

ABBREVIATIONS

CCR2, CC chemokine receptor 2; CCL2, CC chemokine ligand 2; CCR5, CC chemokine receptor 5; TLC, thin layer chromatography

REFERENCES

(1) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3* (11), 935−949.
(2) Novikov, F. N.; Zeifman, A. A.; Stroganov, O. V.; Stroylov, V. S.; Kulkov, V.; Chilov, G. G. *J. Chem. Inf. Model.* **2011**, *51* (9), 2090−2096.
(3) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22* (8), 1420−1426.
(4) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3* (5), 300−313.
(5) *Free Energy Calculations in Rational Drug Design*; Reddy, M. R. E., Mark, D., Eds.; Springer: New York, 2001.
(6) Brandsdal, B. O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J. *Adv. Protein Chem.* **2003**, *66*, 123−158.
(7) (a) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2006**, *16* (3), 663−667. (b) Zeevaart, J. G.; Wang, L.; Thakur, V. V.; Leung, C. S.; Tirado-Rives, J.; Bailey, C. M.; Domaoal, R. A.; Anderson, K. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2008**, *130* (29), 9492−9499.
(8) Mouchlis, V. D.; Mavromoustakos, T. M.; Kokotos, G. *J. Comput.-Aided Mol. Des.* **2010**, *24* (2), 107−115.
(9) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2010**, *6* (12), 3850−3856.
(10) Palma, P. N.; Bonifacio, M. J.; Loureiro, A. I.; Soares-da-Silva, P. *J. Comput. Chem.* **2012**, *33* (9), 970−986.
(11) Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J. A. *Proteins* **2010**, *78* (11), 2523−2532.
(12) Ruiza, F. M.; Grigera, J. R. *Med. Chem.* **2005**, *1* (5), 455−460.
(13) Tirado-Rives, J.; Jorgensen, W. L. *J. Med. Chem.* **2006**, *49* (20), 5880−5884.
(14) Perola, E.; Charifson, P. S. *J. Med. Chem.* **2004**, *47* (10), 2499−2510.
(15) Li, H. Z.; Fajer, M.; Yang, W. *J. Chem. Phys.* **2007**, *127* (2), 024106.
(16) Min, D. H.; Li, H. Z.; Li, G. H.; Bitetti-Putzer, R.; Yang, W. *J. Chem. Phys.* **2007**, *126* (14), 144109.
(17) Jiang, W.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6* (9), 2559−2565.
(18) Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (6), 1937−1942.
(19) Michel, J.; Essex, J. W. *J. Med. Chem.* **2008**, *51* (21), 6654−6664.
(20) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Theory Comput.* **2007**, *3* (4), 1231−1235.
(21) Klimovich, P. V.; Mobley, D. L. *J. Comput.-Aided Mol. Des.* **2010**, *24* (4), 307−316.
(22) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. *J. Mol. Biol.* **2007**, *371* (4), 1118−1134.
(23) Zeifman, A. A.; Titov, I. Y.; Svitanko, I. V.; Rakitina, T. V.; Lipkin, A. V.; Stroylov, V. S.; Stroganov, O. V.; Novikov, F. N.; Chilov, G. G. *Mendeleev Commun.* **2012**, *22* (2), 73−74.
(24) (a) Friedberg, J. W.; Sharman, J.; Sweetenham, J.; Johnston, P. B.; Vose, J. M.; Lacasce, A.; Schaefer-Cutillo, J.; De Vos, S.; Sinha, R.; Leonard, J. P.; Cripe, L. D.; Gregory, S. A.; Sterba, M. P.; Lowe, A. M.; Levy, R.; Shipp, M. A. *Blood* **2010**, *115* (13), 2578−2585. (b) Weinblatt, M. E.; Kavanaugh, A.; Genovese, M. C.; Musser, T. K.; Grossbard, E. B.; Magilavy, D. B. *N. Engl. J. Med.* **2010**, *363* (14), 1303−1312.
(25) Villasenor, A. G.; Kondru, R.; Ho, H.; Wang, S.; Papp, E.; Shaw, D.; Barnett, J. W.; Browner, M. F.; Kuglstatter, A. *Chem. Biol. Drug Des.* **2009**, *73* (4), 466−470.
(26) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26* (16), 1701−1718.
(27) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125* (8), 084902.
(28) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1989**, *90* (6), 3300−3304.
(29) Braselmann, S.; Taylor, V.; Zhao, H.; Wang, S.; Sylvain, C.; Baluom, M.; Qu, K.; Herlaar, E.; Lau, A.; Young, C.; Wong, B. R.; Lovell, S.; Sun, T.; Park, G.; Argade, A.; Jurcevic, S.; Pine, P.; Singh, R.; Grossbard, E. B.; Payan, D. G.; Masuda, E. S. *J. Pharmacol. Exp. Ther.* **2006**, *319* (3), 998−1008.
(30) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(31) Sousa da Silva, A. W.; Vranken, W. F. *BMC Res. Notes* **2012**, *5*, 367.
(32) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25* (2), 247−260.
(33) Charles H, B. *J. Comput. Phys.* **1976**, *22* (2), 245−268.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(35) Michel, J.; Verdonk, M. L.; Essex, J. W. *J. Chem. Theory Comput.* **2007**, *3* (5), 1645−1655.
(36) Ma, H.; Deacon, S.; Horiuchi, K. *Expert Opin. Drug Discovery* **2008**, *3* (6), 607−621.
(37) Radzicka, A.; Pedersen, L.; Wolfenden, R. *Biochemistry* **1988**, *27* (12), 4538−4541.
(38) Hu, X.; Zhang, W.; Carmichael, I.; Serianni, A. S. *J. Am. Chem. Soc.* **2010**, *132* (13), 4641−4652.