JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Is Ring Breaking Feasible in Relative Binding Free Energy Calculations?

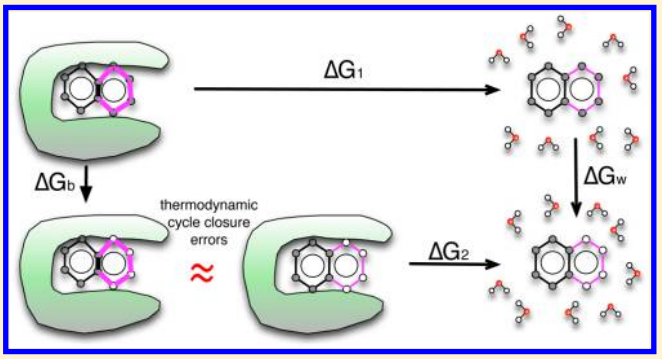
Shuai Liu,[†] Lingle Wang,[‡] and David L. Mobley*,[†]

[†]Department of Pharmaceutical Sciences and Department of Chemistry, University of California, Irvine, 147 Bison Modular, Irvine, California 92697, United States

[‡]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States

Supporting Information

ABSTRACT: Our interest is relative binding free energy (RBFE) calculations based on molecular simulations. These are promising tools for lead optimization in drug discovery, computing changes in binding free energy due to modifications of a lead compound. However, in the "alchemical" framework for RBFE calculations, some types of mutations have the potential to introduce error into the computed binding free energies. Here we explore the magnitude of this error in several different model binding calculations. We find that some of the calculations involving ring breaking have significant errors, and these errors are especially large in bridged ring systems. Since the error is a function of ligand strain, which is unpredictable in advance, we believe that ring breaking should be avoided when possible.



1. INTRODUCTION

Here our interest is in drug lead optimization, where a compound that binds the desired target is known and we seek to create derivatives of this lead compound that either improve the affinity or maintain the affinity while improving other properties. Relative binding free energy (RBFE) calculations based on molecular dynamics (MD) simulations can be used to predict binding free energy differences based on chemical changes in advance of synthesis of the derivative compounds. Thus, they can potentially substantially accelerate the lead optimization process[1] and are of considerable interest for drug discovery applications.

Given a particular model (force field and parameters), alchemical RBFE calculations yield correct relative binding free energies in principle, at least in the limit of adequate sampling, as reviewed elsewhere.[2−4] However, large chemical modifications require substantially more sampling, and hence, with a fixed amount of sampling, increasing the size of the transformation can increase the magnitude of the errors due to sampling. Thus, in order to ensure that typical modifications are relatively small, we recently designed a program, lead optimization mapper (LOMAP),[5] for planning efficient RBFE calculations.

LOMAP automatically selects single-topology RBFE calculations spanning a lead series by pairing similar molecules. In LOMAP, we calculate only the RBFE between molecules that have sufficient similarity. Structural similarities are computed on the basis of a similarity score, which relates to the change in the number of atoms during the transformation between the two molecules in question. Specifically, we identify the maximum common substructure shared by two molecules and determine the change in the number of atoms needed to reach this substructure; we use these changes as the basis for our similarity score. Currently, LOMAP uses two scoring schemes, called "strict" and "loose", which differ only in how we treat transformations that would break polycyclic ring systems[6] (Figure 1).

In the strict scoring scheme, we do not allow any ring breaking when we search for the maximum common

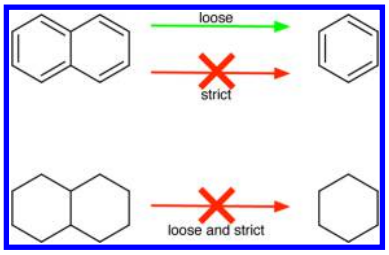


Figure 1. In the strict approach, mutations involving breaking of any ring are not allowed (red arrows). Here, "ring breaking" refers to any transformation in which atoms remaining in the final system were part of any ring that has been removed. In the loose approach, ring-breaking mutations are allowed only when the ring system left behind is rigid (green arrows).

Received: February 2, 2015
Published: April 2, 2015

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

substructure. For example, in the strict approach, if we consider the pair naphthalene and benzene (Figure 1), these have no common substructure because a naphthalene to benzene transformation would involve breaking a ring. On the other hand, in the loose scoring scheme, we allow ring breaking to happen only when the ring system left behind is relatively rigid or planar (typically aromatic). For example, in this case mutation from naphthalene to benzene is allowed because the remaining ring, benzene, is rigid, while the mutation from decalin to cyclohexane (Figure 1) is not allowed because the remaining ring, cyclohexane, is not rigid (indeed, it can undergo significant conformational transitions). LOMAP was designed to use the loose scoring scheme only when absolutely necessary in order to produce RBFE calculations spanning a lead series—for example, if a group of bicyclic molecules could not be connected to single-ring systems via any other means—but it avoids these types of transformations whenever possible. This was done because the effective "deletion" of partial rings in polycyclic (bicyclic, in this example) ring systems can introduce error into the associated thermodynamic cycles (Figure 2).

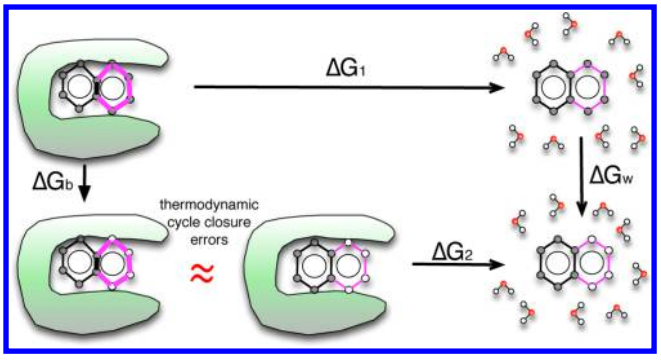


Figure 2. Here we consider the thermodynamic cycle used for single-topology relative binding free energy calculations, where we compare the binding of two ligands 1 and 2 (top and bottom, respectively, where our example shows cartoons of naphthalene and benzene). In this approach, we obtain the difference between the binding free energy of 1, ΔG_1, and the binding free energy of 2, ΔG_2, by perturbing the first ligand into the second in the binding site (ΔG_b) and in water (ΔG_w). These perturbations can involve changing some of the atoms in one or both ligands into *dummy atoms* (shown as hollow spheres), defined as atoms that no longer have charge or Lennard-Jones interactions with the remainder of the system but retain their bonded interactions. In a correct thermodynamic cycle $\Delta G_1 + \Delta G_w - \Delta G_2 - \Delta G_b = 0$. However, this is only strictly true when the free energies of the dummy atoms in the binding site and in solution are identical. This criterion is met when the dummy atoms are a set of masses and springs with a conformation that is not affected by that of the remaining "real" atoms. However, if the conformation of the dummy atoms is coupled to the conformation of the remaining portion of the molecule (i.e., when they are members of a ring), these two free energies may in fact differ, and the thermodynamic cycle will not close. Here, at the bottom left, the binding site is shown to introduce strain into the dummy atom ring (bold bonds), so that the free energy of the left bottom and middle bottom states are no longer equal, causing a cycle closure error (red "approximately equal" sign).

Effectively, the loose scoring scheme means that the thermodynamic cycle on which our RBFE calculation is based in these cases no longer formally closes; we potentially have a missing contribution due to a conformational change in the remaining dummy atoms induced by changes in the conformation of the connected interacting atoms (shown by the red "approximately equal" sign in Figure 2).

More rigorously, we show in the Appendix that the dummy atoms can have a nonzero effect on the free energy change whenever there is more than one bond-stretching interaction between the same group of dummy atoms and the remaining interacting atoms (i.e., more than one connection point between the dummy atoms and the rest of the molecule).

In our previous work,[5] we assumed that any contribution to the free energy change from these dummy atoms would be small in the case where the remaining atoms are in a rigid ring system and larger when these atoms are in a flexible ring system. Thus, we assumed that when it was necessary in order to ensure that all of the compounds in a lead series could be connected, we could break rings in rigid molecules and still introduce only a minimum amount of error in computed binding free energies. However, this was an assumption—the exact magnitude of these contributions is not known. The existence of these errors was well-known, but understanding their magnitude is now essential. This has dramatic implications for how we plan free energy calculations. Specifically, can we allow these types of transformations in special cases? Or do they need to be avoided or implemented via another route such as absolute free energy calculations? Here we aim to answer these questions. In our initial implementation of LOMAP, we assumed that mutations beyond the loose scoring scheme, such as mutation from decalin to cyclohexane, will accumulate substantial errors and in general be unreliable.

Here we compute the error introduced into single-topology RBFE calculations via both the loose and strict scoring schemes in several model binding calculations as a function of the amount of strain in the remaining atoms in the ligand. These errors are limited to the single-topology approach for RBFE calculations, whereas other methods such as the dual-topology approach[7−10] and the separated-topology approach[11] should not suffer from these errors because they handle these transformations without introducing multiply connected dummy atoms that interact with the remaining atoms. However, since the single topology approach is the default method for LOMAP and is widely used in RBFE calculations, these errors are still important.

2. METHOD

As discussed in the Introduction, we test RBFE calculations in two bicyclic systems: the transformation of naphthalene to benzene (Figure 3) and the transformation of decalin to cyclohexane (Figure 4). We also test another case involving a bridged or cagelike ring system: the transformation of adamantane to bicyclo[3.3.1]nonane (Figure 5). Here, to avoid the complexity and potential sampling problems introduced by doing these calculations in a receptor binding site, we model "binding" as the transfer of the ligand from water to a "binding site" consisting of the ligand in the gas phase with conformational restraints that introduce strain.

This is sufficient for our purposes here, as we are solely interested in how introduced strain affects the formal accuracy of RBFE calculations—that is, we seek to determine how much apparent cycle closure error is introduced by ligand strain. Thus, our approach is sufficient, since the error is a function only of the difference in free energy of the dummy atoms when the ligand is under strain (Figure 6). Since the dummy atoms do not interact with the rest of the system, the error is

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

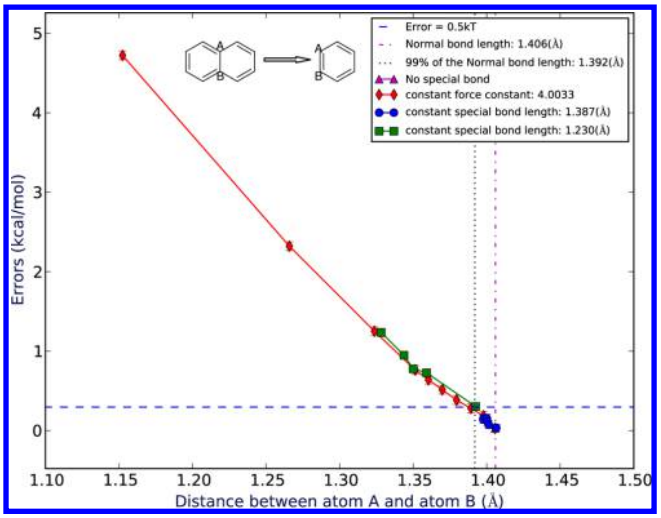


Figure 3. Errors for mutating naphthalene to benzene as a function of the average distance between the shared atoms observed in the simulations. Spots in different colors represent different force constants/bond length groups of the special bond.

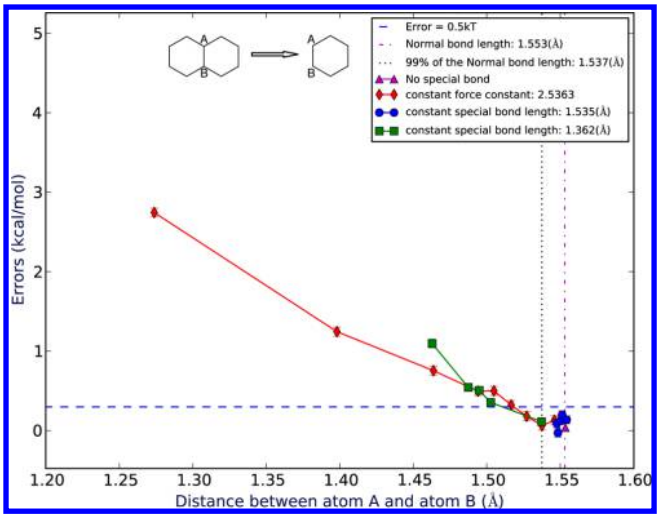


Figure 4. Errors for mutating decalin to cyclohexane as a function of the average distance between the shared atoms observed in the simulations. Spots in different colors represent different force constants/bond length groups of the special bond.

independent of the environment and dependent only on the degree of ligand distortion or strain.

In order to know how much error is introduced by the approximation in question, we need a way to determine the correct "binding" free energy of the molecules in question. We achieve this by computing the absolute binding free energies for every compound considered (Figure 6). Here we compute the free energy to move each molecule from water to the "binding site"—the absolute binding free energy—and then subtract these to obtain the relative binding free energy $\Delta\Delta G_{ab}$. Since absolute binding free energy calculations do not involve any ring breaking, they are correct and provide the gold standard to which we can compare our relative binding free energy results.

Second, we calculate the relative binding free energy using a typical thermodynamic cycle in which we calculate the free energy change by transforming molecule 1 into molecule 2 both in water and in the binding site (Figure 6). From this, we obtain the relative binding free energy as $\Delta\Delta G_{rl} = \Delta G_w - \Delta G_b$.

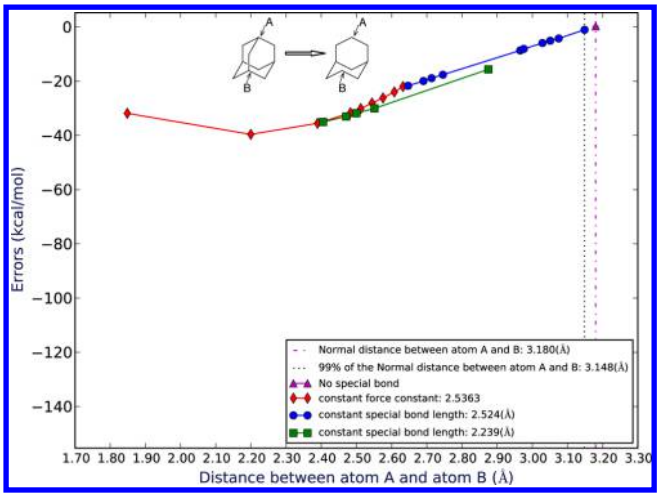


Figure 5. Errors for mutating adamantane to bicyclo[3.3.1]nonane as a function of the average distance between the shared atoms observed in the simulations. This error is for the adamantane−bicyclononane system without NDE. The "original" distance is the distance between atom A and atom B in the absolute free energy calculation of bicyclo[3.3.1]nonane without any special bond restraints. Spots in different colors represent different force constants/bond length groups of the special bond.

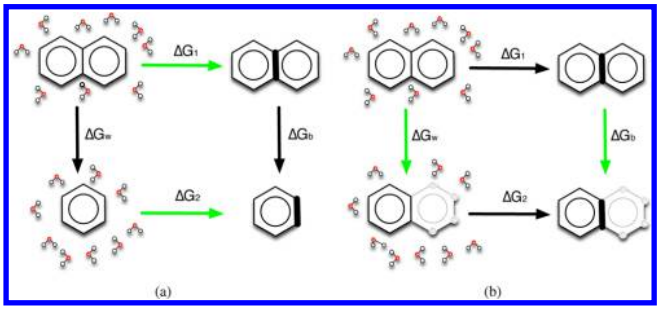


Figure 6. Calculations we run for naphthalene−benzene are shown by the green arrows. (a) We calculate the absolute "binding" free energy for each molecule in going from water to the hypothetical binding site (green arrows) and then get the difference in absolute binding free energies as $\Delta\Delta G_{ab} = \Delta G_1 - \Delta G_2$. Since in this case $\Delta G_1 + \Delta G_b - \Delta G_2 - \Delta G_w = 0$, it follows that $\Delta\Delta G_{ab} = \Delta G_w - \Delta G_b$. (b) We also calculate the relative "binding" free energy (green arrows) by changing atoms to dummy atoms (shown as hollow spheres) and get the relative binding free energy difference as $\Delta\Delta G_{rl} = \Delta G_w - \Delta G_b$. In both panels, the bold bond on the right-hand side represents the bond strained by "binding" to our hypothetical receptor (here represented by introducing restraints).

Since this process includes ring breaking, $\Delta\Delta G_{rl}$ is our target result.

After doing both of these calculations, we compute the overall error as the difference between the reference result (from absolute calculations) and the target result. This measures the difference between the correct free energy change (as determined by absolute calculations) and the free energy change calculated by the relative calculations, which are in error to some degree because the free energy associated with conformational changes in the dummy atoms is neglected.

Here we are interested in determining how the error changes as a function of ligand strain. Particularly, in the limit of no conformational change within the ligand, the conformations and free energies of the dummy atoms in the binding site and in solution will be identical, and no error will be introduced by

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

deleting part of the ring.[12] However, in the limit of very large strain, the bonds between the dummy atoms will be strained in the binding site but not in solution, introducing substantial error. To examine these effects, we added an additional bond between two atoms (bond type 6 in GROMACS), a spring, which is shown in bold in Figure 6. By changing the length and force constant for this additional bond, we can then control the bond length of the shared bond in the bicyclic systems or the distance between the two end atoms in the cagelike system, which controls the strain in the bonds between the remaining dummy atoms. The bond length details are shown in Tables 1, 2, and 3.

Table 1. Errors as a Function of Bond Length (Strain) for Naphthalene−Benzene (The Corresponding Error Figure for This Table is Figure 3)

ID	special bond length (Å)	force constant (kJ mol^{-1} nm^{-2})	bond length in simulation (Å)	error (kcal/mol)
original	none	none	1.4057(8)	0.03 ± 0.04
0	1.387	4.0033	1.3979(6)	0.19 ± 0.06
1	1.370	4.0033	1.3893(7)	0.28 ± 0.06
2	1.350	4.0033	1.3795(7)	0.39 ± 0.05
3	1.330	4.0033	1.3698(7)	0.51 ± 0.05
4	1.310	4.0033	1.3604(6)	0.64 ± 0.05
5	1.290	4.0033	1.3514(8)	0.76 ± 0.05
6	1.230	4.0033	1.3237(6)	1.25 ± 0.05
7	1.110	4.0033	1.2659(6)	2.32 ± 0.05
8	0.870	4.0033	1.1524(5)	4.72 ± 0.05
11	1.387	3.6030	1.3977(6)	0.14 ± 0.05
12	1.387	2.4020	1.4001(8)	0.16 ± 0.05
13	1.387	2.0017	1.4006(6)	0.11 ± 0.05
14	1.387	1.6013	1.4014(6)	0.07 ± 0.05
15	1.387	0.4033	1.4065(7)	0.04 ± 0.04
21	1.230	3.6030	1.3283(6)	1.24 ± 0.05
22	1.230	2.4020	1.3438(6)	0.95 ± 0.05
23	1.230	2.0017	1.3500(6)	0.78 ± 0.05
24	1.230	1.6013	1.3591(6)	0.73 ± 0.04
25	1.230	0.4033	1.3925(8)	0.30 ± 0.04

For each bond length we calculate the cycle closure error as described above. We find that substantial errors are introduced when the change in the bond length (strain) becomes sufficiently large. To provide perspective in terms of how much strain typically is introduced upon ligand binding, we examined simulations of several different protein−ligand systems and measured how much bond length change is typical upon ligand binding.

For all the simulations, we used GROMACS 4.6.[13] The initial structure files were generated using MarvinSketch 5.11.3 and then converted to mol2 files using the OpenEye OEChem toolkits.[14] The OpenEye OEChem Python toolkit and Omega[15] were used to generate three-dimensional conformations and assign AM1-BCC[16,17] partial charges. Antechamber[18] from AmberTools 13 was used to assign GAFF atom types,[19] and then AmberTools' tleap was used to generate the Amber prmtop and crd files, which were converted to GROMACS format using acpype.[20] Small molecules were then set up in GROMACS and, for the solute-in-water case, solvated in TIP3P water[21] in a dodecahedral simulation box with at least 1.2 nm from the solute to the nearest box edge. The numbers of water molecules were 690 for benzene, 554 for napthalene,

Table 2. Errors as a Function of Bond Length (Strain) for Decalin−Cyclohexane (The Corresponding Error Figure for This Table is Figure 4)

ID	special bond length (Å)	force constant (kJ mol^{-1} nm^{-2})	bond length in simulation (Å)	error (kcal/mol)
original	none	none	1.553(1)	0.04 ± 0.05
0	1.535	2.5363	1.5460(7)	0.13 ± 0.06
1	1.517	2.5363	1.5375(8)	0.06 ± 0.06
2	1.494	2.5363	1.5271(8)	0.18 ± 0.06
3	1.472	2.5363	1.5166(9)	0.32 ± 0.06
4	1.449	2.5363	1.5049(7)	0.50 ± 0.06
5	1.428	2.5363	1.4941(7)	0.50 ± 0.06
6	1.362	2.5363	1.4637(7)	0.75 ± 0.06
7	1.228	2.5363	1.3981(7)	1.24 ± 0.06
8	0.964	2.5363	1.2738(7)	2.74 ± 0.05
11	1.535	2.2827	1.5476(7)	0.09 ± 0.05
12	1.535	1.5218	1.5484(8)	−0.03 ± 0.05
13	1.535	1.2682	1.5503(8)	0.12 ± 0.05
14	1.535	1.0145	1.5512(8)	0.20 ± 0.05
15	1.535	0.2563	1.5544(9)	0.14 ± 0.05
21	1.362	2.2827	1.4628(7)	1.10 ± 0.05
22	1.362	1.5218	1.4874(8)	0.54 ± 0.05
23	1.362	1.2682	1.4949(8)	0.51 ± 0.05
24	1.362	1.0145	1.5027(9)	0.35 ± 0.05
25	1.362	0.2563	1.5373(9)	0.11 ± 0.05

Table 3. Errors as a Function of Bond Length (Strain) for Adamantane−Bicyclononane, Using Distances Observed in the Absolute Simulation of Bicyclo[3.3.1]Nonane To Measure the Bond Length (The Corresponding Error Figure for This Table is Figure 5)

ID	special bond length (Å)	force constant (kJ mol^{-1} nm^{-2})	distance between atoms A and B in simulation (Å)	error (kcal/mol)
original	none	none	3.18(7)	0.29 ± 0.03
0	2.524	2.5363	2.63(3)	−21.98 ± 0.10
1	2.494	2.5363	2.61(3)	−23.97 ± 0.10
2	2.456	2.5363	2.58(3)	−26.16 ± 0.10
3	2.421	2.5363	2.54(3)	−28.07 ± 0.10
4	2.383	2.5363	2.51(3)	−30.11 ± 0.10
5	2.347	2.5363	2.48(3)	−31.65 ± 0.10
6	2.239	2.5363	2.39(3)	−35.54 ± 0.10
7	2.019	2.5363	2.20(3)	−39.61 ± 0.10
8	1.585	2.5363	1.85(3)	−31.85 ± 0.10
11	2.524	2.2827	2.65(3)	−21.69 ± 0.08
12	2.524	1.5218	2.69(4)	−19.93 ± 0.05
13	2.524	1.2682	2.71(4)	−18.88 ± 0.05
14	2.524	1.0145	2.75(4)	−17.61 ± 0.04
15	2.524	0.2536	2.97(6)	−8.65 ± 0.03
16	2.524	0.2283	2.98(6)	−8.09 ± 0.03
17	2.524	0.1522	3.03(6)	−5.91 ± 0.03
18	2.524	0.1268	3.05(6)	−5.06 ± 0.03
19	2.524	0.1015	3.08(6)	−4.23 ± 0.03
20	2.524	0.0254	3.15(8)	−1.02 ± 0.03
21	2.239	2.2827	2.41(3)	−35.04 ± 0.08
22	2.239	1.5218	2.47(4)	−33.02 ± 0.05
23	2.239	1.2682	2.50(4)	−31.82 ± 0.05
24	2.239	1.0145	2.55(4)	−30.03 ± 0.04
25	2.239	0.2536	2.88(6)	−15.63 ± 0.03

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

678 for cyclohexane, 552 for decalin, 553 for adamantane, and 597 for bicyclo[3.3.1]nonane.

AMBER combination rules (arithmetic average for σ and geometric average for ε) were used. Simulations were run using Langevin dynamics, as previously,[22] and the simulation time step was 1 fs. Lennard-Jones (LJ) interactions were gradually switched off between 0.9 and 1.0 nm, and analytical corrections were applied to the energy and pressure. Particle mesh Ewald was used for electrostatics, as previously,[22] with a real-space cutoff of 1.2 nm. LINCS was used to constrain bonds to hydrogen. Each system and λ value (where λ is a parameter ranging between 0 and 1, with 0 corresponding to the unmodified system and 1 to the end state of the transformation) was independently minimized for up to 2500 steps of steepest-descent minimization.

Following constant-pressure equilibration, box sizes were adjusted at each λ value by an affine transformation to ensure that each λ value had the correct volume for the target pressure. After this, we conducted an additional 5 ns of constant-pressure production simulation at each λ, discarding the first 100 ps as additional "equilibration", as previously.[22] Here we used Parrinello–Rahman barostat to modulate the pressure.

The parameter λ controls the transformation between end states. In this version of GROMACS, we use three separate λ values, one controlling modification of partial charges (λ_{chg}, turning solute partial charges to zero), the second controlling modification of the bond inducing strain (λ_{bd}, introducing this bond), and the third controlling modification of the LJ interactions (λ_{LJ}, turning solute LJ interactions to zero). The details of the λ spacing can be found in the Supporting Information.

In the case of the bridged ring system, we have to deal with an additional complexity. Because of the absence of a bridging atom in bicyclo[3.3.1]nonane, the internal nonbonded interactions involving atom A and atom B shown in Figure 5 are different in adamantane and bicyclo[3.3.1]nonane—that is, the interactions differ not just in terms of strength but also in terms of which atoms interact. This is the case because the bridging atom changes which interactions are excluded and which are 1–4 interactions. Thus, the end state of the simulation that starts from adamantane (ΔG_w in Figure 7) has different nonbonded interactions from the starting state of the simulation beginning with bicyclo[3.3.1]nonane (ΔG_2 in Figure 7). Unless they are taken into account, these differences in nonbonded interactions will make the thermodynamic cycle fail to close even the absence of strain/conformational change, since a contribution due to the free energy of changing the internal nonbonded interactions would be neglected. We call the error introduced by this change in internal nonbonded interactions "nonbonded discrepancy error" (NDE). The NDE is not a subject of interest here and also is not a necessary feature of binding free energy calculations—that is, if our simulation package allowed us to change the exclusions and pairs lists with λ so as to remove the effects of the presence (or absence) of the bridging atom on 1–4 and excluded interactions, then we could compute relative binding free energies that are unaffected by NDE. Thus, we are interested in understanding errors *aside* from NDE. Therefore, to avoid NDE in GROMACS, we modify the pairs and exclusions sections in our topology files to create a new reference molecule that has the same internal exclusion and 1–4 interactions as adamantane but the same atoms as bicyclo[3.3.1]nonane. This allows us to maintain the same

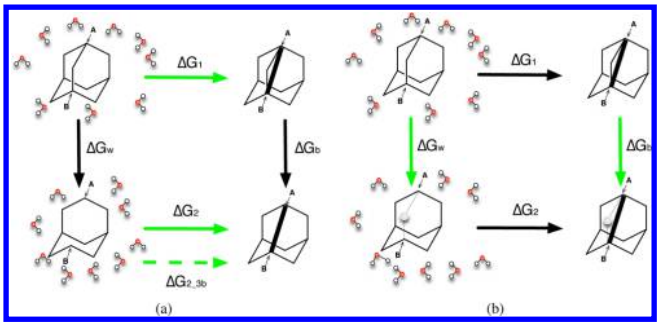


Figure 7. Calculations we run for adamantane–bicyclononane and adamantane–bicyclononane with NDE are shown by the green arrows labeled as ΔG_1, ΔG_b, ΔG_w with the green arrow labeled as ΔG_2 for adamantane–bicyclononane and the dashed green arrow labeled as ΔG_{2_3b} for adamantane–bicyclononane with NDE. The meaning of each simulation is the same as in Figure 6, with panel (a) showing the absolute "binding" free energy for each molecule and panel (b) showing the relative "binding" free energy. For adamantane–bicyclononane specifically, the bottom absolute simulation ΔG_2 starts from the topology based on the structure of bicyclo[3.3.1]nonane plus topology modifications to maintain the same 1–4 and internal interactions with the end state of the left relative simulation ΔG_w. For adamantane–bicyclononane with NDE, the bottom absolute simulation ΔG_{2_3b} starts from the original topology generated from the structure of bicyclo[3.3.1]nonane.

exclusion and 1–4 interactions while transforming a molecule that is like bicyclo[3.3.1]nonane into adamantane. With these adjustments, the simulation ΔG_2 has the same 1–4 interactions and exclusions as the simulation ΔG_w; we call this case "adamantane–bicyclononane". As a comparison, we still run simulations without any adjustments to the topology file (ΔG_{2_3b}). This case, which *does* include NDE, is called "adamantane–bicyclononane with NDE". Adamantane–bicyclononane with NDE is analyzed using the same simulations as adamantane–bicyclononane (we use the same trajectory file and modify the topology file in order to evaluate the desired free energy using different interactions), as discussed in Figure 7 (left bottom green/dashed green arrows in Figure 7).

For all of our systems, we use three different sets of special bonds to induce strain in the ligand. We vary their bond lengths and force constants as follows: (1) varying only the length (keeping the same force constant as in the original bond, or as in a normal carbon–carbon single bond in the case of the cagelike system, which does not initially have a bond between the shared atoms); (2) varying only the force constant (keeping the distance between the shared atoms the same as the original distance; in this case, the initial force constant is that of the original bond or a normal carbon–carbon single bond); (3) varying only the force constant but starting with a reduced distance of 88.7% of the original distance between the shared atoms (this is otherwise the same as case 2). With these combinations, we vary the distance between the shared atoms in the simulations over a wide range. We have eight, five, and five simulations for sets 1, 2, and 3, respectively, for the planar ring systems. In the adamantane–bicyclononane case,[23] we add an additional five simulations for set 2 to obtain better coverage of the space of the bond length.

3. RESULTS

Here we examine how ligand strain in our model "binding" system impacts the error in relative binding free energy calculations. The strain is controlled by an artificial bond that

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727–735

changes a bond length or distance within the ligand as it binds. In our bicyclic systems, we find that as this shared bond deviates from its original value (i.e., the ligand becomes more strained on binding), the error in the computed binding free energy increases. For the bicyclic systems (Figures 3 and 4), in the region close to the normal (unrestrained) bond length, the errors for both systems are relatively small and essentially statistically indistinguishable from zero (smaller than $0.5kT$).

On the other hand, for the cagelike system (the bridged ring case), the errors are bigger. Unlike the bicyclic ring systems, here there is no bond between atom A and atom B in bicyclo[3.3.1]nonane (Figure 5), making the distance between atom A and atom B differ substantially from that in adamantane. That is, the distances between atom A and atom B in the simulations of the absolute free energy calculation of bicyclo[3.3.1]nonane (ΔG_2 in Figure 7) and the relative free energy calculation in vacuum (ΔG_b in Figure 7) are significantly different.

This is not necessarily a problem—it just means that if we want to examine the error as a function of the distance between atom A and atom B, we have two different distances we can use, one which is substantially longer than the other. Thus, we plot the errors versus the distance between atom A and atom B in the simulations based on both of the references: (1) absolute simulations starting from bicyclo[3.3.1]nonane and (2) relative vacuum simulations starting from adamantane. We find that when we use the bond length in the relative calculations from adamantane in vacuum as the "original" bond length, the error is ∼30 kcal/mol when the bond length is 99% of its original value (figures showing this result are presented in the Supporting Information), while using the bond length seen in the absolute calculations from bicyclo[3.3.1]nonane as the "original" bond length gives an error of ∼1 kcal/mol (Figure 5). For both of these simulations, when the bond length changes by 1%, the errors are significant—larger than $1kT$. For adamantane−bicyclononane with NDE, with numbers that include NDE, compared with adamantane−bicyclononane, the errors are similar when the changes in the bond length are small and larger when the changes in the bond length are large (see the Supporting Information). This was expected because adamantane−bicyclononane with NDE includes additional errors beyond those in adamantane−bicyclononane: in addition to contributions due to changes in strain/bonded energies, it also includes NDE errors (Figure 7).

Thus, we find that for the bicyclic systems, large bond length changes do lead to significant errors while small bond length changes (less than 2%) do not result in significant errors in relative binding free energy calculations. For cagelike systems, however, even very small changes in internal distance (1%) can lead to very substantial errors (Figure 5). Here we assess significance on the basis of the point at which the absolute error in the computed relative binding free energy becomes larger than the statistical uncertainty in our calculations.

For the bicyclic systems, we also examined the amount of strain induced by these bond perturbations in order to provide scale. We calculated the average energy difference between the most strained conformation at which the error is still statistically indistinguishable from zero (the "maximum indistinguishable" or "MI" case) and the original, unstrained case ("original"), both for relative calculations in vacuum. For the naphthalene to benzene system (Table 1), case MI is labeled with ID 1. The average potential energy difference between these two cases is 0.31 kcal/mol. For the decalin to cyclohexane system (Table 2), case MI is labeled with ID 2. The average potential energy difference between these two cases is 1.10 kcal/mol.

As noted, we originally expected that for the bicyclic system the flexibility or rigidity of the remaining ring system would have a substantial impact on the magnitude of the error, with rigid rings having substantially smaller errors than flexible rings. However, this is not what we find here—the two approaches seem to have roughly comparable errors. However, we do find that for the flexible cagelike molecule, the errors upon ring breaking are much more substantial.

One possible explanation of this phenomenon is that, in the bicyclic system, the remaining ring is rigid enough—and the structural changes are small enough—to buffer the effect of bond length changes. However, in the bridged ring system, the geometry dictates that changes in distance between the atoms in question cannot easily be absorbed by small changes in other bond lengths, resulting in significant structural discrepancies between the conformation of the dummy atoms in water and in the binding environment, which, as we expected, will lead to larger errors.

We still need some way of determining whether these effects will be significant for real binding free energy calculations, so we examined strain in several real protein−ligand binding systems. Specifically, we examined simulations of several different protein−ligand complexes and the free ligands in solution to determine the magnitude of typical changes in bond length. The simulation trajectories were obtained from our former projects, which include six ligands in trypsin, two ligands bound to DNA gyrase[24] as provided by Vertex Pharmaceuticals, and ibuprofen in human serum albumin (HSA). Trajectories and parameter/coordinate files are provided in the Supporting Information. Our current research efforts do not provide good benchmarks for bond length changes in fused-ring systems, but we believe the systems examined here provide some idea of the amount of bond length change that can be expected in general, at least enough to give a rough idea of the size of the effect.

We found that in most of these simulations the bond length changes were small. The bond lengths in the binding site differ by less than 1% from those in solution. Ibuprofen binding to HSA proved to be an exception—we saw somewhat larger bond length changes, with two over 1% (Table 4). On the basis

Table 4. Bond Length Changes in a Real Protein−Ligand Complex: Ibuprofen Binding to HSA

atom ID 1	atom ID 2	bond length in water (Å)	bond length in complex (Å)	z score	percentage
C1	C3	1.388(1)	1.394(1)	3.4	0.4
C4	C6	1.389(1)	1.399(1)	6.1	0.7
C2	C5	1.388(1)	1.394(1)	4.1	0.5
C11	C5	1.517(2)	1.524(1)	3.3	0.5
C12	C6	1.515(1)	1.528(1)	6.8	0.9
C10	C13	1.535(1)	1.544(1)	4.7	0.6
C7	O1	1.215(1)	1.213(1)	1.6	0.2
C12	C8	1.534(2)	1.546(1)	5.8	0.8
C2	C4	1.388(1)	1.393(1)	2.8	0.3
C7	O2	1.306(1)	1.291(1)	9.4	1.1
C1	C5	1.385(1)	1.394(1)	5.2	0.6
C11	C13	1.535(2)	1.552(1)	6.8	1.1
C3	C6	1.387(1)	1.398(1)	6.9	0.8
C12	C7	1.508(1)	1.521(1)	7.4	0.9
C13	C9	1.536(1)	1.549(1)	6.4	0.8

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

of work in our model systems, bond length changes of this magnitude would be sufficient to cause errors larger than 0.5kT, which is small but notable, in the bicyclic system (Figure 3) and errors as large as 1−30 kcal/mol in the cagelike system depending on how we measure the original bond length (Figure 5; also see the Supporting Information).

4. CONCLUSIONS

Fundamentally, the error introduced by ring breaking results from coupling between dummy atoms in multiply connected groups (such as a ring system that has been turned into dummy atoms) and the conformation of the rest of the system. Specifically, the thermodynamic cycle used for relative free energy calculations assumes that the contribution of the dummy atoms to the free energy of the system is equivalent in the different environments, which is not in general the case. Strain in the ligand or solute induces some degree of conformational change that affects the free energy of the dummy atoms, so this assumption is no longer met.

In this study, we examined how this error introduced by ring breaking in relative free energy calculations grows as a function of ligand strain in a model binding system. We find that for bicyclic ring systems, the errors are relatively small (less than 0.5kT) and typically not statistically significant if the ligand strain is small—that is, if bond length changes caused by the binding environment are small (less than 2%). However, substantial changes in bond length as large as 1% do seem to occur in some real systems we have examined, suggesting that such perturbations ought to be avoided whenever possible. We further find that for cagelike or bridged molecules, when we remove the bridge, the errors grow much more rapidly as a function of ligand strain. In the system we examined here, even distance changes of 1% lead to errors of 1 to many kcal/mol (depending on how the change in bond length is measured). Furthermore, since ligand strain is difficult to predict a priori, there is no way to know in advance how big these errors will be for a specific system of interest. So in all we believe that ring breaking should be avoided in relative free energy calculations whenever possible, even for planar rings, though it is especially critical to avoid breaking bridged rings.

If researchers do need to calculate free energy changes for transformations involving ring breaking, we believe dual- or separated-topology[1,11] relative free energy calculations and absolute free energy calculations[1] may be better options, as these do not suffer from the same limitations.

APPENDIX

In alchemical relative binding free energy calculations, the two molecules before and after mutation usually have different topologies and differing numbers of atoms, and dummy atoms are therefore necessarily introduced into the calculations. To ensure that the effect of the dummy atoms exactly cancels out in the two legs of the simulations, certain rules must be followed regarding which interaction energy terms between the dummy atoms and the physical atoms are kept in the two end points of the relative calculation, specifically, in the states reached at the end-point λ values.

In general, the end-point λ window in alchemical RBFE simulations has the following parts: N physical atoms (labeled 1, 2, ..., N) in the mixed molecule, m dummy atoms (labeled a, b, c, ...,x), and i atoms in the surrounding environment (labeled S_1, S_2, ..., S_i and denoting the solvent, ion, and/or protein). The total interaction energy (bonded and nonbonded) has the following components: the interaction energy between the physical atoms (U_P^{self}), the interaction energy between the dummy atoms (U_D), the interaction between dummy atoms and physical atoms (U_{PD}), the interaction energy between particles in the environment (U_e), and the interaction between the environment and the physical atoms (U_{Pe}):

$$\begin{aligned} U(1, 2..., N; a, b, ..., x; S_1, S_2, ..., S_i) &= U_P^{self}(1, 2, ..., N) + U_D(a, b, ..., x) \\ &+ U_{PD}(1, 2, ..., N; a, b, ..., x) + U_e(S_1, S_2, ..., S_i) \\ &+ U_{Pe}(1, 2, ..., N; S_1, S_2, ..., S_i) \end{aligned} \quad (1)$$

Since the dummy atoms do not interact with the surrounding environment, we can define the following effective potential due to the surrounding environment by integrating over these degrees of freedom:

$$U_P^{env}(1, 2, ..., N) = -kT \ln \int dS_1 dS_2 ... dS_i \exp[-\beta(U_e + U_{Pe})] \quad (2)$$

The effective potentials from the environment in the two legs of the alchemical RBFE simulations are different. With the definition

$$U_P = U_P^{self} + U_P^{env} \quad (3)$$

the configurational part of the partition function for the whole system simplifies to

$$Q = \int d1 d2 ... dN da db ... dx \exp[-\beta(U_P + U_D + U_{PD})] \quad (4)$$

It is easy to show that if there are only one bonded stretching interaction, two bonded angle interactions, and three bonded dihedral angle interactions between the physical atoms and the dummy atoms at the end point, then the effect of the dummy atoms exactly cancels out in the two simulations.[25] (These interactions could be labeled r_{1a}, θ_{21a}, θ_{1ab}, ϕ_{321a}, ϕ_{21ab}, and ϕ_{1abc}, where the subscripts stand for the atom numbers in Figure 8; for example, θ_{1ab} refers to the angle between bond 1a and bond ab.) This is also true for fewer retained interactions. The $3m$ degrees of freedom for the dummy atoms can be decomposed into $3m - 6$ internal degrees of freedom for the dummy atoms (x_1^D, x_2^D, ..., x_{3m-6}^D) and six degrees of freedom joining the dummy atoms with the physical atoms (x_1^{PD}, x_2^{PD}, ..., x_6^{PD}, corresponding to the six degrees of freedom listed above). Therefore,

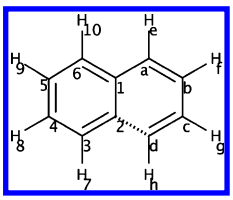


Figure 8. Two-dimensional structure for naphthalene. Considering a transformation from naphthalene to benzene, the dashed line highlights the fact that the (right-hand) ring, which would be transformed into dummy atoms, is doubly connected to the remaining ring.

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

$$
\begin{aligned}
&Q(1, 2, ..., N, a, b, c, ..., x) \\
&= \int \mathrm{d}1\mathrm{d}2 ... \mathrm{d}N\mathrm{d}a\mathrm{d}b ... \mathrm{d}x \exp[-\beta(U_P + U_D + U_{PD})] \\
&= \int \mathrm{d}1\mathrm{d}2 ... \mathrm{d}N \exp(-\beta U_P) \int \mathrm{d}x_1^{PD}\mathrm{d}x_2^{PD} ... \mathrm{d}x_6^{PD} \\
&\quad \exp(-\beta U_{PD}) \int \mathrm{d}x_1^{D}\mathrm{d}x_2^{D} ... \mathrm{d}x_{3m-6}^{D} \exp(-\beta U_D) \\
&= Q(1, 2, ..., N)Q^D
\end{aligned} \tag{5}
$$

If we now suppose that in addition to the interactions involving the six degrees of freedom mentioned above, one more interaction between the dummy atoms and the physical atoms is retained in the end-point λ window, such as the bonded stretching interaction between atoms 2 and d in our example of a ring-closing mutation involving perturbation of a benzene ring into a naphthalene ring (Figure 8). If this extra interaction is retained, then

$$
\begin{aligned}
&Q(1, 2, ..., N, a, b, c, ..., x) \\
&= \int \mathrm{d}1\mathrm{d}2 ... \mathrm{d}N\mathrm{d}a\mathrm{d}b ... \mathrm{d}x \exp[-\beta(U_P + U_D + U_{PD})] \\
&= \int \mathrm{d}1\mathrm{d}2 ... \mathrm{d}N \exp(-\beta U_P) \int \mathrm{d}x_1^{PD}\mathrm{d}x_2^{PD} ... \mathrm{d}x_6^{PD} \\
&\quad \int \mathrm{d}x_1^{D}\mathrm{d}x_2^{D} ... \mathrm{d}x_{3m-6}^{D} \exp[-\beta(U_{PD}^6 + U_{r_{2d}})] \exp(-\beta U_D) \\
&= \int \mathrm{d}1\mathrm{d}2 ... \mathrm{d}N \exp[-\beta(U_P + U_P^{eff})]
\end{aligned} \tag{6}
$$

where $U_P^{eff}(1, 2, ..., N) = -kT \ln \int \mathrm{d}x_6^{PD}\mathrm{d}x_{3m-6}^{D} \exp[-\beta(U_{PD}^6 + U_{r_{2d}} + U_D)]$ is the effective potential (restraint) applied on the physical atoms due to the interactions with the dummy atoms. It should be noted that the bond stretching potential between atoms 2 and d ($U_{r_{2d}}$) is introduced because of the restriction of an extra degree of freedom aside from the six rigid-body degrees of freedom. Thus, the final result in eq 6 cannot be separated into separate integrals as in eq 5 because having only those six degrees of freedom restrained is the prerequisite for ensuring that the thermodynamic properties of the dummy atoms and those of the physical atoms are independent.[25] Thus, with this extra interaction, $U_P^{eff}(1, 2, ..., N)$ is no longer separable into a part for the dummy atoms and a part for the interaction between physical and dummy atoms. Therefore, the inclusion of additional bonded stretching interactions for the ring-opening/closing RBFE calculations introduces a conformational bias for the ligand simulated, and the effect of the dummy atoms does not cancel out in the two legs of the simulations, leading to errors in the calculations.

■ ASSOCIATED CONTENT

Supporting Information

Example GROMACS topology and geometry files for all three systems and detailed information on bond length changes in the protein−ligand systems examined, except those already provided in the main text. This material is available free of charge via the Internet at http://pubs.acs.org. Additional supporting information, in the form of trajectory files for the real protein−ligand systems examined, is available via eScholarship at www.escholarship.org/uc/item/27d9s5j9.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dmobley@mobleylab.org.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We appreciate financial support from the National Institutes of Health (1R15GM096257-01A1 and 1R01GM108889-01) and computing support from the UCI GreenPlanet Cluster, supported in part by NSF Grant CHE-0840513.

■ REFERENCES

(1) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137*, No. 230901.

(2) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous Free Energy Calculations in Structure-Based Drug Design. *Mol. Inf.* **2010**, *29*, 570−578.

(3) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical Prediction of Hydration Free Energies for SAMPL. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 551−562.

(4) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569−1582.

(5) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755−770.

(6) By "ring breaking" we here refer to transformations that turn a ring into dummy atoms connected by bonds. In contrast, transforming complete rings into partial rings by breaking bonds is outside the scope of most current classical molecular dynamics approaches.

(7) Boresch, S.; Karplus, M. The Role of Bonded Terms in Free Energy Simulations. 2. Calculation of Their Influence on Free Energy Differences of Solvation. *J. Phys. Chem. A* **1999**, *103*, 119−136.

(8) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *J. Phys. Chem.* **1994**, *98*, 1487−1493.

(9) Michel, J.; Verdonk, M. L.; Essex, J. W. Protein-Ligand Complexes: Computation of the Relative Free Energy of Different Scaffolds and Binding Modes. *J. Chem. Theory Comput.* **2007**, *3*, 1645−1655.

(10) Michel, J.; Essex, J. W. Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. *J. Med. Chem.* **2008**, *51*, 6654−6664.

(11) Rocklin, G. J.; Mobley, D. L.; Dill, K. A. Separated Topologies—A Method for Relative Binding Free Energy Calculations Using Orientational Restraints. *J. Chem. Phys.* **2013**, *138*, No. 085104.

(12) More precisely, turning part of the ring into dummy atoms.

(13) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(14) *OpenEye Unified Python Toolkit*; OpenEye Scientific Software: Santa Fe, NM, 2012.

(15) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

(16) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.

(17) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(18) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(19) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735

(20) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE—AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367−374.
(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(22) Klimovich, P. V.; Mobley, D. L. Predicting Hydration Free Energies Using All-Atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 307−316.
(23) Except for the NDE case.
(24) Varela, R.; Walters, W. P.; Goldman, B. B.; Jain, A. N. Iterative Refinement of a Binding Pocket Model: Active Computational Steering of Lead Optimization. *J. Med. Chem.* **2012**, *55*, 8926−8942.
(25) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.

DOI: 10.1021/acs.jcim.5b00057
J. Chem. Inf. Model. 2015, 55, 727−735