JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

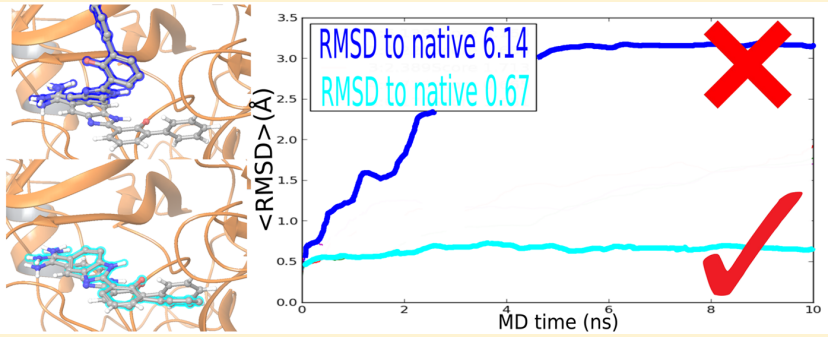
Prediction of Protein−Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations

Anthony J. Clark,[†] Pratyush Tiwary,[*,†] Ken Borrelli,[‡] Shulu Feng,[‡] Edward B. Miller,[‡] Robert Abel,[‡] Richard A. Friesner,[†] and B. J. Berne[†]

[†]Department of Chemistry, Columbia University, New York, New York 10027, United States

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

Supporting Information



ABSTRACT: Ligand docking is a widely used tool for lead discovery and binding mode prediction based drug discovery. The greatest challenges in docking occur when the receptor significantly reorganizes upon small molecule binding, thereby requiring an induced fit docking (IFD) approach in which the receptor is allowed to move in order to bind to the ligand optimally. IFD methods have had some success but suffer from a lack of reliability. Complementing IFD with all-atom molecular dynamics (MD) is a straightforward solution in principle but not in practice due to the severe time scale limitations of MD. Here we introduce a metadynamics plus IFD strategy for accurate and reliable prediction of the structures of protein−ligand complexes at a practically useful computational cost. Our strategy allows treating this problem in full atomistic detail and in a computationally efficient manner and enhances the predictive power of IFD methods. We significantly increase the accuracy of the underlying IFD protocol across a large data set comprising 42 different ligand−receptor systems. We expect this approach to be of significant value in computationally driven drug design.

INTRODUCTION

The problem of accurately and reliably predicting the structure of protein−ligand complexes is a key challenge in computationally driven structure-based drug discovery projects. There are many situations in which a novel, interesting active compound is identified in a screening campaign, but a crystal structure of the compound with the receptor is not available (and may prove difficult to obtain, for various reasons). In lead optimization efforts, modifications to the compound can result in significant and unexpected changes in the binding mode, which can cause problems in understanding structure−activity relationships (SAR) until a high-resolution structure is obtained. Finally, binding mode prediction has obvious utility in the use of computational methods to facilitate the design of new molecules which may bind more advantageously.

In a fraction of cases, binding mode prediction using rigid receptor docking (RRD) (e.g., through the widely used Glide program[1]) is successful, employing a receptor structure obtained from cocrystallization with a ligand differing from the one of interest. Computational experiments[1,2] suggest that a good pose (roughly speaking, one with a root mean squared deviation (RMSD) from experiment of <2 Å) is obtained about 50% of the time when this approach is employed. However, the most critical cases are those in which the effects of induced fit on the receptor play a significant role. Rigid receptor docking cannot address such effects, since the receptor is not allowed to move in response to the presence of the ligand. When induced fit effects exceed some key threshold in magnitude, one typically observes steric clashes of the new ligand with the available receptor conformation, resulting in a failure of RRD to predict the correct binding mode. In some cases, protein side chain movements are also required to make key binding contacts with the ligand.

Induced fit docking (IFD) protocols have been developed to address these cases, in which successful docking requires nontrivial receptor movements. This is a challenging problem which has been addressed by a number of research groups,

Received: February 24, 2016
Published: May 4, 2016

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

using quite different approaches in each case.[3−5] The methodology that we have developed, described in several previous publications,[6,7] combines Glide docking with receptor conformational search using a continuum solvent based molecular mechanics model, available in the Prime protein modeling package. As long as there are no large changes in the protein structure (e.g., huge loop movements, such as a change from a DFG-in to DFG-out conformation in a kinase activation loop), this IFD methodology will usually locate a structure with good accuracy within the top 5−10 results as ranked by a combination of Prime and Glide energetics. However, it has proven difficult to robustly ascertain the correct structure from among this set of many possibilities. The uncertainty associated with this final step significantly reduces the utility of Glide/Prime IFD calculations in actual projects, where actionable drug discovery efforts such as synthetic prioritization require a high confidence in the predicted binding mode of the ligand. The energy functions in both Glide and Prime contain uncontrolled heuristic approximations, and even the IFD sampling protocol has significant limitations (e.g., very limited backbone movement). For these reasons, it has been very challenging to select the correct pose from an IFD calculation by improving the scoring function that can be constructed with these technologies. A better scoring function, and a higher resolution simulation, appears to be required.

If greater accuracy and more complete sampling is desired, an obvious approach is to use all atom, explicit solvent molecular dynamics (MD) simulations to discriminate between the relatively small number of top ranked poses produced by an IFD calculation. However, it is far from clear how this task can be accomplished. A straightforward attempt to equilibrate populations between the different IFD structures will fail in most cases, because there are typically large ($\gg k_B T$) free energy barriers between the alternative poses. Another possible approach is to investigate pose stability via unbiased molecular dynamics; however, as we will discuss further below, incorrect structures are in many cases metastable and cannot be readily distinguished from the correct binding pose by a short MD simulation. Very long (microsecond or more) MD simulations on specialized computers have been shown to be successful in predicting the binding pose in a few cases;[8] the computational effort for simulations of this type is more than 1000 times the cost of an IFD simulation, thus limiting this approach to extremely low throughput.

An alternative MD-based strategy is to devise an algorithm based on enhanced sampling methods, which allow efficient back-and-forth movement across large free energy barriers, and therefore sampling of the relative stability of different poses with practical computing resources, while still maintaining full atomistic resolution. Though there are a number of such methods available, in this work we choose metadynamics,[9−13] since it requires minimal preknowledge of the system being sampled, apart from the choice of collective variables (CV) for biasing. These are slowly varying order parameters that are expected to play a role in the motion of a metastable pose out of its initial configuration. The key idea in the use of metadynamics is to build a time-dependent bias as a function of the chosen CV that samples the ligand movements in and around its binding pose. As the correct IFD pose has (to the extent that it reproduces key features of the native structure) a stronger binding affinity, a properly calibrated metadynamics bias should result in preferred displacement of incorrect poses, while leaving the correct pose relatively stable. The end result of a typical metadynamics simulation is an estimate of the underlying Boltzmann probability distribution or equivalently the potential of mean force, also called free energy.[14] Previous applications[15,16] of metadynamics to complement docking have shown encouraging results, though evidently not in a fashion amenable to processing a large number of calculations, so calculating the full free energy surface is not our aim here. Instead our aim is to use a metadynamics approach to determine only the relative stabilities of different binding poses.

One can anticipate that care is going to be required in defining the CV and other biasing parameters in the metadynamics protocol. While these requirements are typical for effective use of metadynamics to address most interesting problems in complex molecular systems,[17] for the objective of this work a balance must be struck between (i) a careful refining of the choice of metadynamics parameters and (ii) the need for a protocol that can be automated for high-throughput screening of proposed binding modes. Therefore, our choice of CV must be generic enough that it needs minimal system-dependent fine-tuning. The real question then is whether a robust, automated enhanced sampling MD-based protocol can be defined that can be used in practical applications in drug discovery projects.

In the present paper, we investigate a large suite of realistic IFD test cases, comprising 42 different protein−ligand complexes involving targets of pharmacological relevance, and develop a standardized high throughput metadynamics-based all-atom molecular dynamics method for carrying out the task outlined above. The full set of initial poses tested is provided in the Supporting Information. A surprisingly high rate of success is achieved in selecting low RMSD poses from those generated by initial IFD calculations. Given the success and widespread popularity of IFD, we believe that the metadynamics-IFD coupled approach reported here represents a significant step forward in actionable screening of target ligands through computer simulation.

The paper is organized as follows. In the section Methods, we review our IFD methodology and then discuss the optimized metadynamics algorithm that we use to select the best structure from available IFD poses. In the section Data Set, we introduce the data set assembled from Protein Data Bank (PDB) structures used to test the methodology, and in Results we present the results. In the Discussion, we analyze the results in more depth considering their implications. Finally, in the Conclusions, we summarize our findings and outline future directions.

METHODS

The methodology developed and demonstrated in this work has been summarized in Figure 1. We now provide descriptions of the various components of the overall process.

Review of IFD Methodology. Our IFD methodology has been described in detail in a number of previous publications.[6,7] The IFD protocol involves the use of both the Glide docking program and the Prime protein structure modeling code. In broad outline, Glide initially produces possible ligand poses, each of which is refined by Prime using a continuum solvation based molecular mechanics model, via rotamer-based library optimizations of the protein side chain conformations. This process is iterated to produce a list of final poses, rank ordered by a combination of the Prime energy (molecular mechanics force field in Variable-dielectric Surface Generalized Born (VSGB) solvent) and the Glide SP empirical scoring function.

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

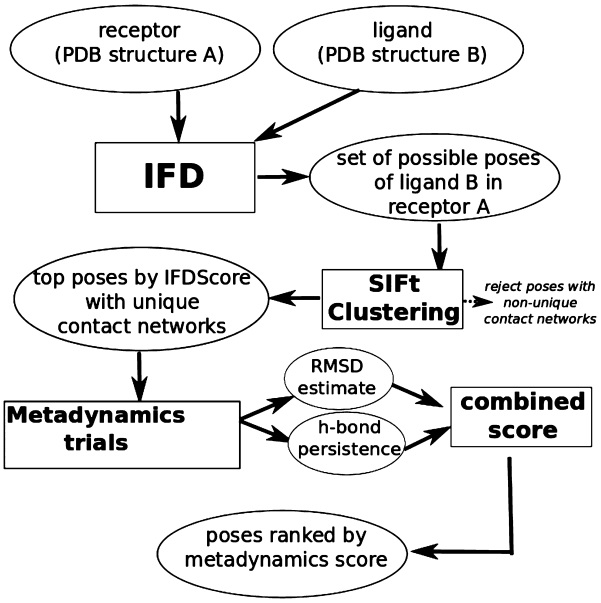


Figure 1. Flowchart summarizing the various components of the structure prediction process described in this work. See the main text for details for the various steps.

The initial docking step requires modification of the receptor model by softening the force field, as in an IFD calculation one has to be able to deal with a situation in which the target ligand will not fit into the available receptor conformation or conformations, due to steric clashes. Additionally, in the current IFD protocol, initial docking is performed in receptor structures in which one or more side chains are mutated to alanine. The list of side chains to mutate is determined by combining information from various sources, e.g., X-ray B factors of the side chain atoms. In the experience of the authors, the present IFD algorithm succeeds in generating a reasonable initial pose somewhere in the top ranked structures in a large majority of cases. Improvements can likely be made in the details of how this initial step is carried out, increasing this success rate to 95% or more, and we are currently investigating promising approaches to making such improvements. However, we will not discuss the initial docking further in the present paper, which is focused exclusively on cases where one or more good initial poses are in fact generated.

A subset of poses are extracted from the initial Glide docking runs and passed on to Prime. All mutated side chains are restored to their original identity, and extensive sampling from a high resolution library is used to search for possible side chain conformations. The OPLS force field, and the VSGB continuum solvent model, described in ref 18, are employed to evaluate the free energies of the various conformations of the complex that are generated. Promising configurations are minimized with the OPLS force field, and the resulting energies constitute one component of the scoring function. Finally, the ligand is redocked into each of the new structures and reminimized, and the final poses are rank ordered by the scoring function given by a linear combination of the Prime energy and GlideScore referred to as the IFDScore.[7]

The total single processor CPU time for a typical IFD job is on the order of 10−20 CPU hours. The algorithm can easily be parallelized to enable the different poses to run on different processors, so the wall clock time using 20 processors is on the order of 1 h. Thus, an IFD calculation can be completed on a time scale compatible with effective use in practical drug discovery projects. Our objective here is to improve the accuracy with which the lowest RMSD pose may be selected from the list of returned IFD structures without increasing the computational cost of the IFD protocol beyond what would be tolerable in a drug discovery campaign. The IFD algorithm is tested by extracting two complexes A and B from the PDB (or proprietary data sets) for a given receptor, each of which binds a different ligand, and such that cross docking of ligand A into receptor conformation B via rigid receptor docking (e.g., via Glide XP or a similar program) fails to yield a pose that is accurate compared to experiment (<2 Å). We currently maintain a standard test set of 100 test cases of this type taken from the PDB, which have been selected to also exclude large changes in receptor structure (e.g., large loop movements such as DFG-in to DFG-out for a kinase activation loop) so that the IFD protocol, which is primarily focused on side chain reorganization, has a chance to produce an appropriate protein conformation. Of these, 42 cases were selected for metadynamics investigation, based on having at least one pose in the top 5 scoring distinct poses by IFDscore within 2.0 Å root-mean-square displacement (RMSD) from the native structure and at least one pose greater than 2.0 Å RMSD from the native (calculated by aligning the receptor in the docked structure with the crystal structure), with some additional triaging of very large receptors to save CPU time. Versions of IFD that include some ability to modify loop conformations, and other backbone structures, are currently under development.

Metadynamics Theory and Methodology. Metadynamics is a widely used enhanced sampling method that allows sampling of complex free energy landscapes.[10−13] By building a time-dependent bias as a function of carefully chosen slowly varying order parameters, called collective variables (CVs), the system is coaxed to escape stable free energy minima where it would normally be trapped. The end result of a metadynamics simulation when fully converged, which is not the aim in this work, is the underlying unbiased probability distribution (or equivalently the free energy), either directly as a function of the biased CV or through a reweighting procedure[14] as a function of any generic observable. The bias $V(s, t)$ is typically constructed in the form of periodically added repulsive Gaussians, where s is the chosen CV which could be multidimensional. At any time t, the free energy $F(s)$ can be obtained from the deposited bias $V(s, t)$ as per the following equation:[11]

$$F(s) = -\frac{T + \Delta T}{T} V(s, t) + C(t) \quad (1)$$

Here T is the simulation temperature (300 K), ΔT is the tempering factor through which the amplitude of the bias deposited at a point s in the collective variable space is tuned down with the number of times the system visits a given point, and $C(t)$ is a time-dependent constant which is irrelevant for the present work[14] since we are interested only in normalized probability distributions resulting from the free energy. Other numerical parameters include the initial Gaussian hill height h, bias deposition interval τ, and Gaussian width w. While a generic metadynamics simulation shows some sensitivity to the various parameters mentioned above, the most critical is the choice of the biasing CV itself. The central requirements for a CV to be useful for practical drug discovery applications are

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

that it should be (1) correlated to the pose stability and (2) require minimal fine-tuning from system to system. With these two requirements in mind, we define the biasing CV s as the RMSD deviation from the starting pose (after equilibration in short unbiased MD), measured over the atoms in a subset K of the ligand−receptor system. This subset K comprises all heavy atoms in the ligand, plus a few protein backbone atoms taken from a well-conserved part of the receptor (see Figure 2). We describe further details of this in the following subsection Choice of Collective Variable.

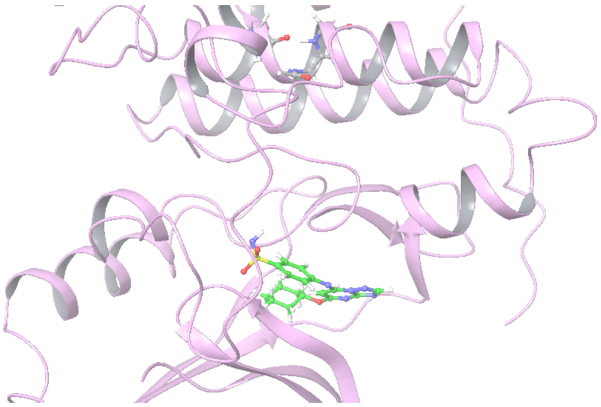


Figure 2. Ligand and anchor atoms are shown for the example case of the ligand from PDB structure 2c6m docked with the cdk2 receptor from PDB structure 1pxj.

At this point, we introduce a simple but key methodological insight. As known through multiple previous studies,[15,16] converging the full free energy landscape while feasible can still be very time-consuming and will probably require system-specific selection of CVs. Instead, in this work we perform several independent short metadynamics simulations, through which we calculate the average or thermodynamically most favored RMSD deviation from the starting pose, calculated for the subset K of the ligand−receptor system. Physically, this is equivalent to doing a much longer unbiased MD run where the starting pose can surmount local barriers and relax into a stable structure. A higher estimate of the average or thermodynamically preferred RMSD can then be considered a proxy for poor stability of the pose initially generated in docking. The basic idea is that the most stable pose is the best pose. This most likely RMSD estimate is calculated as follows:

$$\langle s \rangle = \frac{\int \mathrm{d}s\ s\ \mathrm{e}^{-(F(s)/k_\mathrm{B}T)}}{\int \mathrm{d}s\ \mathrm{e}^{-(F(s)/k_\mathrm{B}T)}} \tag{2}$$

after averaging over independent runs monitored for convergence, where $F(s)$ is the free energy as a function of the CV (see eq 1) and k_B is Boltzmann's constant. In the examples that follow, we will carefully examine the construction of the set K as well as the choice of other parameters.

Choice of Collective Variable. As stated above, the choice of CV is by far the most important requirement for attaining efficient sampling of relevant states through metadynamics. Here we are faced with the additional challenge of constructing a CV that can be straightforwardly applied across a wide variety of ligands and receptors in a manner suitable for high throughput.

Depending on the nature of the specific compound involved, an unstable binding pose can relax from its starting configuration by various mechanisms, including internal deformation and rigid rotation relative to the binding pocket. A broadly applicable CV should be capable of biasing both intraligand conformations and rigid rotations but minimize the biasing of the center of mass displacement of the ligand from the receptor pocket, as strong biasing of the center of mass displacement of the ligand from the binding pocket can lead to the ligand fully leaving the pocket. If the ligand exits the pocket, then either very long simulation times[8] or carefully designed restraints[19] will be needed to facilitate re-entry.

On the basis of these requirements, we choose as our collective variable the aligned RMSD of the ligand heavy atoms plus the set of heavy backbone atoms of a stable region of the receptor away from the binding pocket (see representative illustration in Figure 2). The aligned RMSD is calculated relative to the equilibrated input structure. We term the set of heavy backbone atoms included in the RMSD calculation as the "anchor" atoms. Through this construction we allow both internal and rigid rotational modes to be biased. In order to minimize the biasing of the center of mass displacement mode, we choose the anchor in a fashion that minimizes its alignment with respect to the outward pointing binding pocket axis. This selection criteria is straightforward to automate as part of a high-throughput protocol. For each receptor−ligand case an anchor region is determined from these criteria using the highest ranked IFD pose, and the same anchor was used for all 5 IFD poses. We note that there are likely a variety of CV choices that could be made that would provide biasing suitable for pose stability. One such alternative CV found to give comparable results is discussed in detail in the Supporting Information, where a subset of the receptor atoms is first aligned with the starting structure at each frame and the RMSD is calculated over the ligand atoms from this alignment.

The starting hill height h is another parameter to which we find sensitivity in the simulations. In most cases, a large hill height, 0.3 kcal/mol (around $k_\mathrm{B}T/2$) is found to provide the best discrimination between poses. In a few cases this may prove to be too strong and cause even native-like poses to exit the binding pocket in the short metadynamics runs. Such cases can be diagnosed as having all RMSD estimate values above a cutoff threshold of 2.0 Å. For the small number of compounds where all candidate poses are above this threshold, a rerun of the metadynamics trials is performed with a smaller hill height of 0.05 kcal/mol. The results from this weak rerun replace the strong run for evaluating the poses. Other metadynamics parameters were found to not have a very pronounced effect on the protocol, and we report their suggested values in the Supporting Information. In the Supporting Information, we also report a different construction of the CV that gives similar overall results. All MD and metadynamics runs were performed with the OPLS force-field,[18] see the Supporting Information for further simulation parameters.

Hydrogen Bond Persistence Score. In a small number of cases, non-native-like poses may not be able to fully leave their initial metastable configuration in the short simulations and therefore may not separate from the native-like pose(s) in a very pronounced manner, if estimated solely by the RMSD score. However, even in these cases, metadynamics itself is able to weaken the contact network between the ligand and receptor substantially for the non-native-like pose, even when the ligand may not leave its initial pose. Thus, postprocessing assessment

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

of the persistence of the starting contact network can provide a useful complementary metric that can predict instability of poses in cases where the RMSD estimate has not fully captured information about the instability.

In the data set included here, hydrogen bonds are by far the dominant contacts in the binding modes of the ligand−receptor systems. We therefore propose a simple hydrogen bond persistence fraction metric with which to assess the stability of the ligand contact network. The hydrogen bond persistence (HBP) is defined as the fraction of the hydrogen bonds found in the initial input pose, averaged over the last 2 ns of all 10 trials for each pose. Hydrogen bonds are identified by the Maestro criteria (maximum distance 2.8 Å, donor angle 120°, and acceptor angle 90°). For cases where other types of interactions dominate, this can be readily extended to include hydrophobic contacts, $\pi-\pi$ interactions etc. and will be done in future work. In the Supporting Information, we provide detailed results demonstrating that the hydrogen bond persistence patterns are converged with respect to the number of trials we use in this work.

Metadynamics-IFD Combined Scoring Function. For a final screening, we combine the RMSD stability and hydrogen bond persistence measures, obtained from the metadynamics trajectories into a scoring function whose parameters have been optimized using the data from the 42 test cases discussed in the Results section. This function is defined as follows.

$$S = \frac{\sigma_h(D)}{D} + 2\sigma_l(H)H \tag{3}$$

where D is the RMSD estimate average from eq 2, H is the hydrogen bond persistence score, and $\sigma_h(D) = 1/[\exp(10(D - 1.7))]$ and $\sigma_l(H) = 1/[\exp(-50(H - 0.3))]$ are cutoff functions that reduce the weight of RMSD values exceeding 2.0 Å and hydrogen bond persistence scores below 0.25, as these are taken to be strong indicators of instability. The factor of 2 between the HBP and RMSD contributions makes the weighting roughly equal, as an ideal RMSD estimate of 0.5 Å and an ideal persistence score of 1 contribute equally to the score. No optimization is performed on the relative weighting of HBP. The width of the cutoff functions were postulated and left fixed while the cutoff HBP and RMSD values were optimized in the ranges 0.27−0.33 and 1.6−1.9 Å, respectively, to give the highest fraction of the 42 cases predicted to have a best pose below 2.0. This error function is flat over most of this range, so we have taken the midpoint of both. If no cutoff scheme were imposed, the overall result would be 83.3% sub-2.0 Å structures. With the cutoff in the considered range imposed, the lowest success rate is 85.7%. The number of top scoring structures with RMSDs to native above 3.25 Å is not affected by the choice of cutoff.

DATA SET

The test set is taken from the large set of IFD cross-docking test cases maintained by Schrödinger. These cases have ligands from one publicly available PDB structure docked with a receptor from another PDB structure. Output IFD structures from this set were filtered using the Structural Interaction Fingerprints (SIFt) contact similarity scoring,[20] as implemented in the Schrödinger Suite to eliminate functionally redundant poses. A pair of poses were considered functionally identical if the Tanimoto distance between the contact networks was less than 0.15. A total of 42 cases from the resulting set which contained at least one structure under 2.0 Å and at least one above 2.0 Å were taken for the metadynamics test set. IFD rank alone gives a sub-2.0 Å pose as the top-ranked score in 64% of these cases. The resulting set contains cases involving a total of eight different receptors: aldose reductase (alr2), cyclin-dependent kinase 2 (cdk2), checkpoint kinase 1 (chk1), dipeptidyl peptidase-4 (dpp4), coagulation factor XA (fxa), protein kinase A (pka), peroxisome proliferator-activated receptor gamma (ppar), and thrombin (throm). As the development of IFD is ongoing and the specific results for some cross-docking cases may change, we have provided a full set of the initial poses from IFD that were used as the input to metadynamics in this study in the Supporting Information (zip file).

To evaluate the suitability of the cases for induced fit docking, the 42 cases that composed the metadynamics test set were subjected to rigid receptor docking (RRD) using Glide SP,[1] with default settings as per the 2015-2 Schrodinger Suite release. The poses were ranked by the Glide docking score. Of the 42 cases, there are 36 cases that fail to produce anything under a 2 Å RMSD in the top five poses. These cases demonstrate a necessity for induced fit docking. A single case produced at least one pose under 2 Å, although not as the top-ranked pose. This case would be suitable for a Glide/metadynamics workflow in order to select the correct sub-2 Å pose. And finally five cases where rigid-receptor docking succeeded in returning the top-ranked pose as being under 2 Å. These results are summarized in Table 1. The RMSDs for all of the returned rigid-receptor docking Glide poses for each of the 42 cases is detailed in the Supporting Information.

Table 1. Summary of RRD Results

RRD outcome	no. of cases
no sub-2 Å pose found	36
incorrectly ranked sub-2 Å pose	1
correctly ranked sub-2 Å pose	5
total	42

RESULTS

Overall Results. For every test case, we performed 10 metadynamics runs for each of the 5 candidate poses as described in previous sections, initially using a hill height of 0.3 (the "strong" perturbation case). The system was initially relaxed through a series of short minimizations and restrained MD stages. The metadynamics trial was then performed in the NVT ensemble using a Berendsen thermostat and used a RESPA integration scheme with a time step of 2.0 fs for bonded and near atom pairs and 6.0 fs for far atom pairs. An anchor fragment was chosen for the receptor based on the algorithm discussed in the Choice of Collective Variable section. The results were assessed for both pose stability based on the RMSD of the ligand with respect to the anchor fragment and the initial internal ligand coordinates and for the persistence of hydrogen bonds (designated HBP in what follows) between the ligand and the receptor. Each trial reported here was run using Desmond on a single GPU card, with each trial for a typical system taking 1−2 h per trial. Tests on 8 CPUs in parallel for a small number were performed and show time scaling at least 10 times slower than on a single GPU. Plots for all 42 test cases are available in the Supporting Information.

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

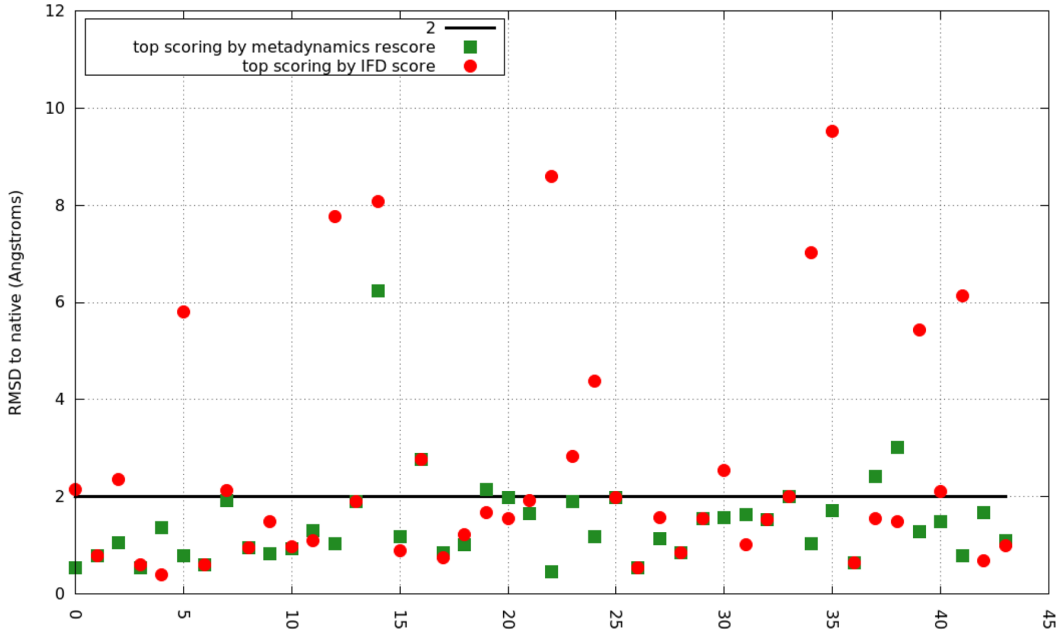


Figure 3. RMSD to native (crystallographically determined binding mode) for top-scoring pose by IFD score (red circles) shown compared to the RMSD to native of the top-scoring pose with metadynamics rescoring (green squares) shows the significant reduction in non-native-like poses detected by metadynamics rescoring. The *x*-axis is numbered in accordance with the structures in Table 1 in the Supporting Information.

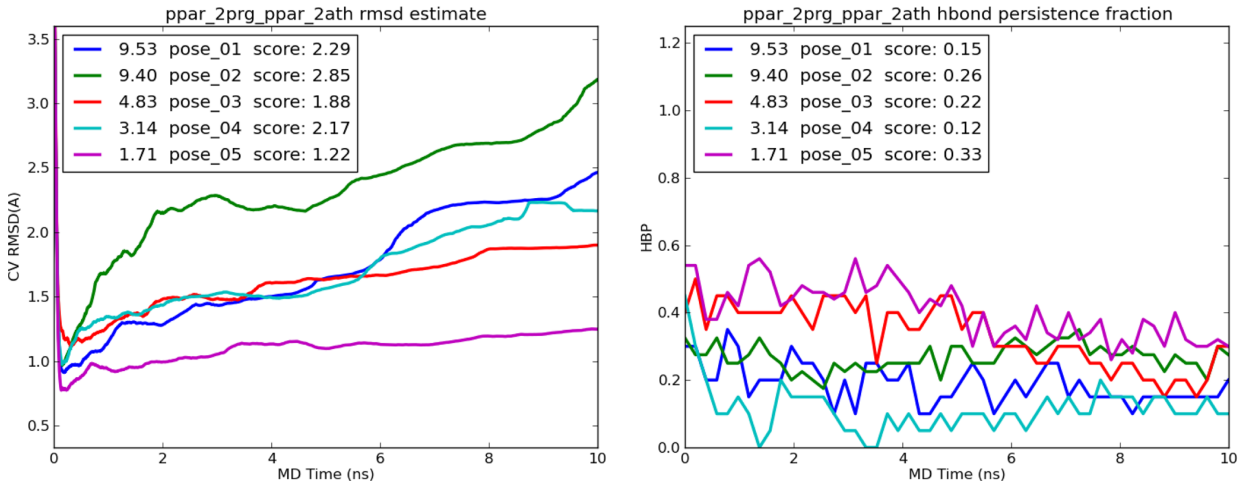


Figure 4. Plots of RMSD estimate (left) and HBP metric (right) averaged over all 10 trials vs simulation time for the ligand from PDB structure 2ath docked with the ppar receptor from PDB structure 2prg. In this and all subsequent plots, the legends can be interpreted as follows. The first column of numbers is the RMSD difference between that particular structure and the correct experimental structure. The second column is the pose numbering as per IFDScore. The third column provides the score as per RMSD (left figure) and HBP (right figure).

Table 1 in the Supporting Information summarizes the overall score for each of the 5 candidate poses obtained from this protocol, along with the RMSD and HBP values for all the cases. The poses are rank ordered by score, and the best scoring pose, listed first, is also indicated in boldface. It can be seen in the great majority of cases (88%), a pose with an RMSD < 2.0 Å from the native structure is selected. A small number of cases yield poses between 2 and 3 Å, which primarily differ from the native structure in a tail region that is close to the solvent. These structures are reasonable and, while not ideal, would in our judgment still be useful in the context of a structure based drug design project, as in many cases key contacts driving potency are identified. In contrast, many of the poses with RMSDs greater than 5 Å would be quite misleading if employed to interpret structure−activity relationships or to guide synthesis. Figure 3 shows the dramatic reduction in such non-native-like poses from IFD scoring alone to metadynamics-rescored IFD poses.

Selected Example Cases. The case of the ligand from PDB structure 2ath docked with the ppar receptor from the 2prg PDB structure provides a good example of where the metadynamics clearly separates out a native-like pose (RMSD from native 1.71 Å). This pose is fifth ranked by the IFD score, and the top four scoring poses by IFD-score are all clearly non-

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

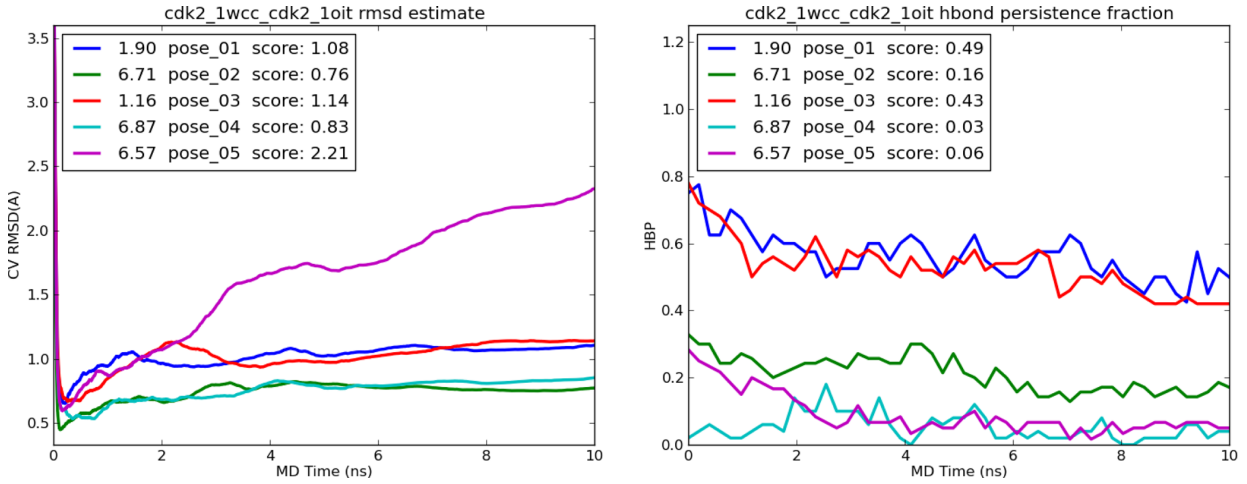


Figure 5. Plots of RMSD estimate (left) and HBP metric (right) averaged over all 10 trials vs simulation time for the ligand from PDB structure 1oit docked with the cdk2 receptor from PDB structure 1wcc. See Figure 4 for a description of the legends.

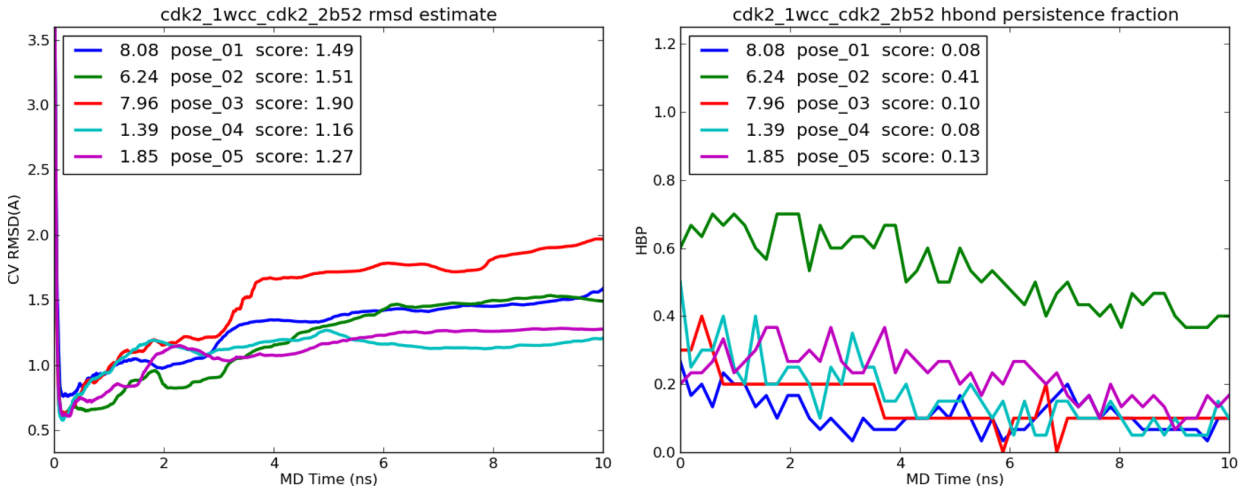


Figure 6. Plots of RMSD estimate (left) and HBP metric (right) averaged over all 10 trials vs simulation time for the ligand from PDB structure 2b52 docked with the cdk2 receptor from PDB structure 1wcc. See Figure 4 for a description of the legends.

native-like. Figure 4 shows the average RMSD estimate and HBP metric vs simulation time for all 5 IFD poses. The HBP metric also favors the 1.71 Å pose, and it is the only pose that receives a score higher than 0.3.

The case of the ligand from PDB structure 1oit docked into the cdk2 structure from PDB structure 1wcc provides an example case where the inclusion of the HBP metric eliminates poses that would otherwise be competitively stable. RMSD estimate and HBP metric plots for these cases are shown in Figure 5. Two non-native-like poses (IFD ranks 2 and 4) show the lowest RMSD estimate in the 10 ns simulations, but their starting hydrogen bond networks with the receptor are very unstable (all end up with less than 20% stability at the end). In contrast, the two native-like poses (RMSDs of 1.90 and 1.16 Å) have both reasonable RMSD stability scores ($\approx$ 1.1 Å) and show strong persistence of their initial hydrogen bonding network by the end of the metadynamics trials, allowing the simple scoring system to identify them.

There remains one case where the simple scoring system here picks out an unambiguously non-native-like structure, that of the ligand from the 2b52 PDB structure docked into the cdk2 receptor from the 1wcc PDB structure (see RMSD and HBP metric plots in Figure 6). In this case, the most native-like pose in the IFD predictions is missing a pair of prominent hydrogen bonds between its backbone and the receptor, and the competitive non-native-like pose has a more consistently stable hydrogen bond network. The RMSD average estimate of the native-like pose is actually lower, but the HBP metric strongly favors the non-native-like pose. In the majority of trials, these hydrogen bonds actually reform during the simulations. Thus, while this case is a failure by the simple scoring outlined here, the contradicting results of the two metrics give some suggestion that this case could be correctly determined with a more sophisticated fitting scheme, or alternatively if an initial low RMSD pose with better hydrogen bonding geometry was available from the original IFD calculations (which is quite possible via improvement in the IFD protocol). For the purposes of the initial test here we wish to present the results with the simple weighted metric to show how strong the raw signal in this data is.

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

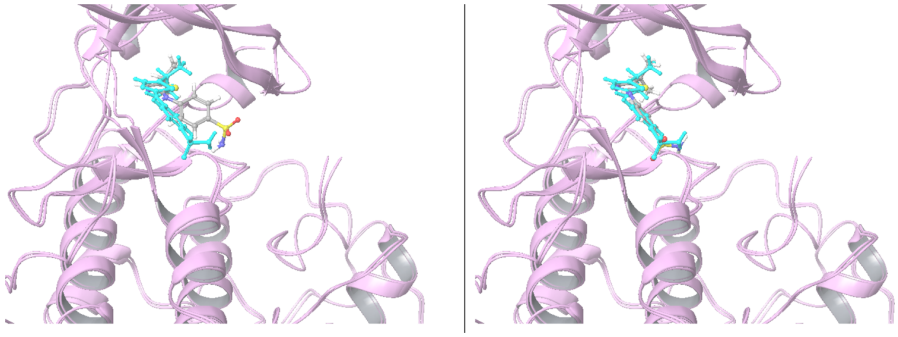


Figure 7. Pose with the exposed tail (left, 2.77 Å RMSD to native) chosen by both IFD and the metadynamics scoring for the cdk2 2bts-1wcc case is shown with the native. The pose ranked second by metadynamics, a native-like pose (right, RMSD 0.69 Å to native and IFD rank 4) is not chosen because the hydrogen bonds between the tail and the receptor are not stable in the metadynamics trials.

The remaining cases where poses above 2.0 Å RMSD from native are chosen are 2−3.1 Å poses that mostly differ from the native-like pose in a potentially solvent exposed tail region. If the tail region in all poses ends up solvent-exposed in most of the metadynamics trials, the initial pose with the most solvent exposed tail region is chosen, as the tail region in the trials is closest to the starting pose. Since the stable part of the hydrogen bond network is essentially identical, the HBP metric does not differentiate such cases. The case of the 2bts ligand docked with the cdk2 receptor, shown in Figure 7 provides a clear example of this.

DISCUSSION

The results presented above demonstrate that a properly designed metadynamics strategy is capable of reliably discriminating the correct, lowest free energy binding mode of a protein−ligand complex from plausible alternatives generated by induced fit docking. The IFD calculations, as noted above, incorporate only limited sampling of many relevant degrees of freedom (e.g., of the protein backbone) and so cannot be expected to rank order binding poses with complete fidelity. Metadynamics simulations allow simultaneous relaxation of all degrees of freedom at a rate that is substantially accelerated as compared to unbiased molecular dynamics, enabling a fair comparison of the alternative poses with a reasonable amount of computational effort. With a relatively modest investment in GPU hardware, or equivalent access to GPU processing power via cloud computing, a project team can obtain a useful structural prediction of the binding mode of any ligand of interest within a few days. Binding mode hypotheses generated in this fashion can then be used to prioritize synthesis and the structure validated via experimental structure−activity relationships.

In the present paper, we have considered cross docking cases emerging from the IFD calculations with at least one low RMSD structure in the top 5 ranked IFD poses. In our test set, there are a subset of cases where the low RMSD pose is ranked lower (typically in the top 10−15 poses), and a smaller number of cases where no good poses is generated. The former cases can be effectively addressed by the methods of this paper via combination of more aggressive clustering (thus moving up the ranking of the native-like pose) and running a larger number of candidates. The latter requires improvement in the IFD protocols themselves; preliminary results indicate that such improvement is possible. We will present a report detailing performance for a larger, more diverse test set, encompassing the above augmentations of the current methodology in a subsequent publication.

The approach proposed in this work can possibly be improved in a range of different ways, some of which we have already investigated (see the Supporting Information) and some which we will be systematically exploring in future works. For instance, the use of gaussians with variable widths that adjusts on-the-fly[21] could lead to even quicker estimates of the quantities involved. Another avenue would be to use the reweighting functionality[14] of metadynamics, where we could estimate unbiased averages of a variety of structural observables without directly biasing them.

It is interesting to note that the appropriate degree of metadynamics biasing needed to obtain a useful signal of pose stability (large enough to significantly displace the ligand but not so large as to completely eject it from the binding site) appears to depend primarily upon the binding affinity of the complex, as opposed to specifics of the receptor binding site and binding mode of the ligand. It is this observation that enables a single protocol to be used to investigate an arbitrary protein−ligand complex with a good degree of confidence. As an even larger set of test cases is investigated, some modifications of the current protocol may be required; however, we believe that the current test set is large and diverse enough to suggest that a robust, widely applicable protocol, albeit with additional sophistication, can be formulated.

CONCLUSIONS

In this work we have introduced a metadynamics-IFD coupled strategy for accurate and reliable prediction of the structures of protein−ligand complexes in a computationally tractable manner. IFD studies are now routinely used across drug design programs to generate a suite of likely stable structures. However, due to a variety of approximations, developing a system to select the one best candidate from this set of structures has been an extremely difficult task. Our strategy allows treating this problem in full atomistic detail, significantly enhancing the predictive power of IFD methods. Our approach is validated across a data set comprising 42 diverse ligand−receptor systems, with 5 candidate structures for each, with a false positive rate on the order of 5−10% (depending upon what is defined as an acceptable structural prediction).

For many years, molecular dynamics simulations have played a very limited, although useful, role in structure based drug discovery projects, primarily due to an inability to make

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998

accurate quantitative predictions concerning structures and energies (as opposed to providing qualitatively interesting snapshots of the system from dynamical trajectories). This situation is now changing rapidly due to the advent of inexpensive GPUs and better models and algorithms for carrying out simulations. In conjunction with the continued improvement in force-fields,[18] we believe that studies such as the present work and the recent free energy perturbation study[22] represent just the beginning of the vast potential in using enhanced sampling molecular dynamics as enabling tools in efficient drug design efforts.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00201.

Additional details and figures (PDF)

Full set of initial poses generated by the IFD protocol used in the metadynamics part of our method (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: pt2399@columbia.edu.

Funding

P.T. and B.J.B. acknowledge support from the National Institutes of Health (Grant NIH-GM4330).

Notes

The authors declare the following competing financial interest(s): B.J.B. is a consultant to Schrodinger, Inc. and is on its Scientific Advisory Board. R.A.F. has a significant financial stake in, is a consultant for, and is on the Scientific Advisory Board of Schrodinger, Inc.

■ REFERENCES

(1) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(2) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(3) Cavasotto, C.; Abagyan, R. *J. Mol. Biol.* **2004**, *337*, 209−225.

(4) Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F. *PLoS Comput. Biol.* **2015**, *11*, e1004586−e1004614.

(5) Gagnon, J. K.; Law, S. M.; Brooks, I.; Charles, L. *J. Comput. Chem.* **2016**, *37*, 753−762.

(6) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. *Bioorg. Med. Chem.* **2006**, *14*, 3160−3173.

(7) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. *J. Med. Chem.* **2006**, *49*, 534−553.

(8) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.

(9) Valsson, O.; Tiwary, P.; Parrinello, M. *Annu. Rev. Phys. Chem.* **2016**, *67*, 1−26.

(10) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(11) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603−020606.

(12) Dama, J. F.; Parrinello, M.; Voth, G. A. *Phys. Rev. Lett.* **2014**, *112*, 240602−240605.

(13) Tiwary, P.; Dama, J. F.; Parrinello, M. *J. Chem. Phys.* **2015**, *143*, 234112−234115.

(14) Tiwary, P.; Parrinello, M. *J. Phys. Chem. B* **2015**, *119*, 736−742.

(15) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. *J. Phys. Chem. B* **2009**, *113*, 4807−4816.

(16) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1467−1472.

(17) Tiwary, P.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2839−2844.

(18) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.

(19) Limongelli, V.; Bonomi, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358−6363.

(20) Deng, Z.; Chuaqui, C.; Singh, J. *J. Med. Chem.* **2004**, *47*, 337−344.

(21) Branduardi, D.; Bussi, G.; Parrinello, M. *J. Chem. Theory Comput.* **2012**, *8*, 2247−2254.

(22) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.

DOI: 10.1021/acs.jctc.6b00201
J. Chem. Theory Comput. 2016, 12, 2990−2998