JCTC Journal of Chemical Theory and Computation

Article

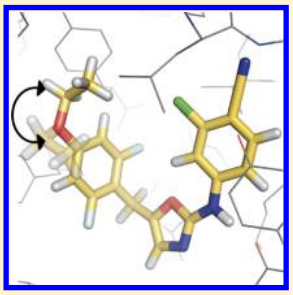
pubs.acs.org/JCTC

Enhanced Monte Carlo Sampling through Replica Exchange with Solute Tempering

Daniel J. Cole, Julian Tirado-Rives, and William L. Jorgensen*

Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United States

ABSTRACT: With a view to improving the consistency of free energy perturbation calculations in Monte Carlo simulations of protein−ligand complexes, we have implemented the replica exchange with solute tempering (REST) method in the *MCPRO* software. By augmenting the standard REST approach with regular attempted jumps in selected dihedral angles, our combined method facilitates sampling of ligand binding modes that are separated by high free energy barriers and ensures that computed free energy changes are considerably less dependent on the starting conditions and the chosen mutation pathway than those calculated with standard Monte Carlo sampling. We have applied the enhanced sampling method to the calculation of the activities of seven non-nucleoside inhibitors of HIV-1 reverse transcriptase, and its Tyr181Cys variant, and have shown that a range of binding orientations is possible depending on the nature of the ligand and the presence of mutations at the binding site.



1. INTRODUCTION

The computational ranking of binding affinities of a congeneric series of ligands to a protein is an invaluable technique in structure-based drug design. Of the many computational methods that have been developed for this purpose, free energy perturbation (FEP) calculations, in combination with molecular dynamics (MD) or Monte Carlo (MC) sampling, are particularly attractive because, in principle, they provide a rigorous means to compute the free energy of binding.[1] In practice, however, the predictive power of FEP calculations is limited by the accuracy of the force field and by finite simulation times that can prevent the exploration of important regions of conformational space.[2,3] In simulations of protein−ligand complexes, in particular, the ligand is often trapped for long times in local minima of the free energy surface, thus leading to quasi-ergodic sampling. This incomplete sampling of the ligand binding modes is problematic in FEP calculations, where the computed free energy of binding may then depend strongly on the starting configuration or the chosen mutation pathway.

Parallel tempering, or the replica exchange method (REM), is a powerful technique for overcoming quasi-ergodicity in small systems.[4,5] In REM, exchange of configurations with high temperature replicas of the system allows more frequent crossing of high potential energy barriers. However, the number of replicas required scales as the square root of the number of degrees of freedom in the system,[6] not only increasing the amount of processing power required for large systems but also limiting temperature diffusion in the system.

Hamiltonian REM is a similar concept to REM except that, instead of scaling the system temperature, the replicas have incrementally scaled potential energy surfaces, thus allowing the user more freedom in scaling selected components of the system Hamiltonian, such as Lennard-Jones interactions.[7,8] Recently, the *replica exchange with solute tempering* (REST) method was suggested as an efficient alternative to REM in large systems.[9,10] In this method, a judicious choice of temperature-dependent scaling of the Hamiltonian allows one to effectively heat the molecule, or fragment, of interest while the remainder of the system remains "cold". In this way, the number of replicas required depends only on a small subset of the total system degrees of freedom.

REST has already been applied to study protein folding[11,12] and dynamics, both in solution[13] and on a crystal surface.[14] By combining REST with λ-hopping (replica exchange between neighboring λ windows),[15] the consistency of binding free energies was found to improve markedly for two problematic cases, namely, the binding of *p*-xylene and benzene to lysozyme L99A, which requires the correct conformational ensemble of side chain dihedral angles in the protein to achieve FEP results that are independent of the starting conformation, and the sampling of ring flips in ligands bound to thrombin.[10] In a study typical of FEP lead optimization projects, the relative binding affinities of 16 ligands for the CDK2-cyclin A receptor were analyzed.[16] In this case, FEP/REST performed better than standard FEP in rank-ordering the ligands, especially in cases for which multiple binding poses were present.

In this paper, we describe the implementation of the REST procedure in the *MCPRO* software.[17] *MCPRO* is a powerful tool for lead optimization, through FEP calculation with Monte Carlo sampling of protein−ligand binding modes. Notable successes have included the computationally guided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase.[1,18−24] Yet, in cases where the receptor and/or the ligand undergo significant conformational change, the reproducibility of the FEP results may be hindered by inadequate sampling. Here, our aim is to improve the consistency of computed FEP results, while maintaining a light computational workload suitable for high throughput lead optimization procedures. All of the

Received: November 17, 2013
Published: January 17, 2014

   dx.doi.org/10.1021/ct400989x | *J. Chem. Theory Comput.* 2014, 10, 565−571

calculations that follow have been run using just four parallel processes on a single desktop machine. As discussed below, though the REST method significantly enhances conformational sampling, further gains are achieved by incorporating the 'flip' option in *MCPRO*, which invokes periodic attempts at large changes in selected dihedral angles.[25]

We apply the combined REST/flip methodology to the optimization of non-nucleoside inhibitors of HIV-1 reverse transcriptase (NNRTIs), with a particular view to enhancing potency against the Tyr181Cys (Y181C) mutant form. These inhibitors are an important component of treatment against HIV infection, but patients who begin NNRTI therapy often develop the Y181C mutation, thus rendering common drugs inactive. In this context, one of the subjects of this lab's design efforts has been the oxazole **1** (Figure 1). In particular, **1b** (R = H, X = Cl) has an EC_{50} of 6 nM toward wild-type HIV-1 and 420 nM toward the Y181C variant. It was expected that substituents at the 4-R position in **1b** might occupy the space vacated by Tyr 181, hence increasing potency against the mutant. Indeed, FEP calculations predicted that bulky alkyl groups, such as ethyl and isopropyl, would give gains in free energy of binding, and it was confirmed experimentally that both analogs have sub-10 nM potency toward both viral strains.[26] However, the bulky nature of the substituents provides a challenge for standard MC sampling. In particular, dihedral angle rotation in the confined binding pocket requires the crossing of substantial free energy barriers, which may not occur within a reasonable computational time. As such, it provides an excellent test of the use of enhanced sampling methods, in combination with MC/FEP, in a typical medicinal chemistry setting. In what follows, we describe the implementation of REST within *MCPRO*, benchmark the performance of the combined REST/flip method in the isopropyl to ethyl FEP transformation in the wild-type protein, and, finally, present activity predictions of seven analogs of **1b** toward the wild-type and Y181C strains.

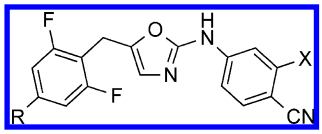


Figure 1. Molecule **1**. For the analogs described here, X = Cl.

2. METHODS

Following the standard REST procedure, the potential energy for a receptor−ligand system is broken down into ligand intramolecular interactions (E_L), the self-interaction energy of the receptor, including water molecules (E_R), and the interaction energy between the ligand and receptor (E_{RL}). A number of replicas m are run at different temperatures T_m, and the three components of the Hamiltonian are scaled differently, according to the temperature

$$E_m(X_m) = E_L(X_m) + \frac{\beta_0}{\beta_m}E_R(X_m) + \sqrt{\frac{\beta_0}{\beta_m}}E_{RL}(X_m) \quad (1)$$

where X_m represents the configuration of the system, $\beta_m = 1/(k_B T_m)$, and T_0 is the temperature of interest (usually room temperature). When $T_m = T_0$, eq 1 reduces to the usual expression for the total potential energy of the system. The temperature of the ligand in each replica is T_m. The scaling factor is such that the Boltzmann factor for the receptor is $\exp(-\beta_0 E_R)$ at all temperatures. Hence, the receptor is kept at the temperature of interest, T_0, while the ligand is heated, allowing it to cross potential energy barriers more rapidly. The scaling factor for E_{RL} is intermediate between those for E_R and E_L and has been shown to prevent the loss of protein secondary structure at high temperatures,[11−13] which was sometimes observed with earlier choices of scaling factor.[9]

The replicas are run in parallel at different temperatures, and, at constant intervals, an exchange of configurations is attempted between neighboring replicas, with the acceptance probability determined by the Metropolis criterion. It can be shown, by imposing detailed balance, that for the particular scaling factors used in the REST method, the exchange probability is independent of the receptor self-interactions (E_R), which explains the relatively small number of replicas that are required to achieve high exchange probabilities when compared with REM.[9] It should be emphasized that the REST method does rigorously sample the correct Boltzmann ensemble, though due to the scaling of the potential energy surface, only at the target temperature T_0.

We have implemented the REST scaling factors (eq 1) in *MCPRO* (a modified version 2.3), while replica exchange is controlled by an external script. The use of Monte Carlo sampling removes the need for velocity rescaling in MD, and hence the replicas can simply be run at different temperatures and replica exchange achieved by swapping the entire system coordinates. To improve exchange acceptance rates, ligand intramolecular bonding interactions are scaled in the same way as E_R − that is, only angle, dihedral, and nonbonded interactions are "heated" as these are the only components likely to contribute to potential energy barriers.

REST has been combined with the standard FEP protocol in *MCPRO*. Namely, the ligand is smoothly mutated between its initial and final states according to a coupling parameter λ. We employ REST to enhance sampling at each λ window separately and sum the free energy differences measured in the ensemble at T_0. We do not use the λ-hopping protocol[15] that is popularly used in other applications of FEP/REST.[10,16] Although such an approach has been shown to further increase solvent configurational sampling and reduce random sampling errors in the calculated free energies, we prefer standard FEP to the λ-hopping approach since there is no need for abundant, identical, synchronized nodes to run all λ windows, and it is, therefore, more suited to moderate computational resources.

MCPRO simulations of the HIV-RT receptor in complex with the benzyloxazole ligands followed the same protocols as described elsewhere.[21,22,24,26] Briefly, the initial coordinates of the complexes were identical to those used in Bollini et al.,[26] being constructed from the 1S9E PDB file[27] using the *MCPRO*[17] and *BOMB*[1] software. The 178 amino acids closest to the ligand were retained, and the bound and unbound structures were solvated in 25 Å caps, comprising 1250 and 2000 water molecules, respectively. In all simulations except the ethyl to methyl transformation, the initial solvent distribution was derived from a stored solvent box, as is typical in MC simulations. In the latter case, the small side chain occasionally allows a water molecule to be placed in the hydrophobic cavity. The JAWS water placement algorithm[28] was run with standard parameters as described elsewhere[29] and confirmed that water placement in the cavity is energetically unfavorable. The solvent distribution derived from JAWS was, hence, subsequently used for the Et to Me simulation. The protein energetics were

described using the OPLS-AA force field, the ligands with OPLS/CM1A, and water with TIP4P.[30] FEP calculations were performed using 11 λ windows of simple overlap sampling,[31] with each window comprising 10 million (M) configurations of equilibration and 30 M (40 M) configurations of averaging for the bound (unbound) simulations. Errors in the computed free energy of binding were estimated by binning the data for each λ window into ten blocks and computing the standard error in the mean.

REST calculations were set up with a view to improving the consistency of results for a standard FEP setup on moderate computational resources. Therefore, for each λ window, four replicas were run in parallel on the same node. The free energy calculations were performed at (T_0 =) 25 °C. The maximum temperature was chosen to allow reasonable temperature diffusion (greater than 20% acceptance rates for exchange) between all replicas and was usually 300 and 250 °C for bound and unbound simulations, respectively. Intermediate temperatures were selected to be approximately exponentially distributed over the temperature range[32]

$$T_i = T_0\left(\frac{T_{N-1}}{T_0}\right)^{i/(N-1)} \quad (2)$$

where N is the number of replicas. In practice, the replica temperatures are occasionally adjusted to ensure that exchange acceptance probabilities exceed 20%. As an example, the average acceptance rates for the *tert*-butyl to isopropyl transformation, which are the largest ligands studied here, are 29% both in the bound and the unbound simulations, using temperature distributions of 25, 90, 190, 315 °C and 25, 85, 160, 260 °C, respectively. Exchange attempts between pairs of neighboring replicas, chosen at random, were attempted after every 10 000 MC configurations. Here, a compromise must be made between ensuring that successive exchange attempts are independent, thus reducing the probability of "back exchange",[32] and improving temperature diffusion by maximizing the number of replica exchanges.

We have found, empirically, that even allowing 10 000 MC configurations between replica exchanges and even at the highest temperatures used here, there is often insufficient time for a dihedral angle transition between energy wells. As such, we have found that the REST method works well with a Monte Carlo dihedral angle 'flip' protocol, in which selected dihedral angles undergo attempted jumps, which are much larger than typical MC moves.[25] The flips are attempted on one out of every six attempted MC moves of the dihedral angle (typically, 1−3 flip attempts are made between replica exchange attempts). As we shall show, the separation between free energy wells in dihedral angle space may differ from the user's intuition. As such, we employ jumps of random size chosen to lie in the range between 60° and 300°.

Where indicated in the text, we have also run standard bound FEP calculations using *MCPRO*. In these cases, the procedure is identical to that of Bollini et al.,[26] except that the simulation is extended to 30 M configurations of averaging for consistency with the REST calculations.

3. RESULTS

3.1. Isopropyl to Ethyl. To demonstrate the potential benefits of the combined REST/flip approach in a typical medicinal chemistry FEP context, we first analyze in detail the energetics of the R = isopropyl to ethyl transformation in the binding pocket of wild-type HIV-RT. In this transformation, either branch of the *i*-Pr group may be mutated to hydrogen to give ethyl, and the free energy of the transformation should, in principle, be independent of the chosen path. In practice, the standard MC/FEP approach, as used in Bollini et al.,[26] can depend quite strongly on which methyl is mutated to hydrogen (labeled paths (1) and (2) in Table 1), with typical differences

Table 1. Differences in Free Energy for the *i*-Pr to Et Transformation in the WT Protein[a]

		run 1	run 2	run 3	run 4	av
standard	path (1)	1.48	0.92	1.33	0.74	**1.12**
	path (2)	2.50	2.45	2.90	2.53	**2.59**
standard/flip	path (1)	1.49	1.90	1.25	1.48	**1.53**
	path (2)	1.56	1.24	2.07	1.61	**1.62**
REST/flip	path (1)	1.38	1.62	1.46	1.32	**1.45**
	path (2)	1.52	1.44	1.59	1.64	**1.55**

[a]Units are kcal/mol. Standard errors on each run are approximately 0.09, 0.14, and 0.12 kcal/mol, for standard, standard/flip, and REST/flip, respectively.

in excess of 1.5 kcal/mol. In addition, to assess the consistency of the computed free energies, we have performed the same FEP calculation four times, each with slightly different starting conditions (labeled runs 1 to 4 in Table 1). The computed free energies are less sensitive to the starting conditions than the mutation pathway, but again, the differences, in excess of 0.5 kcal/mol, are larger than expected based on the computed random error estimates (approximately 0.09 kcal/mol).

As discussed in the Methods, standard FEP calculations in *MCPRO* may be run with the 'flip' option, which attempts large jumps in angle for specified dihedrals (in this case, the angle labeled ϕ in Figure 2). Using random dihedral jumps between 60 and 300°, the consistency of the computed free energies of *i*-Pr to Et mutation in the protein (Table 1) are improved over

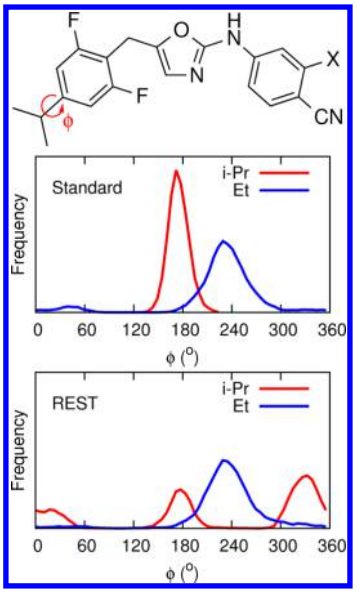


Figure 2. Dihedral angle distributions along the R = *i*-Pr to Et transformation in WT HIV-RT, using standard and REST sampling. Configurations are analyzed from all eight simulations described in Table 1. For isopropyl, REST explores more conformational space than standard sampling.

dx.doi.org/10.1021/ct400989x | J. Chem. Theory Comput. 2014, 10, 565−571

standard FEP. Indeed, the average computed free energy over the four runs is now independent of the mutation pathway. However, there is still an unacceptable variability in the individual runs, with differences as large as 0.8 kcal/mol, indicating that a number of long runs is required to gain confidence in the computed results.

Finally, the same *i*-Pr to Et FEP simulations were run using the REST/flip method in *MCPRO*. Again, the average computed free energy is independent of the chosen mutation pathway and is in very good agreement with standard FEP using flip (Table 1). Importantly, the results are very consistent across all eight FEP simulations with a maximum difference of 0.3 kcal/mol, which is consistent with the estimated random errors. Figure 2 examines the differences between the standard and REST configurations. The key degree of freedom, in this respect, is the dihedral angle labeled ϕ. In the standard simulation, the dihedral angle distribution for both *i*-Pr ($\lambda = 0$) and Et ($\lambda = 1$) shows just one peak, centered at 180 and 240°, respectively. As shown in Figure 3(a,c), this corresponds to *i*-Pr having one methyl group pointing toward Tyr 181 and Et pointing away. In the REST calculations, Et maintains the same geometry, indicating a deep energetic well. However, *i*-Pr shows a much larger distribution of geometries, with a broad peak appearing at $\phi = 330°$, which corresponds to one methyl group of *i*-Pr pointing toward Tyr 188 (Figure 3(b)). Similar analysis may be performed for each intermediate λ-window along the transformation, but the net result of the quasi-ergodic sampling observed in the standard FEP calculations is the inconsistent binding free energy data presented in Table 1.

In the context of the optimization of inhibitors of the Y181C mutant of HIV-RT, the distribution of binding poses shown in Figure 2 is particularly interesting since bulky alkyl groups that are able to occupy the space vacated by Tyr 181 are expected to be an effective means of targeting the variant form. Indeed, Figure 4 shows that both the *i*-Pr and Et analogs favor conformations that orient methyl groups toward Cys 181 in Y181C. For *i*-Pr, the distribution of dihedral angles at $\phi = 180°$ grows at the expense of the peak at $\phi = 330°$. On the other hand, the distribution of dihedrals for Et moves almost entirely to $\phi = 40°$, which corresponds to a conformation pointing toward Cys 181. The energetic consequences of these observations will be discussed in the next section.

Finally in this section, we discuss the advantage of the proposed FEP scheme in which 'flip' is used in combination with REST to enhance sampling at each λ window. Figure 5 shows the number of accepted flip attempts during one λ window of the *i*-Pr to Et transformation in wild-type HIV-RT. The first point to note is that, as expected, the flip acceptance rate is significantly enhanced for the high temperature REST replica, compared with room temperature, due to the scaled intra- and intermolecular interactions. Thus, exchange of replicas between 25 °C and higher temperatures will enhance dihedral angle sampling in the ligand, as we have seen in the *i*-Pr and Et analogs. Second, the flip rate is relatively independent of the attempted dihedral angle flip. Indeed, Figure 2 reveals that bound conformations of the *i*-Pr analog have dihedral angle distributions separated by 160°, rather than the more intuitive 180°.

3.2. NNRTI Optimization. Table 2 lists the relative free energies of binding of seven benzyloxazole analogs with hydrocarbon substituents at the 4-R position (Figure 1) calculated using REST/flip and compares them with the previously reported standard FEP and experimental EC_{50} results.[26] The main conclusions are qualitatively similar to those arrived at using standard FEP. Namely, in the WT protein, replacing R = methyl by an ethyl group should enhance binding, while any further enlargement is detrimental. In the Y181C variant, greater gain in affinity is seen for larger hydrocarbon chains when compared with the WT, as expected, and peak binding is predicted to occur for R = isopropyl.

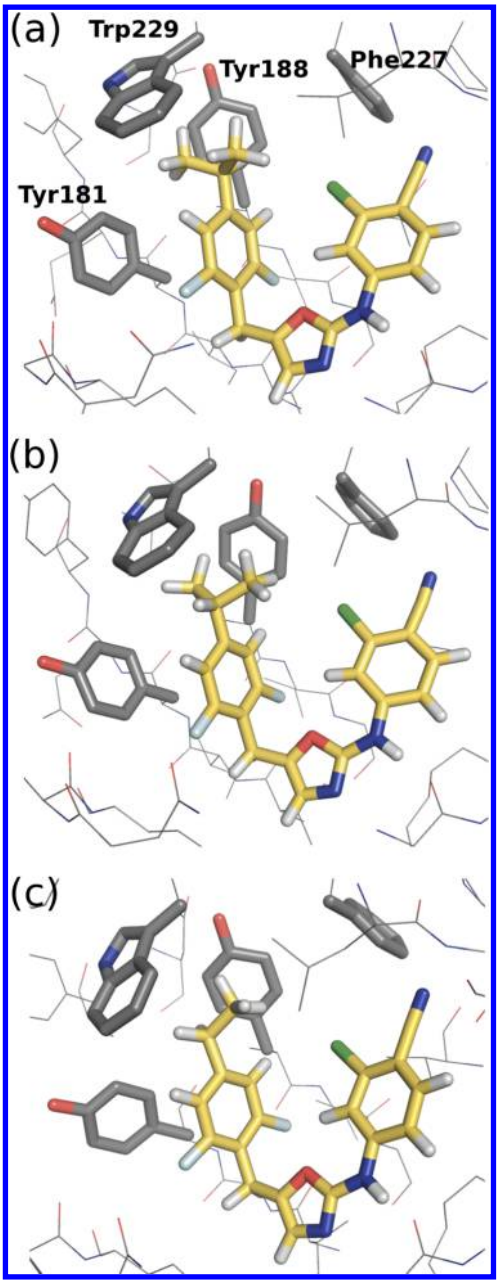


Figure 3. Snapshots from the REST FEP calculations showing (a) R = *i*-Pr ($\phi = 180°$), (b) R = *i*-Pr ($\phi = 330°$), (c) R = Et ($\phi = 240°$). For isopropyl, REST samples two distinct conformations. Ethyl always points away from Tyr 181.

The computed relative free energies of binding can be compared with experiment, with the serious caveat that the latter are obtained in a cell-based assay. The ethyl substituent is confirmed as the most potent in these assays for the WT, while, for Y181C, ethyl and isopropyl are essentially equally potent.[26] Although simulation and experiment agree on the identities of the strongest binders, there are some discrepancies between the

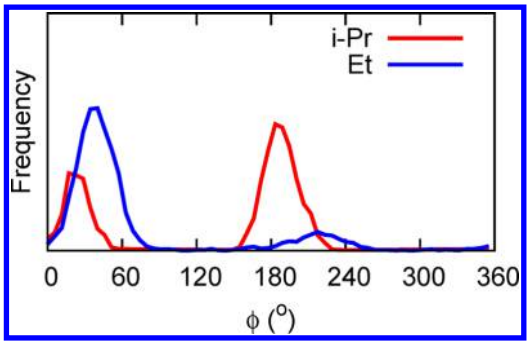


Figure 4. Distribution of dihedral angles in *i*-Pr and Et analogs bound to Y181C from REST simulations. Both inhibitors favor conformations that occupy the space vacated by Tyr 181.

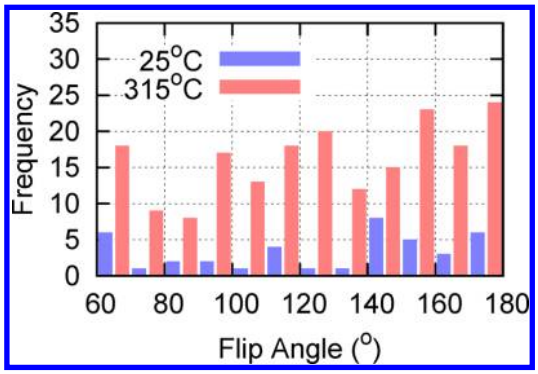


Figure 5. Distribution of accepted flips in one window ($\lambda = 0.2$) of the *i*-Pr to Et transformation in WT HIV-RT. Data is collected from all eight runs in Table 1. Flipping is enhanced in the high temperature REST replica compared to the room temperature replica. Flip successes are not strongly dependent on the flip angle.

Table 2. Computed Relative Free Energies of Binding (kcal/mol) Using Standard FEP and REST[a]

	standard[26]		REST		EC_{50}[26]	
R	WT	Y181C	WT	Y181C	WT	Y181C
Me	0.00	0.00	0.00	0.00	0.011	0.210
Et	−1.60	−2.90	−0.74	−2.28	0.0013	0.0069
Pr	−0.23	**−4.46**	0.69	**−1.60**	–	–
i-Pr	−0.82	−5.45	0.14	−4.40	0.0052	0.0072
OEt	**0.97**	−1.35	**3.15**	0.01	0.028	0.048
CH_2OMe	2.67	0.78	1.72	1.02	0.0036	0.690
t-Bu	0.21	−4.15	0.73	−3.39	1.3[b]	–

[a]Discrepancies larger than 1.5 kcal/mol between the two methods are highlighted in bold. Note that isopropyl REST results are averages over the two mutation pathways and four runs described in Table 1; *tert*-butyl results are averages over three mutation pathways (Table 3). The estimated uncertainties in the REST results range from 0.08 (Et) to 0.22 kcal/mol (OEt). The estimated uncertainties in the standard results are listed in full elsewhere.[26] Also shown for comparison are experimental EC_{50} data (μM).[26] [b]The 4-*tert*-butyl-2,6-dichlorobenzyl analog of **1b**.

two data sets, even using enhanced sampling of the ligand degrees of freedom. In particular, the CH_2OMe analogue is more potent against the WT protein in the cell assay than predicted computationally, while the opposite is true for *i*-Pr against the Y181C variant.

It is instructive to look more closely at the differences between the standard FEP and REST results. In this respect, the greatest discrepancies (more than 1.5 kcal/mol) are in the predicted affinities of Pr for Y181C and OEt for WT. In REST simulations of the WT protein, Pr is oriented exclusively away from Tyr 181, but in the Y181C variant it is observed to explore much more conformational space, including the vacancy left by Tyr 181. This is best illustrated by a 2D histogram showing the distribution of dihedral angles in the Pr chain (Figure 6(a)), which shows extensive sampling in three regions of dihedral space. In contrast, in standard MC sampling, the propyl group is trapped in an energetic well (pointing toward Cys 181), which may contribute to the more favorable binding affinity of Pr in standard FEP.

Figure 6(b) shows the equivalent 2D dihedral angle distributions for the OEt substituent bound to the WT protein. Again, the conformational sampling is much more extensive in the REST simulation, and, in fact, the inhibitor appears to be trapped in a local energy minimum in the standard simulation, which may explain the difference in the computed binding free energies. We have also computed the relative binding free energy of the similar substituent CH_2OMe, which was not included in the original study.[26] REST predicts that CH_2OMe is the more potent inhibitor in the WT protein, but that OEt is more strongly favored in the Y181C variant (Table 2). Interestingly, OEt is oriented exclusively toward Cys 181 in the REST simulations, but CH_2OMe always points away (Figure 7). The relative affinities of OEt and CH_2OMe for the WT and Y181C variants predicted by REST are in good agreement with the experimental EC_{50} results,[26] which underlines the importance of constructively occupying the space vacated by Tyr 181 when targeting the mutant protein.

Finally, Table 3 summarizes the energetic components of the computed *t*-Bu to *i*-Pr substitution in solution, the WT protein and its Y181C variant, which is the most demanding studied here since both end points comprise bulky hydrophobic groups. As in Table 1, the three sets of results correspond to a different methyl to hydrogen mutation pathway, which, in the limit of complete sampling, should give identical results. The standard FEP results in solution and the WT protein are quite consistent. However, in Y181C, the computed free energy difference in the bound state ranges from 0.77 to 3.50 kcal/mol, probably due to incomplete sampling of the possible *i*-Pr conformations shown in Figure 4 (red line). The situation is much improved using our REST/flip implementation, and the Y181C free energies vary over a much smaller range (1.52 to 1.79 kcal/mol). One of the computed free energy differences in the WT protein (3) appears to be higher than expected, but increasing the number of MC configurations from 30 to 60 M reduces this value to 2.16 kcal/mol.

4. CONCLUSIONS

If FEP simulations are to be useful in guiding lead optimization in a typical medicinal chemistry setting, then the aim must be to perform high-throughput FEP calculations on moderate computational resources, with minimal user intervention. Under these conditions, consistency of the obtained results is vital, and the computed activities must be independent of small changes in starting orientations and chosen mutation pathways. A typical example of such a lead design effort is the optimization of NNRTIs for the inhibition of HIV-RT and its variant forms (such as the Y181C protein discussed here). In this case, the shape of the NNRTI binding pocket indicates that the addition of bulky hydrophobic groups to molecule **1** should be energetically favorable, and this hypothesis has been

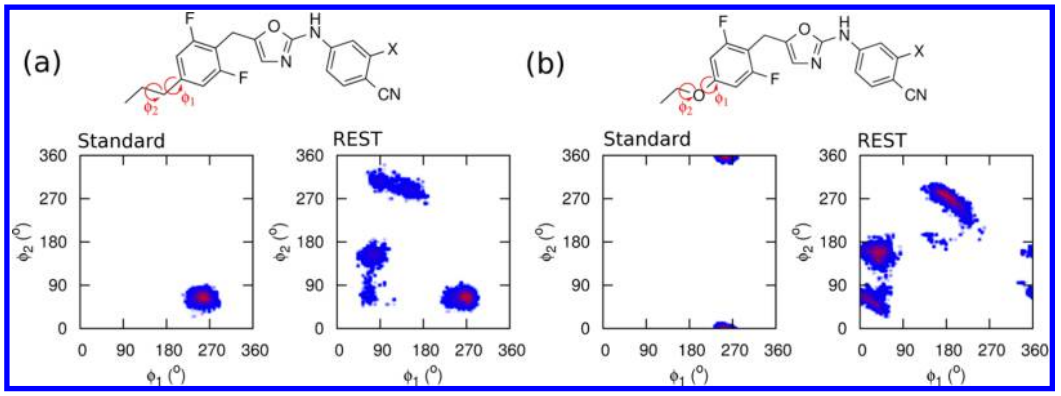


Figure 6. Distribution of (a) R = Pr and (b) R = OEt dihedral angles from typical standard and REST MC simulations.

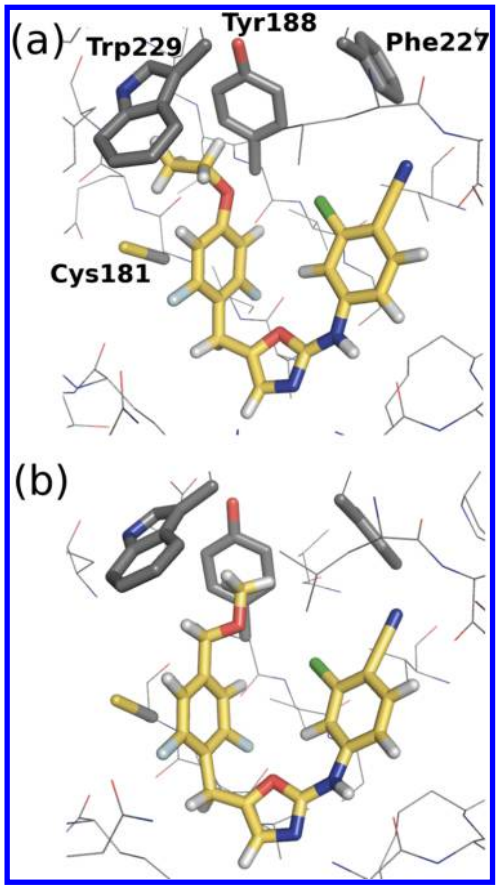


Figure 7. Snapshots from REST simulations of (a) R = OEt and (b) R = CH_2OMe. Computational simulations using REST sampling predict that the OEt group fills the space vacated by Tyr 181, while CH_2OMe is oriented toward Phe 227.

previously confirmed by FEP simulations and experiment.[26] However, the nature of the bulky hydrophobic substituents presents a challenge to standard MC/FEP, since dihedral angle rotation in the binding pocket may be prevented by large free energy barriers, leading to quasi-ergodic sampling and inconsistent binding free energies.

To improve sampling of protein−ligand binding modes in these difficult cases, we have implemented the REST method in *MCPRO*. By allowing exchange of configurations with high temperature replicas of the system, sampling of the possible ligand binding modes is facilitated. We have shown that the consistency of computed free energy changes, that is the independence of the results from the starting conditions and the chosen mutation pathway, is much higher for FEP/REST than standard calculations. The differences between the results obtained by the two techniques are shown to be linked to the sampling of key dihedral angles in the ligand, and we recommend the use of REST in combination with the 'flip' algorithm to facilitate crossing of high free energy barriers in torsional space. Benchmark tests were performed on bulky isopropyl to ethyl and *tert*-butyl to isopropyl substitutions in a confined binding pocket, which are particularly problematic cases for standard MC sampling.

Table 3. Components of the Free Energies of the Transformation of R = *t*-Bu to *i*-Pr in Solution (aq), the WT Protein, and the Y181C Variant[a]

	standard[26]			REST		
	ΔG-aq	ΔG-WT	ΔG-Y181C	ΔG-aq	ΔG-WT	ΔG-Y181C
t-Bu to *i*-Pr(1)	2.71	2.29	0.96	2.60	1.84	1.65
t-Bu to *i*-Pr(2)	3.36	1.78	0.77	2.68	1.94	1.52
t-Bu to *i*-Pr(3)	2.74	1.65	3.50	2.70	2.44[b]	1.79

[a]Three mutation pathways are calculated in each environment. The estimated uncertainty on each value is approximately 0.10 kcal/mol. [b]After increasing the number of averaging configurations to 60M, ΔG-WT is 2.16 kcal/mol.

It should be emphasized that our current implementation in *MCPRO* limits the high temperature region to the ligand and its interactions, which assumes that there are no slow degrees of freedom in the binding pocket itself. This is partly for computational convenience but is also to ensure that simulations may be run on very moderate computational resources. Indeed, the motivation of the current study has been to study how the consistency of computed binding free energies may be improved using the REST method, while retaining an otherwise standard *MCPRO* FEP setup. All of the calculations described here were run on four parallel processors, which is significantly smaller than in typical REM simulations. The protocol may be trivially extended to allow more replicas or modified to be used as part of a λ-hopping FEP scheme if desired.

The computed activities of the seven NNRTIs are qualitatively similar to the original computational study.[26] R = ethyl is favored for the WT protein, while the Y181C variant supports more bulky substituents. However, some differences have been observed, particularly for Pr in Y181C and OEt in the WT protein, where the ligands' bulky hydrophobic groups are observed to be trapped in energetic wells for the entire

dx.doi.org/10.1021/ct400989x | J. Chem. Theory Comput. 2014, 10, 565−571

simulations using standard sampling techniques. A particularly interesting case is the comparison between R = OEt and CH_2OMe in Y181C, where a small change in the substituent results in an entirely different binding mode (Figure 7). The relative affinities of these two substituents, measured in cell-based assays[26] and explained using REST/flip simulations, highlights the importance of a thorough exploration of the ligand's binding modes for computationally guided lead optimization.

AUTHOR INFORMATION

Corresponding Author

*E-mail: william.jorgensen@yale.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Markus Dahlgren for helpful discussions. D.J.C. is supported by a Marie Curie International Outgoing Fellowship within the seventh European Community Framework Programme. The authors are grateful for financial support provided by the National Institutes of Health (GM32136 and AI44616).

REFERENCES

(1) Jorgensen, W. L. *Acc. Chem. Res.* **2009**, *42*, 724−733.
(2) Zuckerman, D. M. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.
(3) Barril, X.; Javier Luque, F. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 81−86.
(4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(5) Paschek, D.; Garcia, A. E. *Phys. Rev. Lett.* **2004**, *93*, 238105.
(6) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.
(7) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2008**, *128*, 144121.
(8) Itoh, S. G.; Okumura, H.; Okamoto, Y. *J. Chem. Phys.* **2010**, *132*, 134105.
(9) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749−13754.
(10) Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1937−1942.
(11) Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(12) Terakawa, T.; Kameda, T.; Takada, S. *J. Comput. Chem.* **2011**, *32*, 1228−1234.
(13) Moors, S. L. C.; Michielssens, S.; Ceulemans, A. *J. Chem. Theory Comput.* **2011**, *7*, 231−237.
(14) Wright, L. B.; Walsh, T. R. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4715−4726.
(15) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703−13710.
(16) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. *J. Chem. Theory Comput.* **2013**, *9*, 1282−1293.
(17) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689−1700.
(18) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. *Bioorg. Med. Chem.* **2006**, *16*, 663−667.
(19) Ruiz-Caro, J.; Basavapathruni, A.; Kim, J. T.; Bailey, C. M.; Wang, L.; Anderson, K. S.; Hamilton, A. D.; Jorgensen, W. L. *Bioorg. Med. Chem.* **2006**, *16*, 668−671.
(20) Thakur, V. V.; Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. *Bioorg. Med. Chem.* **2006**, *16*, 5664−5667.
(21) 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*, 9492−9499.
(22) Leung, C. S.; Zeevaart, J. G.; Domaoal, R. A.; Bollini, M.; Thakur, V. V.; Spasov, K.; Anderson, K. S.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2485−2488.
(23) Bollini, M.; Domaoal, R. A.; Thakur, V. V.; Gallardo-Macias, R.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L. *J. Med. Chem.* **2011**, *54*, 8582−8591.
(24) Jorgensen, W. L.; Bollini, M.; Thakur, V. V.; Domaoal, R. A.; Spasov, K. A.; Anderson, K. S. *J. Am. Chem. Soc.* **2011**, *133*, 15686−15696.
(25) Thomas, L. L.; Christakis, T. J.; Jorgensen, W. L. *J. Phys. Chem. B* **2006**, *110*, 21198−21204.
(26) Bollini, M.; Gallardo-Macias, R.; Spasov, K. A.; Tirado-Rives, J.; Anderson, K. S.; Jorgensen, W. L. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 1110−1113.
(27) Das, K.; Clark, A. D.; Lewi, P. J.; Heeres, J.; de Jonge, M. R.; Koymans, L. M. H.; Vinkers, H. M.; Daeyaert, F.; Ludovici, D. W.; Kukla, M. J.; De Corte, B.; Kavash, R. W.; Ho, C. Y.; Ye, H.; Lichtenstein, M. A.; Andries, K.; Pauwels, R.; de Béthune, M.-P.; Boyer, P. L.; Clark, P.; Hughes, S. H.; Janssen, P. A. J.; Arnold, E. *J. Med. Chem.* **2004**, *47*, 2550−2560.
(28) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Phys. B* **2009**, *113*, 13337−13346.
(29) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2010**, *6*, 3850−3856.
(30) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665−6670.
(31) Jorgensen, W. L.; Thomas, L. L. *J. Chem. Theory Comput.* **2008**, *4*, 869−876.
(32) Abraham, M. J.; Gready, J. E. *J. Chem. Theory Comput.* **2008**, *4*, 1119−1128.