Published in final edited form as:

J Chem Theory Comput. 2018 June 12; 14(6): 3279–3288. doi:10.1021/acs.jctc.8b00031.

Enhanced Monte Carlo Methods for Modeling Proteins Including Computation of Absolute Free Energies of Binding

Israel Cabeza de Vaca, Yue Qian, Jonah Z. Vilseck, Julian Tirado-Rives, and William L. Jorgensen*

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

Abstract

The generation of a complete ensemble of geometrical configurations is required to obtain reliable estimations of absolute binding free energies by alchemical free-energy methods. Molecular dynamics (MD) is the most popular sampling method, but the representation of large biomolecular systems may be incomplete owing to energetic barriers that impede efficient sampling of the configurational space. Monte Carlo (MC) methods can possibly overcome this issue by adapting the attempted movement sizes to facilitate transitions between alternative local-energy minima. In this study, we present an MC statistical mechanics algorithm to explore the protein-ligand conformational space with emphasis on the motions of the protein backbone and side chains. The parameters for each MC move type were optimized to better reproduce conformational distributions of 18 dipeptides and the well-studied T4-lysozyme L99A protein. Next, the performance of the improved MC algorithms was evaluated by computing absolute free energies of binding for L99A lysozyme with benzene and seven analogs. Results for benzene with L99A lysozyme from MD and the optimized MC protocol were found to agree within 0.6 kcal/mol, while a mean unsigned error of 1.2 kcal/mol between MC results and experiment was obtained for the seven benzene analogs. Significant advantages in computation speed are also reported with MC over MD for similar extents of configurational sampling.

Graphical Abstract

*Corresponding Author: william.jorgensen@yale.edu.

Supporting Information. It consists of five Figures and a Table with Ramachandran trajectory plots, lysozyme C α RMSF evolution in the MD trajectory, all dipeptide Chi distributions, lysozyme-benzene free energy results per window, and lysozyme-benzene analog free energy evolution, and lysozyme-benzene analog free energies results for 1.14*CM1A and OPLS charge models. This information is available free of charge on the ACS Publications website at <http://pubs.acs.org>

Notes

The authors declare no competing financial interest.

INTRODUCTION

In the last thirty years, many free energy methods have been explored to study biomolecular phenomena including protein folding,¹ hydration,² and often, free energies of binding.^{3,4} Most of these calculations have featured computing free energy differences with Thermodynamic Integration (TI),^{5,6} Free Energy Perturbation (FEP),^{7,8} Bennett Acceptance Ratio (BAR)⁹ and Weighted Histogram Analysis Methods (WHAM).¹⁰ The precision of the approaches is dependent on performing an extensive sampling of the configuration space and in having an accurate energy evaluation through molecular mechanics force fields.¹¹ In biomolecular simulations the exploration of the conformational landscape has been commonly done through molecular dynamics (MD) methods, both due to their sampling efficiency and the easy availability of software packages. However, the convergence of free energy calculation is still hindered by the large number of possible geometrical configurations for protein or nucleic acid complexes. In order to improve their efficiency, different strategies have been developed such as umbrella sampling¹², replica exchange,^{13,14} grand canonical free energy computations,¹⁵ and metadynamics.¹⁶ Moreover, leading MD software packages, such as GROMACS,¹⁷ AMBER,¹⁸ NAMD,¹⁹ CHARMM²⁰ and OpenMM,²¹ have been ported to graphic processing units (GPUs) to increase their performance.

Metropolis Monte Carlo statistical mechanics (MC) is an alternative sampling strategy that generates configurational ensembles using discrete, localized, random variations of geometrical degrees of freedom (moves) rather than integration of Newton's equations of motion. MC methods avoid possible instabilities associated with the calculation of forces during the simulations, and they may converge faster than MD by more efficient exploration of disparate local minima.²² In addition, the local moves allow for focused sampling in the regions of interest, e. g., binding-site residues in proteins, with the concomitant gains in speed. On the other hand, the large number of possible moves in biomolecular systems and their interdependence makes it challenging to find an efficient way to explore the available conformational space, or a sufficient subset thereof. The locality of moves in a long chain, such as a protein backbone, implies that small angular or torsional changes may result in large displacements far from the site of modification. Such large moves usually result in high-energy configurations that lead to poor acceptance rates and sampling. These problems are avoided in MD simulations due to the nature the algorithm, which features small Cartesian displacements of all atoms in the system as directed by the forces. The global

character of the motion, however, makes it difficult to enhance performance by focusing the sampling in regions of interest.

Different MC-based packages have been developed that generate configurations based on different move types in order to study thermodynamic properties including free energies. From our group, MCPRO and BOSS are two software packages where MC methods have been implemented and successfully applied to many problems including evaluation of relative binding free energies for host-guest systems, free energies of solvation, and QM/MM free-energy surfaces for organic and enzymatic reactions.²³ MCPRO is used for biomolecular systems that feature macromolecules consisting of multiple residues, while BOSS is directed towards modelling organic molecules in the gas-phase and in solution. Both packages perform MC simulations using a combination of localized moves augmented with global volume changes, when the isothermal-isobaric ensemble is invoked, to generate the configurations needed to calculate free energy changes via free energy perturbation (FEP) theory. Other groups have developed software that implements similar methodology such as ProtoMS²⁴ and SIRE,²⁵ while with PELE^{26,27} a combination of local and global moves (side chain rotamer exploration, ligand rotation/translation and normal modes) are used to estimate absolute binding free energies using Markov state models (MSM).²⁸ This approach has been successfully applied to study ligand binding for proteins and nucleic acids.^{29,30}

The accuracy of energetic evaluations has also been the subject of research for a long time. In our laboratories, extensive efforts have been devoted to steady improvement of the OPLS force fields for proteins,³¹ nucleic acids,³² and organic ligands,³³ and similar efforts have been conducted in other groups for the AMBER,³⁴ OPLS3,³⁵ and CHARMM³⁶ force fields. However, this report focuses on the sampling issue, and presents an improved MC algorithm, which includes both protein and ligand flexibility to explore thoroughly the conformational space while maintaining the ability to focus the sampling on regions of interest. This new MC algorithm is composed of five independent types of moves: volume moves, conformational and rigid body motions for the ligand or solvent molecules, protein side-chain variations, and two types of backbone moves applied with different frequencies. As the parameters for ligand/solvent and volume moves have been widely used and optimized through studies of pure liquids and free energies of hydration,³⁷ this work focuses on improvement of the parameters controlling the side-chain and backbone moves. Given the lack of direct experimental measurements of the structural fluctuations in solution, these parameters were optimized to reproduce the conformational distributions obtained from MD simulations of 18 standard amino acid dipeptides and a well-known protein, T4 lysozyme L99A.

As a more stringent test of its capabilities, the optimized MC algorithm was used to calculate absolute free energies of binding via the “double annihilation” procedure. The complexation of the L99A mutant of T4 lysozyme (hereinafter referred to as lysozyme) with benzene and seven analogs was chosen due to its relative simplicity, well-known structure, and the wealth of available data, both from experiments and simulations. The mutated lysozyme was created artificially to study the ligand binding process with pure hydrophobic interactions.³⁸ As precise experimental binding free energies for several small aromatic

ligands have been obtained by calorimetry,³⁸ this system has been widely used to test the accuracy of binding-affinity calculations.^{39,40} Good agreement is found between MD and the present MC simulations for the free energy of binding of benzene to lysozyme. The same MC/FEP parameters were then used in computing absolute binding free energies for seven benzene analogs. A mean unsigned error (MUE) of 1.23 kcal/mol with respect to the experimental measurements is obtained. Comparison with results from previous MD studies demonstrates the viability of the present MC/FEP method.^{15,41,42} The benefits for convergence of the MC approach are demonstrated by analysis of the MUE results along the simulations.

THEORY and METHODOLOGY

Monte Carlo Sampling Algorithm.

The MC simulations in this study have been carried out with MCPRO (Monte Carlo for PROteins), version 3.2, which includes the described modifications. In this program, the MC moves are controlled by a user-defined set of move frequencies and ranges to sample protein backbone, side-chain, ligand, and solvent degrees of freedom. Water molecules are represented by the rigid TIP4P model;⁴³ thus, solvent moves only involve random rigid-body translations and rotations. The ligand moves are a combination of rigid-body translations and rotations with variations in bond lengths, bond angles and dihedral angles. The protein sampling includes sidechain moves, where bond angles and dihedrals angles of the side chains are modified, while keeping the residue's backbone fixed. And, backbone sampling is performed with two MC algorithms: concerted rotations with variable angles (CRA) and a pivot algorithm (described below). After each move the final configuration is accepted or rejected using the Metropolis criterion at a temperature of 300 K. All parameters controlling the moves were adjusted to produce global acceptance rates of 30–40%.

Concerted Rotations with Angles.

Concerted rotations with angles (CRA) is a biased MC algorithm composed of two parts, prerotations and chain closure.^{44–46} It provides for crankshaft-like modifications of a polymer backbone. Briefly, each CRA step selects at random a continuous chain of five consecutive residues of the protein to perform localized moves. Prerotations, which are applied to the first four of the selected residues consist of random moves of backbone bond angles and dihedrals. The ranges are controlled by three adjustable parameters where two parameters (c_1 , c_2) guide the dihedral moves, and one parameter (c_3) controls the bond-angle changes. The three parameters are then scaled with the move size, chosen from a Gaussian random distribution. Their values were selected to balance the acceptance ratio and the largest movement possible. In the original publication,⁴⁴ optimal values for c_1 , c_2 , and c_3 were found to be 100, 8, and 20 respectively. These values arose from evaluations of the folding of (Ala)₁₄ and the villin headpiece (36 residues) yielding a 30% acceptance.⁴⁴ On the other hand, the chain-closure adjustments applied between the fourth and fifth residues come from a geometrical algorithm based on adjustments to the bond and dihedral angle of the nitrogen connected to the last carbon atom perturbed by the prerotation to remove discontinuities. Although this particular type of move was already included in MCPRO, the existing implementation was replaced by a more streamlined version. The current

implementation of this algorithm also substitutes the original, iterative determination of the chain closure parameters with an analytical solution. These changes, and the addition of the pivot move warranted the re-evaluation of the optimal ranges for the rotation.

The MC simulations that were performed to optimize the values of c_1 , c_2 , and c_3 were carried out using the recent OPLS-AA/M force field,⁴⁷ a spherical water cap of 25-Å radius centered on lysozyme's binding site containing 1516 TIP4P water molecules, and a temperature of 300 K. The total charge of lysozyme was adjusted to zero by neutralizing the 8 positively charged residues farthest from the binding site, Lys16, 19, 43, 48, 60, and 65 and Arg14 and 52 (all atoms of these residues are separated by more than 20 Å from the ligands). The nonbonded energy terms were evaluated using a residue-based 10-Å cutoff. The frequencies for CRA and side chain moves were set to every 71 and 10 configurations to minimize the overlap between different move types and allow significant movement of the water molecules. A total of 1 million (1M) and 50M MC steps were performed for the equilibration and averaging phases, respectively.

Pivot Axial Rotation.

In this work an additional type of move is added to improve the backbone conformational sampling. It is termed Pivot, since it is comprised of an axial rotation (pivoting) of the atoms between two consecutive alpha carbons. It is equivalent geometrically to the Backrub⁴⁸ move used in modelling of conformational changes in proteins and the fixed end point method used in coarse-grained folding simulations.⁴⁹ When a Pivot move is chosen in a given MC step, a residue is selected at random, and the vector between the alpha carbons of the selected and the previous residue is used as an axis for a random rotation of the atoms between them, namely the amide group, in the range between the negative and positive values of a maximum value specified by the user. As the amide group is rotated it modifies the backbone ψ dihedral of the preceding residue and ϕ of the selected residue as shown in Figure 1. The optimal maximum rotation angle was determined to be about 25° by running systematic trials with the decaalanine peptide and the full lysozyme protein, which maintained an approximate 30% acceptance ratio in both systems.

Side-chain Sampling.

To optimize and analyse the χ distributions for the standard amino acids, a set of 18 dipeptides (glycine and alanine were excluded) were considered. For each dipeptide, a MC simulation was performed using the OPLS-AA/M force field in a cubic TIP4P water box created with a padding of 9 Å around the entire solute. All nonbonded energy cutoffs were set to 9 Å, and periodic boundary conditions were applied. The system preparation included a short conjugate gradient minimization and 200K steps of NVT MC simulation, followed by 28M NPT steps at 300 K and 1 atm. The production phase was then continued for 500M steps in the NPT ensemble. These runs were then repeated at different values of the maximum allowed dihedral moves. The total running time of each simulation was around 5 hrs in a single Core2 Quad 3.3 GHz processor.

Molecular Dynamics.

All MD simulations were performed with NAMD,¹⁹ using the OPLS-AA/M force field.⁴⁷ The simulations were run in a truncated octahedral box using the TIP3P water model.⁴³ The minimal distance between solute atoms and the box edges was 10 Å, and when necessary, the system was neutralized by adding the needed number of Cl⁻ or Na⁺ counterions. No additional salt was added since the binding site of lysozyme is sequestered from solvent and it would not be expected to influence the results. For all simulations, a time step of 2 fs was used, and the SHAKE algorithm⁵⁰ was applied to constrain all bonds to hydrogen atoms. Nonbonding interactions were truncated using a 12-Å cutoff for the dipeptides, and 10-Å for lysozyme. Long-range electrostatic interactions were corrected using the Particle-Mesh-Ewald (PME) procedure.^{51–53} The equilibration protocol consisted of 5000 steps of conjugate-gradient minimization followed by 100 ps of equilibration maintaining the temperature and pressure at 300 K and 1 atm using a Langevin barostat and thermostat.^{54,55} The production phases were then run for 10 and 100 ns for the dipeptides and lysozyme, respectively.

ABSOLUTE BINDING FREE ENERGIES

Monte Carlo Protocol.

The calculations of absolute free energies of binding were carried out by annihilating the ligand both unbound in water and bound to the protein using the free energy perturbation (FEP) method in a single topology mode. The protein and ligands were represented with the OPLS-AA/M force field.⁴⁷ In addition, a semiempirical quantum mechanics based charge model, 1.14*CM1A, as generated by the LigParGen server,⁵⁶ was also tested for the ligands. The ligand annihilations were performed using simple overlap sampling (SOS)^{57,58} with decoupling of the electrostatic and Lennard-Jones (LJ) terms. First the charges were scaled to zero linearly with the lambda parameter, followed by the LJ terms, which was performed using the 1-1-6 soft-core potential⁵⁹ to avoid instabilities.

The annihilations of the charge and LJ parameters were split into 15 and 18 λ -windows, respectively. Each window covered 80M steps of equilibration and 180M steps of averaging. A hard-wall (HW) potential constraint was applied to the center of the ligand during the LJ annihilation to restrain its movement to a sphere centered in the binding site. The radius of the sphere was set at 2.8 Å and its origin was placed in the geometric center of the benzene molecule in the initial X-ray structure. The radius of the constraint was chosen so that in the course of the simulation the ligand could reach the entire binding pocket. A correction term corresponding to the HW constraint was included in the binding free energy estimation (eq 1), where V_{eff} is the effective volume of ligand, i.e., the volume of the HW

$$\Delta G_{HW} = -k_B T \ln \left(\frac{V_{eff}}{V_0} \right) \quad (1)$$

sphere and V_0 is $1660 \text{ \AA}^3$ (corresponding to a 1M standard state), T is the temperature and k_B is the Boltzmann constant.⁶⁰ The final absolute binding free energy was therefore computed using eq 2.

$$\Delta G_{bind} = \Delta G_{unbound} - \Delta G_{bound} + \Delta G_{HW} \quad (2)$$

A 25- $\text{\AA}$ radius cap with 1514 TIP4P water molecules centered on the benzene ligand was used to solvate the lysozyme-benzene system for the MC simulations, and the nonbonded energy terms used a 10- $\text{\AA}$ cutoff. The frequencies for backbone moves were set to and 11 for CRA and Pivot moves, and side chain and ligand move frequencies were set to 10 and 3. As discussed below, the CRA parameters c_1 , c_2 and c_3 were assigned values of 100, and 20, respectively. Additionally, the temperature was set to 300 K. The unbound ligand simulations were run in a periodic cube containing 1574 water molecules with a ligand move frequency of 60 configurations; the remaining parameters were kept the same as in the bound simulations.

Molecular Dynamics Protocol.

The MD simulations needed to calculate the absolute free energy of binding were performed with the same settings as described above. For all unbound ligand simulations, the equilibration protocol consisted of 5000 steps of conjugate gradient minimization followed by 100 ps of equilibration at 300 K and 1 atm. For the bound simulations, the system was first subjected to 2 ns of pre-equilibration with the protein-ligand complex stationary. Minimization and equilibration protocols were then performed for 500K steps and 2 ns, respectively.

After equilibration, MD/FEP simulations were performed in both growing and shrinking directions using 32 and 43 λ -windows for the unbound and bound states, respectively. In the bound-state calculation, benzene was restrained to its initial position relative to the protein using harmonic restraints for six degrees of freedom.⁶¹ The harmonic restraints were the three spherical coordinates and three Euler angles needed to constrain three atoms of the ligand with respect to three atoms of the protein. Analytical corrections to the standard state were again included to obtain the final absolute free energy of binding.⁶² For the lysozyme-benzene system, each production window covered 5 ns, and all simulations were run in triplicate. It should be noted that for the purposes of this study the typical protocols used in each type of simulation were followed, rather than trying to replicate the settings among them.

RESULTS AND DISCUSSION

CRA Parameter Optimization.

The optimization of the CRA parameters was performed using L99A lysozyme (PDB ID: 4W52), with the benzene ligand removed, as the test case. This 164-residue protein features an engineered, solvent excluded, hydrophobic binding pocket and has been extensively studied both experimentally and computationally.^{41,42,62–67} A range of values for the three

CRA adjustable parameters, which control the prerotation stage, were tested to improve the backbone sampling. The values tried for the two parameters related with the dihedral movement were 100, 150 and 200° for c_1 and 6, 8, 10, 12, 15, 17 and 20° for c_2 , while values of 10, 15, 20, 25 and 30° were tested for the bond angle parameter (c_3). Since the parameters are interdependent, all possible combinations of these values were considered by running 105 independent MC simulations, as described in the section on Monte Carlo methods. The results were compared to a 100-ns MD trajectory performed as described in the Molecular Dynamics protocol section above. The trajectory comparison was performed using two different metrics: the overlap between the distributions in Ramachandran plots for 16 residues selected randomly around the binding site and the root mean square fluctuations (rmsF) of C^α positions.

Many parameter combinations gave similar results. As an illustration, the Ramachandran plot for Val111 (located in the binding site) from the simulation with the CRA parameters $c_1=150$, $c_2=17$ and $c_3=30$ is shown in Figure 2A. It can be seen that the overlap is significant between the 50M MC and 100-ns MD simulations, which demonstrates that the sampling of ψ/ϕ space explores similar regions. Consistent results were obtained for all 16 residues that were analysed (see Supplementary Figure S1).

The C^α rmsF analyses between MC and MD simulations also indicated similar sampling. The configurations generated with CRA over a 50M step run showed an average rmsF difference of 0.3 Å to the 100 ns-long MD simulation. As a key goal is to apply this MC method to binding free energy calculations, a more appropriate comparison is to MD runs of 5–20 ns, the typical times used in each window of MD/FEP simulations. Figure 2B shows the rmsF of all residues from a 20 ns MD run and 50M MC with CRA using $c_1=150$, $c_2=17$ and $c_3=30$. In this case, the average rmsF difference between MC and MD is reduced to 0.2 Å. The results for configurations generated with the other choices of CRA parameters ranged from 0.2 to 0.8 Å. Figures 2C and 2D then show the backbone variations from 50 equally distributed trajectory frames over 20 ns of MD and 50M steps of MC simulation. As expected from the rmsF plots, the structures from 20 ns of MD display larger fluctuations on average than the snapshots from the 50M MC simulation. The spread on the backbone atoms is larger in the bottom “domain” (residues 1–12 and 61–164) of the MD simulation, but narrower in the top (residues 13–60). The evolution of the rms fluctuations for different lengths of the MD trajectory is shown in Supplementary Figure S2.

Overall, the analyses of Ramachandran plots and rmsF did not reveal substantial differences among the alternative choices for the CRA parameters, and the results were consistent with the ranges found with MD. Thus, it was decided to use default parameters of $c_1=100$, $c_2=8$ and $c_3=20$ for the calculations of absolute binding free energies.

Sidechain Dihedral Move Optimization.

The ranges for moves of the amino acids' χ angles were optimized to increase the side-chain conformational sampling and improve the MC convergence. The 18 dipeptides were capped with acetyl- and N-methyl- groups at the N- and C- termini. To optimize the dihedral moves, the maximum move ranges were varied from 0 to 360° every 30° for each χ . For residues with more than one χ all possible combinations were evaluated, thereby generating 144,

1728 and 20736 possible ranges for residues with two, three and four χ s, respectively. Dihedral distributions for each 500M MC simulation were then compared against 10-ns MD trajectories to find an optimal set of ranges. The length of the MD simulation was chosen based on its sampling most of the available dihedral space. As an example, the χ_1 , χ_2 and χ_3 distributions obtained with the best set of parameters found for methionine (Met) are shown in Figure 3.

As reflected in Supplementary Figure S3, parameters could be found that produce good agreement between MD and MC for all dipeptides. The resultant optimal χ ranges are recorded in Table 1. As could be expected, not all reasonable minima were explored when using some of the smaller ranges, but instead the system tended to be trapped in the local minimum closest to the starting point during the entire MC simulation. While this result could be changed by increasing the run lengths, the goal was to find ranges that produce good distributions in the shortest time. Indeed, it was found that larger ranges for dihedral moves were preferred. This contrasts our former default practice of using ranges of $\pm 15^\circ$ for variations of side-chain dihedral angles.²³ The 10-ns MD simulations were performed in a single core (intel Core2 Quad 3.3Ghz) and took between 19–25 hours depending on the dipeptide size. The 500M MC simulations of the same set of dipeptides with similar water box sizes took between 4.5–5 hours using the same processor. Therefore, MC produced equivalent χ distributions 4–5 times faster than MD; however, the MD simulations used a longer cutoff (12 vs 9 Å). The impact of this difference was addressed by running a sample MD simulation for the Arg dipeptide with the shorter cutoff. With the shorter cutoff the time for the MD simulation was halved, so the MC simulations produce equivalent conformational distributions 2–2.5 times faster than MD with the same cutoffs. If only a simple sampling simulation is desired, most MD packages leverage the use of GPUs to accelerate the calculation. In our hands, the same MD simulations on the lysine dipeptide using a single core and a single Tesla K20m GPU provided a speedup of a factor of 2, while utilizing 4 cores and one GPU speeds up the calculation by a factor of 23.

Lysozyme-benzene Absolute Free Energy of Binding.

Once the MC sampling algorithm was optimized, the next step was to test its application for calculation of the absolute binding free energy of the L99A lysozyme-benzene complex. Benzene was chosen for the initial test owing to its small size and rigidity, which facilitate the annihilations, and to the availability of accurate experimental data.^{39,40}

As detailed in the methods section, the ligand annihilations in the bound state and unbound in water were done in two stages, decoupling the charge and Lennard-Jones portions. The free energy changes for the unbound benzene annihilation in water were determined by MC/FEP to be 0.23 kcal/mol for the Coulomb term and –2.3 kcal/mol for the LJ part, while the corresponding values for the bound system were –1.34 and 7.42 kcal/mol, and the hard wall correction is 1.71 kcal/mol (eq 1). The absolute free energy of binding can be calculated from these values using eq 2. The evolution of the final result as a function of the number of MC steps is shown in Figure 4A. The result is well converged with negligible fluctuations after 30M configurations, though the simulations were continued through 180M configurations per window to ensure that sampling was complete. After 180M steps, the

resulting free energy of binding is -6.44 ± 0.22 kcal/mol. Additionally, two other simulations were performed to verify the appropriateness of the HW size, using larger radii of 4 and 5 Å; the results were indistinguishable from those with the 2.8-Å radius. For this system, the variation of computed binding free energies was small (Figure 4A), ranging from a low value of -6.75 kcal/mol at 10M steps, well before convergence, to values very close to -6.4 kcal/mol from 30M configurations on.

The MD simulations were conducted in triplicate and used harmonic restraints for six degrees of freedom. After the corrections for the restraints and symmetry were applied, the absolute binding free energies obtained for each run were -6.02 , -5.58 and -5.94 kcal/mol, to give an average of -5.85 ± 0.19 kcal/mol. The evolution of the average binding free energy as a function of the simulation time for each window is shown in Figure 4B. There is only a change of 0.1 kcal/mol after 3 ns, though there continues to be an upward drift to 5 ns. The difference between the final results using the MC and MD methods was only 0.59 kcal/mol. The agreement is notable given the methodological differences. In addition to the fundamental difference in the MC and MD algorithms, the setups were chosen to reflect typical FEP simulations conducted for each method. In particular, the MC simulations were run in a cap of TIP4P water using a residue-based truncation for the non-bonded energies and a simple hard-wall restraint for the ligand. The MD simulations, on the other hand, utilized a periodic truncated-octahedral box of TIP3P water with Ewald corrections for the long-range electrostatic interactions, and the restraint of six degrees of freedom for the ligand.

Still, in the present study, the free energies of binding calculated by MC (-6.44) and MD (-5.85) had differences of 1.25 and 0.66 kcal/mol to the experimentally measured affinity of -5.19 ± 0.16 kcal/mol.⁴⁰ These are well within the 2 kcal/mol rms error found for typical calculations of free energies of binding for proteins.⁶⁸ For the specific case of benzene-lysozyme, some of the published results using different methods and force fields are -4.01 ,⁶⁹ -4.26 ,⁷⁰ -5.96 ,⁴¹ -6.0 ,⁷¹ and -8.53 kcal/mol.⁷² Since benzene represents a simple case, routine calculation of absolute free energies of binding for protein-ligand systems with errors less than 1 kcal/mol remains a formidable challenge.

A set of additional MD and MC simulations were conducted in order to probe the effects of the allowed sampling, as presented in Table 2. In MD, two FEP calculations were run in which all heavy atoms in the backbone were constrained to their original positions with a force constant of 10 kcal/mol-Å² and another where all C^α atoms were constrained in the same fashion. The resultant calculated values for G_{bind} are -9.93 and -9.45 kcal/mol, which are significantly lower than the result with no constraints, -5.85 kcal/mol. The sampling in MC simulations was controlled by first not allowing any backbone moves, and then by adding only CRA moves, and compared to the final run including both CRA and pivot moves. The values calculated for G_{bind} systematically improved from -9.97 to -6.82 , and to the final value of -6.44 kcal/mol with full sampling. It is interesting to note that the result of the backbone-constrained MD simulation (-9.93) is very similar to the MC simulation with rigid backbone (-9.97). Examination of Table 2 shows that in both algorithms the increase of allowed mobility, going down in the table, produces less negative binding free energies that are closer to the experimental value.

It is important to note the computational times used for the two approaches in the time-dominant bound calculations. In the MC case, the FEP results converged within 30M steps, requiring one cpu-day of computation time per window. In contrast, a 3-ns MD run required ~6.6 cpu-days per window with the same 10-Å cutoffs to provide reasonable convergence in the same 2.2 GHz Intel Xeon E5-2660 V2. Considering that the MC calculations used 33 windows and the MD calculations used 43, the MC approach provided similarly converged results about 9 times faster than by MD with the software utilized, MCPRO and NAMD. The latter package does not have FEP calculations implemented for a GPU.

Benzene Derivatives.

The same MC/FEP sampling scheme and parameters from the lysozyme-benzene simulations were then tested further for seven benzene analogs (toluene, *o*-xylene, *p*-xylene, ethylbenzene, benzofuran, indene, and indole). The final run lengths were again 180M configurations per window providing around 6000M MC steps per ligand in each bound and unbound simulation. The seven new complexes were prepared by directly replacing the benzene ligand by the desired analogs using maximum overlap of the non-hydrogen atoms. Initially, the ligands were represented using OPLS-AA Lennard-Jones parameters with 1.14*CM1A charges, which produced the results in column 2 of Table 3 with their associated errors in column 5. The average MUE including benzene was 2.16 kcal/mol.

However, the value for indole deviated from the experimental result by 6 kcal/mol. Omitting it reduces the MUE to 1.58 kcal/mol. Since it is known that nitrogen atoms in aromatic rings can be problematic for CM1A charges,⁷³ it was decided to test as well the original OPLS-AA force field charges of all ligands, which were parameterized by computing the structural and thermodynamic properties of pure organic liquids.⁷⁴ The results for the charge perturbation for indole (Supplementary Table S1) suggested that the molecule is too polar with the scaled CM1A charges. As there is no experimental value for the free energy of hydration of indole, a definitive test was not possible. However, calculations using the Generalized Born continuum method⁷⁵ confirmed the trend with G_{hyd} values of -10.99 and -5.53 kcal/mol using the 1.14*CM1A and OPLS-AA charge models, respectively. As an additional comparison, the hydration free energy for 3-methylindole was calculated using the FEP/MC methodology. The 1.14*CM1A and OPLS-AA charge models produced G_{hyd} values of -10.90 and -3.02 kcal/mol, respectively, with the latter closer to the reported experimental value of -5.88 kcal/mol.⁷⁶ Thus, the much too favorable free energy of hydration with the 1.14*CM1A charges is consistent with the too unfavorable free energy of binding of indole with lysozyme. The deficiency is in the charge model for indole, since the same Lennard-Jones parameters are used in both cases. Particularly suspect are the charges for the N-H moiety, -0.673 and 0.457 e with the 1.14*CM1A model, significantly greater in magnitude than the -0.500 and 0.376 e with the OPLS-AA force field.

The results for the absolute free energies of binding with the OPLS-AA charge model for the eight complexes are shown in columns 3 and 6 of Table 3. Reassuringly, the free energies are in better agreement with the experimental results, giving a much improved MUE of 1.23 kcal/mol. The smaller MUE better reflects the potential accuracy of this MC method to predict absolute free energies of binding, although it should be emphasized that the results

are both very sensitive to the charge choices, and the relative differences between different ligands fall short of the desired precision. Given the small differences in binding free energies, the latter calculations would be better done by simulations of the direct alchemical transformations between ligand pairs that focus the sampling in the more relevant regions of configurational space. Still, these findings compare favorably to results obtained using other methods and force fields. As a comparison, the values of the MUEs obtained for the same set of eight ligands in the literature are 1.67,⁴¹ 5.44,⁷⁷ 2.05,⁴² and 1.25¹⁵ (the latter value excludes ethylbenzene).

The free energies of binding computed with the OPLS-AA parameters for the seven analogs are also compared graphically with the experimental values in Figure 5A. As expected from the MUEs, the correlation is good, with an R^2 of 0.81. The principal outlier is now benzene itself, and including it lowers the R^2 to 0.66. The R^2 value of 0.20 for this ligand set with 1.14*CM1A charges is significantly poorer. For comparison, the R^2 values published in the literature for this set of eight ligands are 0.27,⁴¹ 0.02,⁷⁷ 0.54,⁴² and 0.06¹⁵ (the latter value excludes ethylbenzene). As it was the case with benzene, shown in Figure 4A, the evolution of the total binding free energies with number of MC steps shows that the results are generally well converged within 30M configurations (see Supplementary Figure S5). The unsigned errors, plotted in Figure 5B, show longer-lived variances of individual values, but after 60M MC steps, they become essentially constant for all molecules except the two largest ones, indole and indene. Interestingly, there is a partial cancellation between these fluctuations, such that an almost constant average MUE of *ca.* 1 kcal/mol is obtained from the very first 10M configurations to the final value at 180M steps.

CONCLUSION

Significant improvements to MC protocols have been introduced to better explore protein-ligand conformational space including receptor and ligand flexibility. The sampling parameters determined in this study for the protein backbone and side chain motions generate competitive sampling of conformational space for MC and MD. For the CRA algorithm, reasonable changes to the three controlling parameters did not alter significantly the extent of sampled conformational space, while side-chain sampling did show substantial dependence on the size of attempted moves. For all residues, the ranges for dihedral changes had to be larger than 120° to provide higher probabilities of moving between local minima. As CRA, pivot and side chain moves are purely geometric algorithms, the results obtained from this work can be generalized for any protein system.

The calculation of absolute binding free energies is a challenging problem due to the size of the conformational space that needs to be sampled. With FEP, such calculations require the total annihilation of the ligand, which often leads to LJ-energy instabilities produced by atomic overlaps. Here, we have established a successful MC/FEP procedure to overcome these problems combining a Coulomb and LJ decoupling scheme, 1-1-6 soft core potentials, and a hard wall constraint for the ligand. The MC/FEP approach and MD produced results within 0.6 kcal/mol for the lysozyme-benzene complex. The computed binding free energies were shown to be well converged after 30M MC steps per λ -window. In comparison, MD required at least 3 ns per λ -window to achieve convergence of the free energy, requiring 5 –

10 times more computation time. The successful application of this MC/FEP approach to seven benzene analogs showed the robustness of the protocol for estimation of absolute binding free energies. The average MUE of 1.23 kcal/mol with respect to experimental data compares very favourably with prior results from alternative MD^{41,42} or MC^{15,77} studies. Current efforts are being directed to application to other proteins and more complex ligands.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

Gratitude is expressed to the National Institutes of Health (GM32136) and to the Yale University Faculty of Arts and Sciences High Performance Computing Center for support.

REFERENCES

1. Zhou R Trp-cage: folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. U. S. A* 2003, 100, 13280–13285. [PubMed: 14581616]
2. Jorgensen WL; Nguyen TB Monte Carlo simulations of the hydration of substituted benzenes with OPLS potential functions. *J. Comput. Chem* 1993, 14, 195–205.
3. Wong CF; McCammon JA Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc* 1986, 108, 3830–3832.
4. Essex JW; Severance DL; Tirado-Rives J; Jorgensen WL Monte Carlo simulations for proteins: Binding affinities for trypsin–benzamidine complexes via free-energy perturbations. *J. Phys. Chem. B* 1997, 101, 9663–9669.
5. Berens PH; Mackay DH; White GM; Wilson KR Thermodynamics and quantum corrections from molecular dynamics for liquid water. *J. Chem. Phys* 1983, 79, 2375–2389.
6. Gilson MK; Given JA; Bush BL; McCammon JA The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J* 1997, 72, 1047–1069. [PubMed: 9138555]
7. Zwanzig RW High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys* 1954, 22, 1420–1426
8. Jorgensen WL; Ravimohan C Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys* 1985, 83, 3050–3054.
9. Bennett CH Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys* 1976, 22, 245–268.
10. Chodera JD; Swope WC; Pitera JW; Seok C; Dill KA Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput* 2007, 3, 26–41. [PubMed: 26627148]
11. Mobley DL; Klimovich PV Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys* 2012, 137, 230901. [PubMed: 23267463]
12. Torrie GM; Valleau JP Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys* 1977, 23, 187–199.
13. Hansmann UH Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett* 1997, 281, 140–150.
14. Sugita Y; Okamoto Y Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett* 1999, 314, 141–151.
15. Clark M; Meshkat S; Wiseman JS Grand Canonical Free-Energy Calculations of Protein–Ligand Binding. *J. Chem. Inf. Model* 2009, 49, 934–943. [PubMed: 19309088]

16. Laio A; Parrinello M Escaping free-energy minima. *Proc. Natl. Acad. Sci U. S. A* 2002, 99, 12562–12566. [PubMed: 12271136]
17. Abraham MJ; Murtola T; Schulz R; Pall S; Smith JC; Hess B; Lindahl E GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015, 1–2, 19–25.
18. Case DA; Cheatham TE, 3rd; Darden T; Gohlke H; Luo R; Merz KM, Jr.; Onufriev A; Simmerling C; Wang B; Woods RJ The Amber biomolecular simulation programs. *J. Comput. Chem* 2005, 26, 1668–88. [PubMed: 16200636]
19. Phillips JC; Braun R; Wang W; Gumbart J; Tajkhorshid E; Villa E; Chipot C; Skeel RD; Kale L; Schulten K Scalable molecular dynamics with NAMD. *J. Comput. Chem* 2005, 26, 1781–802. [PubMed: 16222654]
20. Brooks BR; Brooks CL; MacKerell AD; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S; Caflisch A; Caves L; Cui Q; Dinner AR; Feig M; Fischer S; Gao J; Hodoscek M; Im W; Kuczera K; Lazaridis T; Ma J; Ovchinnikov V; Paci E; Pastor RW; Post CB; Pu JZ; Schaefer M; Tidor B; Venable RM; Woodcock HL; Wu X; Yang W; York DM; Karplus M CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem* 2009, 30, 1545–1614. [PubMed: 19444816]
21. Eastman P; Friedrichs MS; Chodera JD; Radmer RJ; Bruns CM; Ku JP; Beauchamp KA; Lane TJ; Wang LP; Shukla D; Tye T; Houston M; Stich T; Klein C; Shirts MR; Pande VS OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput* 2013, 9, 461–469. [PubMed: 23316124]
22. Jorgensen WL; Tirado-Rives J Monte Carlo vs Molecular Dynamics for Conformational Sampling. *J. Phys. Chem* 1996, 100, 14508–14513.
23. Jorgensen WL; Tirado-Rives J Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem* 2005, 26, 1689–700. [PubMed: 16200637]
24. Woods CJMJ; Bodnarchuk M; Genheden S; Bradshaw R; Ross GA; Cave-Ayland C; Bruce-Macdonald H; Cabedo Martinez AI; Graham J Protoms 3.3 <http://protoms.org/> (accessed November 15, 2017).
25. Woods C; Michel J, Sire: An advanced, multiscale, molecular simulation framework <http://siremol.org/> (accessed November 15, 2017).
26. Borrelli KW; Vitalis A; Alcantara R; Guallar V PELE: protein energy landscape exploration. A novel Monte Carlo based technique. *J. Chem. Theory Comput* 2005, 1, 1304–1311. [PubMed: 26631674]
27. Madadkar-Sobhani A; Guallar V PELE web server: atomistic study of biomolecular systems at your fingertips. *Nucleic Acids Res* 2013, 41, W322–W328. [PubMed: 23729469]
28. Noé F; Fischer S Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol* 2008, 18, 154–162. [PubMed: 18378442]
29. Takahashi R; Gil VA; Guallar V Monte Carlo free ligand diffusion with Markov state model analysis and absolute binding free energy calculations. *J. Chem. Theory Comput* 2013, 10, 282–288. [PubMed: 26579911]
30. Cabeza de Vaca I; Lucas MF; Guallar V New Monte Carlo based technique to study DNA–ligand interactions. *J. Chem. Theory Comput* 2015, 11, 5598–5605. [PubMed: 26642982]
31. Robertson MJ; Tirado-Rives J; Jorgensen WL Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput* 2015, 11, 3499–3509. [PubMed: 26190950]
32. Robertson MJ; Tirado-Rives J; Jorgensen WL Improved treatment of nucleosides and nucleotides in the OPLS-AA force field. *Chem. Phys. Lett* 2017, 683, 276–280. [PubMed: 29479109]
33. Dodda LS; Vilseck JZ; Tirado-Rives J; Jorgensen WL 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* 2017, 121, 3864–3870. [PubMed: 28224794]
34. Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput* 2015, 11, 3696–3713. [PubMed: 26574453]

35. Harder E; Damm W; Maple J; Wu C; Reboul M; Xiang JY; Wang L; Lupyan D; Dahlgren MK; Knight JL; Kaus JW; Cerutti DS; Krilov G; Jorgensen WL; Abel R; Friesner RA OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput* 2016, 12, 281–296. [PubMed: 26584231]
36. Best RB; Zhu X; Shim J; Lopes PE; Mittal J; Feig M; Mackerell AD, Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput* 2012, 8, 3257–3273. [PubMed: 23341755]
37. Dodda LS; Vilseck JZ; Cutrona KJ; Jorgensen WL Evaluation of CM5 Charges for Nonaqueous Condensed-Phase Modeling. *J. Chem. Theory Comput* 2015, 11, 4273–82. [PubMed: 26575922]
38. Eriksson A; Baase W A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* 1992, 355, 371. [PubMed: 1731252]
39. Morton A; Matthews BW Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* 1995, 34, 8576–8588. [PubMed: 7612599]
40. Morton A; Baase WA; Matthews BW Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* 1995, 34, 8564–8575. [PubMed: 7612598]
41. Deng Y; Roux B Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput* 2006, 2, 1255–1273. [PubMed: 26626834]
42. Mobley DL; Graves AP; Chodera JD; McReynolds AC; Shoichet BK; Dill KA Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol* 2007, 371, 1118–1134. [PubMed: 17599350]
43. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys* 1983, 79, 926–935.
44. Ulmschneider JP; Jorgensen WL Monte Carlo backbone sampling for polypeptides with variable bond angles and dihedral angles using concerted rotations and a Gaussian bias. *J. Chem. Phys* 2003, 118, 4261–4271.
45. Ulmschneider JP; Jorgensen WL Polypeptide folding using Monte Carlo sampling, concerted rotation, and continuum solvation. *J. Am. Chem. Soc* 2004, 126, 1849–1857. [PubMed: 14871118]
46. Vitalis A; Pappu RV Methods for Monte Carlo Simulations of Biomacromolecules *Annu. Rep. Comput. Chem* 2009, 5, 49–76. [PubMed: 20428473]
47. Robertson MJ; Tirado-Rives J; Jorgensen WL Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput* 2015, 11, 3499–509. [PubMed: 26190950]
48. Davis IW; Arendall WB; Richardson DC; Richardson JS The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* 2006, 14, 265–274. [PubMed: 16472746]
49. Betancourt MR Efficient Monte Carlo trial moves for polypeptide simulations. *J. Chem. Phys* 2005, 123, 174905. [PubMed: 16375567]
50. Ryckaert J-P; Ciccotti G; Berendsen HJ Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys* 1977, 23, 327–341.
51. Kolafa J; Perram JW Cutoff errors in the Ewald summation formulae for point charge systems. *Mol. Simul* 1992, 9, 351–368.
52. Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A smooth particle mesh Ewald method. *J. Chem. Phys* 1995, 103, 8577–8593.
53. Darden T; York D; Pedersen L Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys* 1993, 98, 10089–10092.
54. Toda M; Kubo R; Saito N; Hashitsume N *Statistical Physics II: Nonequilibrium Statistical Mechanics*, Second Edition; Springer Series in Solid-State Sciences, Vol. 31; Springer-Verlag: Heidelberg, 1991; pp 14–21.
55. Brunger A; Brooks CL; Karplus M Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett* 1984, 105, 495–500.

56. Dodda LS; Cabeza de Vaca I; Tirado-Rives J; Jorgensen WL LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res* 2017, 45, W331–W336. [PubMed: 28444340]
57. Jorgensen WL; Thomas LL Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput* 2008, 4, 869–876. [PubMed: 19936324]
58. Lu N; Kofke DA; Woolf TB Improving the efficiency and reliability of free energy perturbation calculations using overlap sampling methods. *J. Comput. Chem* 2004, 25, 28–40. [PubMed: 14634991]
59. Steinbrecher T; Mobley DL; Case DA Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys* 2007, 127, 214108. [PubMed: 18067350]
60. Hermans J; Wang L Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc* 1997, 119, 2707–2714.
61. Woo H-J; Roux B Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A* 2005, 102, 6825–6830. [PubMed: 15867154]
62. Wang J; Deng Y; Roux B Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J* 2006, 91, 2798–2814. [PubMed: 16844742]
63. Wei BQ; Baase WA; Weaver LH; Matthews BW; Shoichet BK, A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol* 2002, 322, 339–355. [PubMed: 12217695]
64. Wei BQ; Weaver LH; Ferrari AM; Matthews BW; Shoichet BK Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol* 2004, 337, 1161–1182. [PubMed: 15046985]
65. Mobley DL; Chodera JD; Dill KA On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J. Chem. Phys* 2006, 125, 084902. [PubMed: 16965052]
66. Mobley DL; Chodera JD; Dill KA Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput* 2007, 3, 1231–1235. [PubMed: 18843379]
67. Jiang W; Roux B Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput* 2010, 6, 2559–2565. [PubMed: 21857813]
68. Perez A; Morrone JA; Simmerling C; Dill KA Advances in free-energy-based simulations of protein folding and ligand binding. *Curr. Opin. Struct. Biol* 2016, 36, 25–31. [PubMed: 26773233]
69. Gallicchio E; Lapelosa M; Levy RM Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein–Ligand Binding Affinities. *J. Chem. Theory Comput* 2010, 6, 2961–2977. [PubMed: 21116484]
70. Wang K; Chodera JD; Yang Y; Shirts MR Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput.-Aided Mol. Des* 2013, 27, 989–1007. [PubMed: 24297454]
71. Jo S; Jiang W; Lee HS; Roux B; Im W CHARMM-GUI Ligand Binder for absolute binding free energy calculations and its application. *J. Chem. Inf. Model* 2013, 53, 267–277. [PubMed: 23205773]
72. Rodinger T; Howell PL; Pomès R, Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J. Chem. Phys* 2008, 129, 155102. [PubMed: 19045232]
73. Yan XC; Tirado-Rives J; Jorgensen WL Hydration Properties and Solvent Effects for All-Atom Solutes in Polarizable Coarse-Grained Water. *J. Phys. Chem. B* 2016, 120, 8102–8114. [PubMed: 26901452]
74. Jorgensen WL; Maxwell DS; Tirado-Rives J Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc* 1996, 118, 11225–11236.
75. Jorgensen WL; Ulmschneider JP; Tirado-Rives J Free Energies of Hydration from a Generalized Born Model and an All-Atom Force Field. *J. Phys. Chem. B* 2004, 108, 16264–16270.
76. Wolfenden R; Andersson L; Cullis P; Southgate C Affinities of amino acid side chains for solvent water. *Biochemistry* 1981, 20, 849–855. [PubMed: 7213619]

77. Clark MW; Jeffrey S Free Energy Computation by Monte Carlo Integration 2017, ArXiv:physics.chem-ph/1701.05073. ArXiv.org e-Print archive, <https://arxiv.org/ftp/arxiv/papers/1701/1701.05073.pdf> (accessed November 15, 2017).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. For, a backbone pivot move, the atoms between two consecutive C^α s are rotated about the axis connecting them.

Figure 2. Lysozyme L99A protein conformational analysis. (A) Ramachandran plot of the 100-ns MD (green dots) and 50M MC (blue dots) trajectories for Valine 111. (B) C α rmsF per residue for 20 ns MD (blue) and 50M MC (yellow) runs and the difference (red, MD-MC). The dashed lines represent the average rmsF values using the same color codes. (C) and (D) backbone representations over 50 equally distributed frames of 20 ns MD and 50M MC, respectively. The red, white, and blue traces are the initial, intermediate, and final frames.

Figure 3.

Panels A, B, and C show the Met dipeptide χ_1 , χ_2 and χ_3 distributions for 10-ns MD and 500M MC steps. Optimal ranges for the χ moves were 210, 120 and 180 degrees for χ_1 , χ_2 and χ_3 . Blue and green bars show the MC and MD distributions, respectively. The darker blue shade indicates full overlap between MD and MC. Same color codes are used for the average distribution values represented by dashed lines. Red dashed line is the initial χ value.

Figure 4.
Evolution of the absolute free energy of binding with simulation length for lysozyme–benzene. (A) MC/FEP. (B) MD/FEP.

Figure 5. Analysis for the seven benzene analogs. (A) Computed vs. experimental absolute binding free energies. (B) MUE evolution with respect to the number of MC steps.

Table 1.

Optimal Dihedral Ranges (°) for Side-Chain Moves

	χ_1	χ_2	χ_3	χ_4
Thr	150			
Cys	240			
Ser	120			
Val	120			
Trp	330	210		
Tyr	360	360		
His	300	150		
Asp	300	180		
Phe	180	330		
Leu	300	150		
Ile	120	120		
Pro	210	120		
Asn	360	120		
Glu	300	120	180	
Gln	120	180	360	
Met	210	120	180	
Lys	240	360	300	360
Arg	270	300	270	270

Table 2.

Computed and Experimental Absolute Binding Free Energies (kcal/mol) for Benzene Using Different Algorithms.

	backbone DOF	G_{bind}
MD	N,C $^{\alpha}$,C constrained	-9.93 ± 0.15^a
MD	C $^{\alpha}$ constrained	-9.45 ± 0.33^a
MD	free	-5.85 ± 0.19
MC	rigid	-9.97 ± 0.42
MC	CRA	-6.82 ± 0.20
MC	CRA+pivot	-6.44 ± 0.22
Experimental		-5.19 ± 0.16

^aUncertainty calculated from a single simulation.

Table 3.

Computed and Experimental Absolute Binding Free Energies (kcal/mol) for Benzene and Seven Analogs with L99A Lysozyme.

	$G_{\text{bind, CMIA}}$	$G_{\text{bind, OPLS}}$	$G_{\text{bind, exptl}}^{40}$	UE_{CMIA}	UE_{OPLS}
Benzene	-6.44 ± 0.22	-7.68 ± 0.22	-5.19 ± 0.16	1.25	2.49
Toluene	-5.76 ± 0.24	-6.60 ± 0.24	-5.52 ± 0.04	0.24	1.08
<i>o</i>-Xylene	-1.36 ± 0.35	-2.90 ± 0.38	-4.6 ± 0.06	3.24	1.70
<i>p</i>-Xylene	-4.11 ± 0.30	-4.98 ± 0.30	-4.67 ± 0.06	0.56	0.31
Ethylbenzene	-3.04 ± 0.27	-6.95 ± 0.25	-5.76 ± 0.07	2.72	1.19
Benzofuran	-4.57 ± 0.32	-7.21 ± 0.26	-5.46 ± 0.03	0.89	1.75
Indene	-2.96 ± 0.37	-5.87 ± 0.40	-5.13 ± 0.01	2.17	0.74
Indole	1.29 ± 0.30	-4.35 ± 0.32	-4.89 ± 0.06	6.18	0.54
MUE				2.16	1.23