Free Energy Calculations Using the Movable Type Method with Molecular Dynamics Driven Protein–Ligand Sampling

Wenlang Liu, Zhenhao Liu, Hao Liu, Lance M. Westerhoff, and Zheng Zheng*

ABSTRACT: Fast and accurate biomolecular free energy estimation has been a significant interest for decades, and with recent advances in computer hardware, interest in new method development in this field has even grown. Thorough configurational state sampling using molecular dynamics (MD) simulations has long been applied to the estimation of the free energy change corresponding to the receptor–ligand complexing process. However, performing large-scale simulation is still a computational burden for the high-throughput hit screening. Among molecular modeling tools, docking and scoring methods are widely used during the early stages of the drug discovery process that they can rapidly generate discrete receptor–ligand binding modes and their individual binding affinities. Unfortunately, the lack of thorough conformational sampling in docking and scoring protocols leads to difficulty discovering global minimum binding modes on a complicated energy landscape. The Movable Type (MT) method is a novel absolute binding free energy approach which has demonstrated itself to be robust across a wide range of targets and ligands. Traditionally, the MT method is used with protein–ligand binding modes generated with rigid-receptor or flexible-receptor (induced fit) docking protocols; however, these protocols are by their nature less likely to be effective with more highly flexible targets or with those situations in which binding involves multiple step pathways. In these situations, more thorough samplings are required to better explain the free energy of binding. Therefore, to explore the prediction capability and computational efficiency of the MT method when using more thorough protein–ligand conformational sampling protocols, in the present work, we introduced a series of binding mode modeling protocols ranging from conventional docking routines to single-trajectory conventional molecular dynamics (cMD) and parallel Monte Carlo molecular dynamics (MCMD). Through validation against several structurally and mechanistically diverse protein–ligand test sets, we explore the performance of the MT method as a virtual screening tool to work with the docking protocols and as an MD simulation-based binding free energy tool.

INTRODUCTION

Successful drug discovery generally relies on receptor binding active site determination, ligand virtual screening, receptor–ligand binding mode analysis, and binding free energy estimation. Structure based drug design (SBDD) supports this effort with atomic scale knowledge of one or more three-dimensional (3D) molecular structures coupled with receptor–ligand recognition mechanisms discovered from using computational tools. Computational (in silico) molecular modeling protocols such as docking and scoring and Monte Carlo (MC) and molecular dynamics (MD) simulations support the needs of large-scale virtual screening during the hit-to-lead stage of drug discovery in which these methods have long been applied to help control the trial-and-error cost of the design and synthesis of drug leads.

Docking and scoring methods are widely used during the early stages of the drug discovery process by rapidly generating discrete receptor–ligand binding modes and quickly calculating their individual binding affinities. Such methods provide high-throughput virtual screening filters to reduce the number of compounds for synthesis and subsequent in vitro or in vivo evaluation. The goal of docking and scoring is therefore often to capture global minimum binding modes using a “one shot” pose-generation or a restricted local sampling protocol followed by estimation of receptor–ligand binding free energies using the calculated potential energies and conforma-...
tional entropy-related features according to these binding modes. However, the lack of thorough conformational sampling in the docking and scoring computations brings difficulty in exploration of the complicated energy landscape of the protein–ligand binding process. On one hand, correlated motif movements, especially at the receptor–ligand recognition interface, during the binding process compromise the accuracy of the global minimum binding mode prediction from using only local pose-sampling protocols. On the other hand, binding affinity prediction based on a single binding mode conformation usually leads to ligand-size-dependence, in which larger molecules outperform smaller ones simply because there are more intermolecular pairwise contacts. The binding free energy estimation is therefore contaminated by these “one shot” pose-generation strategies due to lack of knowledge of the converged conformational distributions for both bound-state and free-state protein–ligand pairs. However, notwithstanding the inherent disadvantages of the docking and scoring methods, they are still among the most popular SBDD tools because of their computational speed and reasonable binding mode prediction capabilities for many cases.

Thorough configurational state sampling using MD simulations has been applied for decades to the estimation of the free energy change corresponding to the receptor–ligand complex process. These methods are often quite reliable for revealing molecular movement trajectories and generating binding mode probability distributions. However, even today with software optimized for fast central processing units (CPUs) and graphical processing units (GPUs), performing large-scale simulations at scale is still a significant computational burden for high-throughput hit screening. There are two primary binding free energy estimation regimes: absolute and relative, with the former term suggesting the calculation procedures directly generate the binding free energy change ($\Delta G$) involving conformational sampling of end-point states or along the pathway collective variables, and the latter term indicating the protocols first estimating the relative binding free energy change ($\Delta \Delta G$) of a new molecule referencing a known protein–ligand complex and then generating the binding free energy change ($\Delta G$) using the thermodynamic cycle.

Regarding applicable scenarios and prediction accuracy, relative binding free energy methods benefit from using the thermodynamic cycle that common interaction energy errors among similar molecules can be canceled. Given the binding free energy of a reference ligand, relative binding free energy protocols usually have good accuracy performance from using the free energy perturbation involving alchemical structural transformations. FEP and TI based relative binding free energy methods have been recently reported to achieve good simulation with errors around 1 kcal/mol on average against large-scale validations with test sets including congeneric ligands targeting common protein targets. However, relative binding free energy methods usually have limited application in the cases where no known ligands were discovered for the protein target, or structural differences were huge among the molecules. On the other hand, without depending on the reference molecule and structural modifications, absolute binding free energy methods are more robust against wider protein and ligand space but require more comprehensive conformational sampling for both free-state and bound-state receptor and ligand systems.

Fast, effective free energy methods are in high demand to fulfill a growing need in the fast-paced drug discovery efforts. The MovableType (MT) method is a recently introduced absolute binding free energy method which has shown itself to be competitive with other, much more computationally expensive methods by utilizing ensembles of protein–ligand structures generated using docking methods to mimic protein–ligand binding; but docking is often limited to rigid or semirigid target protein structures, and therefore relying on docking alone can limit the applicability of MT to fairly rigid protein systems. In computational chemistry, there is a number of binding mode sampling methods available ranging from docking to large-scale molecular dynamics protocols which are generally distributed on a spectrum of performance between high efficiency and high accuracy. In this study, we evaluate the impact of the chosen sampling protocol on protein–ligand binding free energy estimation by challenging the MT method. Two validation benchmarks are used in this study including the well-known CASF-2016 set which covers a wide variety of protein and ligand structures and the recently published Merck KGaA benchmark which consists of a set of 264 ligands with minor structural differences targeting eight pharmaceutically valuable receptors. Since the MT free energy method has been developed to use bound-state and free-state conformations of protein ligand molecular systems for the evaluation of absolute binding free energies, we use the CASF and Merck KGaA sets to explore the compatibility of the MT free energy method working with different sampling methods and show that the method is applicable not only as a virtual screening algorithm by using docking generated poses but also as a flexible free energy evaluation algorithm by using protein–ligand complex trajectories from molecular dynamics simulations.

## METHOD

The MT Method for Protein–Ligand Absolute Binding Free Energy Estimation. The MT method is a fast absolute binding free energy approach showing its robustness across a wide range of targets and ligands. It combines global sampling methods for generating ensembles of free-state and bound-state protein–ligand molecular conformations with a novel local sampling method which is used to numerically estimate local partition functions with respect to these sampled conformations. The current version of the MovableType software (available at [http://www.quantumbioinc.com/](http://www.quantumbioinc.com/)) contains two protocols for the protein–ligand binding free energy estimation. The full protocol, denoted as MT$_{ScoreES5P}$ performs partition function estimations for the bound-state complex (hole) and free-state protein (apo) and ligand and then calculates the protein–ligand binding free energy with the ratio of the approximate partition functions. On the other hand, the simplified protocol, denoted as MT$_{ScoreESP}$ predicts the protein–ligand binding affinity only using one single accurate complex conformation as the representative energy state. This simplified protocol is much faster than the full protocol and is positioned for higher-throughput virtual screening tasks. In this project, we applied the full MT protocol, MT$_{ScoreES}$ as the free energy evaluation tool coupled with different conformational sampling approaches for the estimation of the protein–ligand binding free energies. Its computational procedure is explained in the following paragraphs. The simplified protocol, MT$_{ScoreES5P}$ is also briefly summarized at the end of this subsection. The MT$_{ScoreES}$ Protocol is applied to the linear scaling parameter set
generation with respect to the results rescale during the Merck KGaA set evaluation. This protocol is also applied to score the energy-minimal snapshots selected from the simulation trajectories to compare with the conformational ensemble-based binding free energy estimation results.

The full MT protocol first employs built-in or external molecular energy-state sampling methods (e.g., docking or dynamics) to generate significant free-state and bound-state protein—ligand conformations. Then, for each sampled molecular conformation, local numerical integrations are performed to generate an atomic pairwise potential ensemble which samples the local space of the atomic coordinates of the globally sampled molecular conformation. The local pairwise potential integrations are multiplied through all the atom pairs in the molecular conformer (e.g., protein, ligand, and complex) to estimate the local molecular energy states in the vicinity centered at each sampled conformation. These local energy states approximate local molecular partition functions, and when the MT method is supplied with multiple sampled conformations of the protein (apo), ligand, and complex (holo), the method sums over all the local partition functions regarding the free-state and bound-state NVT ensembles. This calculation yields an estimate of the approximated total molecular partition functions of the protein—ligand system before and after complexation. The ratio of the partition function approximations is then used to estimate the gas-phase binding free energy as represented in eq 1:

\[ \Delta G_{\text{binding}} \approx -RT \log \left( \frac{Z_{PL}^M}{Z_{PL}^M} \right) \]  

(1)

In this equation, each partition function \( Z_{PL} \), \( Z_{PL} \), and \( Z_{PL} \) for protein (apo), ligand, and complex (holo) respectively is calculated using eq 2

\[ Z_{PL} = \sum_i Z_{PL,i}^{N_{PL}} = Z_{PL,i}^{N_{PL}} + Z_{PL,i}^{N_{PL}} + ... + Z_{PL,i}^{N_{PL}} \]  

(2)

where \( M \) indicates the sampled molecule at different states, and every \( Z_{PL,i}^{N_{PL}} \) is estimated using numerical integration against all the atom pairwise interactions of each sampled conformational state \( i \). For this study, we employed DivCon Discovery Suite v.815 in which potential energies for every sampled conformation along with the partition functions represented in eqs 1 and 2 are calculated using two sets of energy functions implemented in the DivCon Discovery Suite: (1) GARF energy function developed by Zheng et al. (GARF set)\(^{30}\) and (2) Amber ff19SB and General Amber Force Field (GAFF) force field optimized for the MT method (Amber set).\(^{31,32}\) When Amber ff19SB was employed, the ligand atomic charges were supplied as Mulliken charges calculated using the built-in PM6 Hamiltonian.

This MT method, referred to in the balance of this work as MTScoreES or ensemble-score, employs conformational sampling protocols to generate energy-state ensembles for a protein—ligand pair in both unbound and bound state, and it approximates the unbound-state and bound-state NVT partition functions using the generated energy-state ensembles to derive the Helmholtz free energy using a ratio of partition functions.

In this work, the partition function approximation for the bound-state complex is calculated as

\[ Z_{PL}^{N_{PL}} = \sum_{i,j,k} Z_{PL,i}^{N_{PL}'} = Z_{PL,i}^{N_{PL}} + Z_{PL,i}^{N_{PL}} + ... + Z_{PL,i}^{N_{PL}} \]  

(3)

where \( N_{PL} \) denotes the number of sampled complex-state conformations. The conformational sampling protocols introduced in this work for the bound-state and free-state partition function approximations will be visited in the following subsection.

Similarly, the partition function approximations for the free-state protein and ligand are calculated respectively as

\[ Z_{PL}^{N_{PL}} = \sum_{j,k} Z_{PL}^{N_{PL}'} = Z_{PL}^{N_{PL}} + Z_{PL}^{N_{PL}} + ... + Z_{PL}^{N_{PL}} \]  

(4)

and

\[ Z_{PL}^{N_{PL}} = \sum_{k} Z_{PL}^{N_{PL}'} = Z_{PL}^{N_{PL}} + Z_{PL}^{N_{PL}} + ... + Z_{PL}^{N_{PL}} \]  

(5)

\( N_{PL} \) and \( N_{PL} \) in eqs 4 and 5 denote the numbers of sampled free-state protein and ligand conformations, respectively. Applying the approximated partition functions \( Z_{PL}^{N_{PL}} \), \( Z_{PL}^{N_{PL}} \), and \( Z_{PL}^{N_{PL}} \) to eq 6, the gas-phase protein—ligand binding free energy is then calculated using the logarithm of the ratio of bound-state and free-state partition functions.

\[ \Delta G_{\text{binding}}^{\text{gas}} \approx -RT \log \left( \frac{Z_{PL}^{N_{PL}}}{Z_{PL}^{N_{PL}} Z_{PL}^{N_{PL}}} \right) \]  

(6)

The KMTISM solvation model is then employed for the solvation free energy change during the binding procedure.\(^{56}\) KMTISM is an implicit solvation model that generates multiple solvent layers around each sampled solute conformation and calculates the solute—solvent interaction potentials by adding up the layer-based solvent—solvent contact energies. Ensembles of solvent—solvent interaction potentials are then generated for the protein—ligand complex and the free-state protein and ligand. Hence the solvation free energy change is calculated using the following equation

\[ \Delta G_{\text{sol}} = \frac{1}{\beta} \ln \left( \sum_i \exp(-\beta(E_{PL,i}^{sol})) \right) - \frac{1}{\beta} \ln \left( \sum_i \exp(-\beta(E_{PL,i}^{sol})) \right) \]  

(7)

where \( (E_{PL,i}^{sol}) \) and \( (E_{PL,i}^{sol}) \) are the solvent—solvent interaction potentials for the sampled complex, protein, and ligand conformations, respectively. \( i, j, \) and \( k \) are the index numbers for the corresponding conformations. The protein—ligand binding free energy is then calculated as

\[ \Delta G_{\text{binding}} = \Delta G_{\text{binding}}^{\text{gas}} + \Delta G_{\text{sol}} \]  

(8)

With more careful and thorough sampling protocols, the MT method has a better chance to obtain the most significant conformations in the molecular conformational ensemble. On the other hand, we can also introduce fast and restricted sampling methods to the MT method for higher-throughput virtual screening tasks. In addition to MTScoreES, a simplified MT protocol named MTScoreE, where "ES" denotes "end-state", has also been developed which utilizes a single binding pose as the representative energy state for the binding free energy prediction.\(^{53}\) This approach only calculates the intermolecular
atom pairwise potentials between the protein and ligand pair
referring the input protein–ligand binding pose. The
binding free energy is then approximated in eq 9
\[
\Delta G_{\text{binding}} \approx -RT \log(\mathcal{Z}_{\text{InterPL}}^0)
\]
where \(\mathcal{Z}_{\text{InterPL}}^0\) is the protein–ligand binding pose’s local
partition function considering only the intermolecular atomic
pairwise interaction potential.

Conformation Sampling Protocols. When provided
proper ensembles of molecular conformations for unbound-
state and bound-state protein ligand systems, MT is
theoretically compatible with many different conformational
sampling protocols. In our previous studies, we mainly focused
on introducing docking protocols to work with MT for virtual
screening purposes. In the present work, we further explore the
performance of the MT method when it is challenged with
conformers generated with different simulation-based con-
formational sampling protocols. While docking+MT is robust and
flexible for many cases, without thorough configurational
space coverage or reliable energy gradient searching, discrete
conformational sampling protocols have inherent limitations
locating significant molecular configurations. One would
expect that introducing more globally significant conforma-
tions or landscape minima should improve the free energy
prediction accuracy of the MT method with the addition of
more converged energy states. Therefore, in addition to
docking, this work utilizes conventional MD (cMD) and
Monte Carlo MD (MCMD) simulations. Protein–ligand complex “snapshots” from these sampling simulations are
provided to MT to improve binding free energy estimation
accuracy. With this support in place, we evaluate the MT
method as a flexible free energy tool for its compatibility with
different energy functions and sampling strategies. This work
open further opportunities for using this method to analyze
MD trajectories and receptor–ligand complexing mechanisms.

The MT_{CS} and MT_{Flex} protocols are built-in modules in the
DivCon Discovery Suite to generate the free-state ligand and
protein conformations based on corresponding prebuilt small-
molecule and amino-acid rotamer libraries, respectively, followed by local structural minimization. Both modules
contain corresponding molecular fragment libraries, in which
every fragment is collected from structural fragmentation
against the ligands and proteins from Protein Data Bank. Every
molecular fragment contains one single rotatable bond, along
which the rotational energy curve for every fragment is
precalculated using the built-in energy functions, i.e., Amber
ff19SB and General AMBER Force Field (GAFF) force field,
or GARF energy function. When using the MT_{CS} protocol for
the conformational sampling of ligands, the energy-minimum
rotamers for necessary structural fragments from the libraries
are selected to construct the conformations of the correspond-
ing molecule. Local energy minimization is then performed to
generate reasonable conformations according to their energy-
based scores. On the other hand, due to the vast conforma-
tional space of proteins, MT_{Flex} performs local conformational
search based on the input protein configuration, instead of
constructing the protein conformations totally based on
fragment rotamer assembly in a “de novo” manner. For every
amino-acid conformation from the input protein conformation,
the MT_{Flex} protocol selects the close-by energy-minimum
rotamers and hence assembles the protein conformation,
followed by local structural minimization. In such a case, the
MT_{Flex} protocol generates a cluster of conformations, resembling the configurational features of the input protein
structures.

In this project, we select the same number of free-state protein conformations as the number of bound-state complex
conformations for the corresponding partition function approximations \((N_P \approx N_{PL})\), mostly because the protein
usually dominates conformational space in a protein–ligand
complex. To explore the convergence of the approximated
partition function estimation, and hence the reliability of the
\(\Delta G_{\text{binding}}\) prediction, we validate the \(\Delta G_{\text{binding}}\) estimation
convergence by introducing different numbers of sampled
conformations, to decide the proper \(N_P\) and \(N_{PL}\) numbers for sampling the bound-state complex and the free-state protein.

See the detailed validation procedure and result analysis in the
Supporting Information.

On the other hand, following the default setting introduced in
our previous work, the top five lowest-energy conformers for
the free-state ligand are selected from the MT_{CS} protocol.
Compared to the number of protein conformations, much
fewer free-state ligand conformations are chosen because the
conformational energies for most ligands increase significantly
soon after the top 3 to 5 conformations.

We employed three different protocols to generate binding
modes for use in this project: high-throughput with induced-fit
docking as implemented in the Molecular Operating Environ-
ment (MOE) v2019 package; exhaustive with molecular
dynamics simulations using the Amber18 suite; and balanced
(sampling efficiency and exhaustiveness) with MCMD as
captured with the parallel Consecutive Histograms Monte
Carlo (CHMC) method combined with MD simulation. The
following conformational sampling methods are employed for
the protein–ligand binding modes generation:

MT with MOE Induced Fit Docking. The coordinate files
utilized in the present study structures were downloaded from the
RCSB/PDB, and preparation was performed using the
Molecular Operating Environment (MOE) v2019 package, using the Structure Prepare application followed by the
Protonate3D application to add and optimize protons. After
structure preparation was complete, docking for this protocol
was performed using the same version of MOE. For this
calculation, the ligand parameters were calculated using the
Extended Hückel Theory (Amber10:EHT) as implemented in
MOE, and the “Induced Fit” docking (IFD) protocol was
chosen for each protein–ligand model. Under the IFD protocol, initial placement was performed using the Triangle
Matcher approach, and the London \(\Delta G\) score was used as the
initial score followed by the GBV/WSA \(\Delta G\) score used as the
final filter. In each case and in order to maximize the number
of initial poses, 250 poses were provided by the Triangle
Matcher, and the active site was optimized using the
Amber10:EHT force field as implemented in MOE. Finally, 25
poses were passed to MT_{ScoreE} as landscape minima for
ensemble scoring.

MT with the Conventional Molecular Dynamics (cMD) Simulation. Molecular dynamics simulations for all
the protein–ligand complexes were performed using Amber18
with the ff99SB force field. The Generalized AMBER Force Field (GAFF) was used for ligands and for nonstandard
residues as characterized using the Antechamber program.
Parameters for any metal ions (e.g., Zn^{2+}, Mg^{2+}, and Ca^{2+})
were introduced from the work of Li et al., and the partial
charges of ligands and nonstandard residues were calculated

using the Antechamber program with the AM1-BCC method.\textsuperscript{44,45} For each molecular simulation, the initial conformation for each complex was the best MOE docking pose (determined using the GBVI/WSA $\Delta G$ score) generated from the above-noted induced-fit docking protocol. The complex was solvated in a rectangular TIP3P water box with periodic boundary conditions and neutralized with Na+ or Cl− (as required) using the tleap program. The SHAKE algorithm\textsuperscript{41} was applied to constrain bonds containing hydrogen atoms with a 2 fs time step, and the electrostatic interactions were calculated using the PME (particle mesh Ewald)\textsuperscript{42} with a cutoff of 10.0 Å for long-range interactions. Prior to beginning the simulation, each complex was treated with an energy minimization consisting of 10000 steps of steepest descent algorithm followed by 10000 steps conjugate gradient algorithm. Each molecular model was gradually heated from 0 to 300 K in the isothermal-isovolumetric (NVT) ensemble for 200 ps and then equilibrated for 1 ns in the isothermal–isobaric (NPT) ensemble (300 K and 1.01325 MPa). Finally, the molecular dynamics simulations were carried out in the NPT ensemble (300 K).

**MT with the Consecutive Histograms Monte Carlo (CHMC) Protocol Followed by Parallel cMD Simulation.** Recently, our laboratory developed the CHMC method employing a mixed Monte Carlo/Molecular Dynamics protocol to estimate the potential of mean force (PMF) regarding the protein–ligand binding or dissociation process.\textsuperscript{45} CHMC first generates a series of consecutive Monte Carlo sampling units with equal volume representing the receptor–ligand association/dissociation space. These sampling units are stacked from within the receptor’s binding active site (which is designated as the ligand’s initial position), along a dissociation path vector extending to the ligand’s final location which is usually in the open space representing the fully dissociated state for the protein–ligand system.

In this study, we divided each protein model’s active site into five sampling units with equal volume. Each sampling unit was then used for modeling equivalent and comparable NVT ensembles for the estimation of the partition functions in each sampling unit. CHMC generates a set of initial binding modes within each sampling unit using a docking-like protocol, and in the present project, the top 50 binding modes (based on the protein–ligand interaction energies) per sampling unit were selected for the next calculation stage. Geometric minimization was performed against these binding modes using the Newton–Raphson method to obtain an appropriate receptor–ligand complex ensemble for each sampling unit. The top-five separate minimum-energy protein–ligand complex geometries at each of the sampling units were chosen based on the protein–ligand binding mode energies. According to the free energy estimation protocols introduced in the following subsection, Amber and GARF energy functions were both introduced in the geometric minimization and protein–ligand binding mode energy calculations. The selected five protein–ligand complex geometries were thus used as initial conformations for 50 ns cMD simulations using the same

---

**Figure 1.** Illustration of the free energy estimation protocols applied to (A) the CASF-2016 benchmark and (B) the Merck benchmark.
setup, structure preparation, and protocol described in the “MT with Conventional Molecular Dynamics (cMD) Simulation” subsection above.

**General Setting for the MT Free Energy Method.** In this work, we utilized the three sampling protocols mentioned above coupled with two pairwise potential functions available in the MT package for protein–ligand binding free energy estimation. In the following paragraphs, we use “MOE-MT-amber” to designate the MT ensemble scoring protocol challenged with MOE induced-fit docking poses in which the partition function is calculated using the Amber energy function; “MOE-MT-garf” to designate the MT protocol with MOE docking poses and the GARF energy function; “cMD-MT-amber” to designate the MT protocol performing 250 ns conventional MD simulation and calculating the MT partition functions using the Amber energy function; the term “cMD-MT-garf” denoting the MT protocol performing CHMC parallel sampling using the GARF energy function; the term “CHMC-cMD-amber” denoting the MT protocol performing CHMC parallel sampling using the Amber energy function followed by 50 ns conventional MD simulation and calculating the MT partition functions using the Amber energy function; the term “CHMC-cMD-MT-garf” denoting the MT protocol performing CHMC parallel sampling using the GARF energy function followed by 50 ns conventional MD simulation and calculating the MT partition functions using the GARF energy function.

From the MOE docking process, 25 top ranked binding poses are selected as energy minima for the bound-state partition function estimation in the MT method. From the MD simulations, including the cMD and CHMC-cMD calculations, 500 equally spaced snapshots of every protein–ligand complex (every 10 ps) are selected from the converged region of the production MD trajectories. For each protein–ligand complex, 5 different sets of the 500 snapshots from the converged MD trajectories are selected for the estimation of error bars (standard deviations) and 90% confidence intervals.

**RESULTS AND DISCUSSION**

We applied the MT free energy method and different conformational sampling protocols through the treatment of two validation benchmarks: (1) the industry-standard Comparative Assessment of Scoring Functions v2016 set (the CASF-2016 benchmark) containing 57 protein targets and 285 ligands; and (2) the validation benchmark published by Merck KGaA originally for the evaluation of Schrödinger FEP+ including 8 protein targets and 264 ligands (the Merck benchmark). The CASF-2016 benchmark covers various protein and ligand structures for validating the robustness of different sampling protocols across a broad range of protein and ligand classes. On the other hand, the Merck benchmark focuses on pharmaceutically relevant targets and ligand series containing ligand structures with subtle structural differences.

Figure 1 summarizes the calculation procedures applied to the CASF-2016 benchmark and the Merck benchmark.

For statistical analysis of the free energy calculation results compared to the experimental binding free energies, we utilized Pearson’s R and Kendall’s τ for the measure of linear and rank correlation between the predicted and experimental binding free energies and Mean Absolute Error (MAE) for measuring the calculation errors. They are calculated using the following equations respectively

\[
R = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

\[
\tau = \frac{2 \sum_{i<j} \text{sgn}(x_i - x_j)\text{sgn}(y_i - y_j)}{n(n-1)}
\]

where \(\text{sgn}(x_i - x_j) = -1\) if \(x_i < x_j\) and \(\text{sgn}(x_i - x_j) = 1\) if \(x_i > x_j\) and analogously for \(y\).

\[
\text{MAE} = \frac{\sum_{i=1}^{n} |(y_i - x_i)|}{n}
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}}
\]

To compare the error performance of the MT protocols with other published free energy methods, for all the target sets from the Merck benchmark, we also utilized Root-Mean-Square Error (RMSE) evaluated using the following equation:

Table 1. Results for Protein-Ligand Binding Free Energy Estimation Using Different MT-Based Protocols

| MT protocols         | Pearson’s R | Kendall’s τ | MAE (kcal/mol) |
|----------------------|-------------|-------------|----------------|
| cMD-MT-amber         | 0.66        | 0.49        | 2.10           |
| cMD-MT-garf          | 0.70        | 0.52        | 1.68           |
| CHMC-cMD-MT-amber    | 0.66        | 0.49        | 2.10           |
| CHMC-cMD-MT-garf     | 0.71        | 0.52        | 1.67           |

As noted in the Methods section, the top-scored docked pose from MOE was chosen as the initial complex conformation for both the cMD-MT-amber and the cMD-MT-garf protocols, and the CHMC method was used to generate five different initial complex conformations to initiate the MD simulations in the CHMC-cMD-amber and CHMC-cMD-MT-garf protocols. Using the AMBER09SB force field, both the cMD-MT-amber and the CHMC-cMD-MT-amber protocols achieve Pearson’s R values of 0.66, MAE values of 2.10 kcal/mol, and Kendall’s τ values of 0.49. Using the GARF energy function, the cMD-MT-garf and CHMC-cMD-MT-garf protocols generated improved results, with Pearson’s R values of 0.70 and 0.71, MAE values of 1.68 and 1.67 kcal/mol, and Kendall’s τ values of 0.52. Introducing simulation-based sampling methods in this study provided us with converged energy ensembles of each protein–ligand complex. Therefore, the initial conformations from different sampling methods were converged through the MD simulation into similar conformational ensembles, resulting in similar free energy estimation values. Figure 3 illustrated the free energies obtained from both cMD-MT and CHMC-cMD-MT against the CASF-2016 benchmark and discovered very high
correlations between these two protocols when using the same energy functions.

In 2020, we reported free energy estimation results using MOE-MT-amber and MOE-MT-garf against the CASF-2016 benchmark. In this prior work, based on 25 docking poses for each ligand, the MOE-MT-amber and MOE-MT-garf protocols generated Pearson’s R values of 0.63 and 0.64, respectively. In another publication, we reported the binding free energy reproduction against the CASF-2016 benchmark using the CHMC method with the GARF energy function, achieving the Pearson’s R as 0.61 and the MAE as 2.06 kcal/mol. The present project shows that MD simulation-based sampling methods improve MT binding free energy prediction accuracy compared to the docking methods. Furthermore, the CHMC-cMD-MT-garf protocol shows that the free energy estimation accuracy with the parallel sampling strategy in CHMC is improved by introducing multiple short MD simulations and pairing this increased sampling with numerical integration and atom pairwise local sampling in MT.

While the present work represents the most exhaustive exploration of the impact of improved sampling on MT-based free energy predictions performed to date, there are still potential sources of error in these receptor–ligand binding free energy calculations. Specifically, the presented results utilized bound-state conformational sampling and therefore calculation of the bound, \( Z_{PL} \) partition function; however, accurate binding free energy estimation also depends on conformational sampling for the free-state receptor and ligand molecules along with the change in solvation free energy during the binding process. Therefore, against test cases with flexible receptor and ligand structures, the lack of exhaustive conformational sampling for the free-state protein and ligand may introduce significant error in the predicted binding free energies. In the present effort, we focused on comparing the

---

**Figure 2.** Experimental binding free energies versus predicted binding free energies for the 285 protein–ligand complexes from the CASF-2016 test benchmark. Standard deviations for the test cases were shown as error bars.
different binding mode sampling protocols by looking at the calculation results against the subsets targeting the identical protein receptor. In subsequent work, we will explore the question of whether we can further improve our predictions if we use apo as well as holo protein sampling.

The CASF-2016 benchmark contains 57 subsets, each of which has five different ligands targeting the same protein receptor. In our previous publication, we reported the binding free energy estimation results of 52 out of the 57 subsets from the CASF-2016 benchmark when the MT method was provided with 25 MOE dock-based poses for each of the protein:ligand complexes in those 52 subsets. To compare the performance of these earlier MOE-MT calculations with the present cMD-MT and CHMC-cMD-MT calculations on a case-by-case basis, Figure 4 shows the Pearson’s R values ranked according to each protocols’ performance and represented in pairs. For the cMD-MT-amber protocol, 36 subsets had Pearson’s R values higher than 0.7, and 40 subsets had Pearson’s R values higher than 0.6. Using the parallel, CHMC-cMD-MT-amber simulation protocol, 35 subsets had Pearson’s R values higher than 0.7, and 40 subsets had Pearson’s R values higher than 0.6. These results compared to

Figure 3. Correlations between predicted binding free energies using cMD-MT and CHMC-cMD-MT protocols against the CASF-2016 benchmark. The correlation on the left illustrates the predicted binding free energies using cMD-MT-amber (x-axis) vs CHMC-cMD-MT-amber (y-axis). The correlation on the right illustrates the predicted binding free energies using cMD-MT-garf (x-axis) vs CHMC-cMD-MT-garf (y-axis).

Figure 4. Pearson’s R coefficients comparisons for simulation-based and docking-based protocols against 52 subsets from the CASF-2016 benchmark. Pearson’s R values for each protocol were ranked individually in their own descending orders.
the earlier MOE-MT-amber protocol in which 31 subsets had Pearson’s $R$ values higher than 0.7, and 33 subsets had Pearson’s $R$ values higher than 0.6. When performing the cMD-MT-garf protocol, 35 subsets had Pearson’s $R$ values higher than 0.7 and 41 subsets had Pearson’s $R$ values higher than 0.6. Using the parallel CHMC-cMD-MT-amber protocol, 36 subsets had Pearson’s $R$ values higher than 0.7 and 41 subsets had Pearson’s $R$ values higher than 0.6. This compares favorably to the original MOE-MT-garf protocol for which 29 subsets had Pearson’s $R$ values higher than 0.7, and 32 subsets had Pearson’s $R$ values higher than 0.6.

From the calculation results comparison regarding the CASF-2016 subsets, we observe that the free energy protocols with simulation-based sampling methods generally obtain better agreement with the experimental data. When using the GARF energy function, the advantage for performing simulation-based sampling methods is even more pronounced. Moreover, the single-trajectory, exhaustive (250 ns) MD simulations (cMD-MT-amber and cMD-MT-garf) and the multitrajectory, short-time (50 ns) MD simulations (CHMC-cMD-MT-amber and CHMC-cMD-MT-garf) achieved similar free energy estimation accuracies suggesting that the free energy estimation scheme using multiple local samplings

| Protein         | No. of ligands | MOE-MT-amber | MOE-MT-garf | cMD-MT-amber | cMD-MT-garf | CHMC-cMD-MT-amber | CHMC-cMD-MT-garf |
|-----------------|----------------|--------------|-------------|--------------|-------------|-------------------|------------------|
| CDK8            | 33             | 0.407        | 0.477       | 0.463        | 0.505       | 0.620             | 0.639            |
|                 |                | 0.396        | 0.420       | 0.399        | 0.440       | 0.585             | 0.572            |
|                 |                | 0.726        | 0.854       | 0.757        | 0.852       | 0.757             | 0.763            |
| c-Met           | 24             | 0.724        | 0.822       | 0.749        | 0.852       | 0.757             | 0.843            |
|                 |                | 0.720        | 0.783       | 0.740        | 0.812       | 0.749             | 0.805            |
| Eg5             | 28             | 0.483        | 0.394       | 0.582        | 0.584       | 0.537             | 0.707            |
|                 |                | 0.480        | 0.378       | 0.567        | 0.546       | 0.513             | 0.668            |
| Eg5 with an alternative loop | 28 | 0.504 | 0.505 | 0.494 | 0.542 | 0.651 | 0.697 |
|                 |                | 0.496        | 0.488       | 0.473        | 0.520       | 0.633             | 0.682            |
| HIF-2α          | 42             | 0.107        | 0.237       | 0.257        | 0.507       | 0.177             | 0.263            |
|                 |                | 0.090        | 0.134       | 0.182        | 0.428       | 0.095             | 0.167            |
| PFKFB3          | 40             | 0.010        | 0.056       | 0.689        | 0.840       | 0.628             | 0.741            |
|                 |                | -0.005       | -0.005      | 0.567        | 0.698       | 0.509             | 0.644            |
| SHP-2           | 26             | -0.065       | 0.156       | 0.232        | 0.416       | 0.391             | 0.501            |
|                 |                | -0.092       | 0.014       | 0.148        | 0.249       | 0.279             | 0.348            |
| SYK             | 44             | 0.204        | 0.300       | 0.321        | 0.411       | 0.288             | 0.272            |
|                 |                | 0.202        | 0.279       | 0.296        | 0.365       | 0.273             | 0.265            |
| TNKS2           | 27             | 0.591        | 0.541       | 0.705        | 0.640       | 0.671             | 0.597            |
|                 |                | 0.588        | 0.522       | 0.644        | 0.611       | 0.611             | 0.531            |
| Total           | 264            | 0.551        | 0.554       | 0.620        | 0.656       | 0.658             | 0.626            |
|                 |                | 0.547        | 0.519       | 0.589        | 0.598       | 0.633             | 0.581            |
followed by parallel short-time MD simulations could obtain comparable accuracy to the free energy estimation scheme using the normal, longer running single-trajectory simulation strategy.

In the CASF-2016 benchmark, each subset contains only 5 test cases making the Pearson’s $R$ values sensitive to protein–ligand complex selection and individual outliers, but with a sufficiently large test set, we can still obtain meaningful statistical results from the subsets study. The simulation-based methods achieved Pearson’s $R$ values higher than 0.7 for 67% to 69% of all the 52 subsets and Pearson’s $R$ values higher than 0.6 for 77% to 79% of the 52 subsets. The simulation-based MT free energy protocols show significant correlations with the experimental binding free energies in most of the homologous protein–ligand test sets in the CASF-2016 benchmark, and these methods show improved prediction profiles compared to the previous docking-based MT methods.

To further explore the performance of the MT free energy method with different binding-mode sampling protocols, we employed another public validation benchmark, first introduced by Merck KGaA for FEP+ evaluation, which includes 8 protein targets and their corresponding congeneric ligand series.

Table 3. Kendall’s $\tau$ Values with 90% Confidence Intervals for All the Target Sets Using Different Free Energy Protocols

| Protein      | No. of ligands | MOE-MT-amber | MOE-MT-garf | cMD-MT-amber | cMD-MT-garf | CHMC-cMD-amber | CHMC-cMD-garf |
|--------------|----------------|--------------|-------------|--------------|-------------|----------------|---------------|
| CDK8         | 33 0.287       | 0.379        | 0.389       | 0.508        | 0.579       | 0.545          | 0.472         |
|              |                | 0.283        | 0.295       | 0.306        | 0.341       | 0.468          | 0.395         |
|              |                | 0.571        | 0.703       | 0.645        | 0.799       | 0.684          | 0.689         |
| c-Met        | 24 0.527       | 0.662        | 0.603       | 0.752        | 0.674       | 0.623          | 0.601         |
|              |                | 0.498        | 0.637       | 0.579        | 0.681       | 0.623          | 0.601         |
|              |                | 0.241        | 0.209       | 0.354        | 0.273       | 0.322          | 0.429         |
| Eg5          | 28 0.241       | 0.188        | 0.349       | 0.209        | 0.293       | 0.257          | 0.427         |
|              |                | 0.236        | 0.166       | 0.322        | 0.177       | 0.257          | 0.386         |
|              |                | 0.225        | 0.241       | 0.220        | 0.209       | 0.365          | 0.498         |
| Eg5 with an | 28 0.231       | 0.248        | 0.241       | 0.228        | 0.368       | 0.499          | 0.509         |
| alternative  |                | 0.225        | 0.241       | 0.220        | 0.209       | 0.365          | 0.498         |
| HIF-2α       | 42 0.061       | 0.209        | 0.137       | 0.395        | 0.104       | 0.173          | 0.297         |
|              |                | 0.052        | 0.145       | 0.082        | 0.311       | 0.054          | 0.156         |
|              |                | 0.112        | 0.172       | 0.597        | 0.759       | 0.531          | 0.626         |
|              |                | 0.066        | 0.095       | 0.436        | 0.544       | 0.382          | 0.490         |
|              |                | -0.034       | 0.188       | 0.157        | 0.385       | 0.169          | 0.360         |
| SHP-2        | 26 -0.106      | 0.087        | 0.145       | 0.331        | 0.114       | 0.304          |               |
|              |                | -0.126       | 0.003       | 0.102        | 0.249       | 0.052          | 0.200         |
|              |                | 0.110        | 0.199       | 0.206        | 0.297       | 0.178          | 0.208         |
| SYK          | 44 0.105       | 0.170        | 0.180       | 0.264        | 0.171       | 0.179          |               |
|              |                | 0.102        | 0.161       | 0.161        | 0.231       | 0.161          | 0.161         |
|              |                | 0.434        | 0.333       | 0.485        | 0.399       | 0.439          | 0.359         |
| TNKS2        | 27 0.414       | 0.325        | 0.418       | 0.383        | 0.374       | 0.343          |               |
|              |                | 0.394        | 0.314       | 0.331        | 0.359       | 0.354          | 0.297         |

Total 264 0.433 | 0.434 | 0.483 | 0.489 | 0.510 | 0.537 | 0.459 | 0.418
Table 4. RMSE Values in kcal/mol with 90% Confidence Intervals for All the Target Sets Using Different Free Energy Protocols

| Protein            | No. of ligands | MOE-MT-amber | MOE-MT-garf | cMD-MT-amber | cMD-MT-garf | CHMC-cMD-MT-amber | CHMC-cMD-MT-garf |
|--------------------|----------------|--------------|-------------|--------------|-------------|-------------------|-----------------|
| CDK8               | 33             | 2.049        | 1.317       | 1.371        | 1.457       | 1.311             | 1.434           |
|                    |                | 1.912        | 1.269       | 1.317        | 1.119       | 1.310             | 1.428           |
| c-Met              | 24             | 2.434        | 0.984       | 1.886        | 0.982       | 1.104             | 2.023           |
|                    |                | 2.259        | 0.890       | 1.811        | 0.874       | 1.846             | 1.160           |
| Eg5                | 28             | 4.863        | 1.951       | 1.515        | 1.297       | 1.879             | 1.225           |
|                    |                | 4.465        | 1.810       | 1.394        | 1.157       | 1.729             | 1.073           |
| Eg5 with an loop   | 28             | 4.320        | 1.598       | 1.656        | 1.530       | 1.772             | 2.002           |
|                    |                | 4.167        | 1.467       | 1.585        | 1.455       | 1.638             | 1.906           |
| HIF-2α             | 42             | 1.903        | 1.194       | 1.263        | 1.343       | 1.194             | 1.477           |
|                    |                | 1.738        | 1.128       | 1.209        | 1.086       | 1.359             | 1.386           |
| PFKFB3             | 40             | 2.610        | 1.435       | 0.858        | 0.588       | 0.774             | 0.833           |
|                    |                | 2.396        | 1.295       | 0.782        | 0.476       | 0.840             | 0.744           |
| SHP-2              | 26             | 2.257        | 1.504       | 1.501        | 1.501       | 1.201             | 1.264           |
|                    |                | 2.090        | 1.299       | 1.321        | 1.005       | 1.089             | 0.981           |
| SYK                | 44             | 4.653        | 1.565       | 2.173        | 1.463       | 2.704             | 1.769           |
|                    |                | 4.498        | 1.400       | 2.092        | 1.191       | 2.625             | 1.694           |
| TNKS2              | 27             | 3.652        | 1.979       | 0.821        | 0.924       | 1.933             | 1.162           |
|                    |                | 3.439        | 1.777       | 0.754        | 1.733       | 1.056             | 1.582           |
| Total              | 264            | 3.280        | 1.557       | 1.609        | 1.339       | 1.787             | 1.497           |
|                    |                | 3.107        | 1.394       | 1.434        | 1.116       | 1.611             | 1.289           |

The sampling protocols and the free energy calculations using the MT method were performed following the same protocols as discussed in the Method section. To better illustrate and compare the performance of different sampling protocols applied to the receptor–ligand complex state and to compensate for the lack of apo-state conformational sampling in the absolute binding free energy protocols, the raw MT results were rescaled. We selected 8 sets of protein–ligand cocrystal structures from the PDBBind database v2019 having the same 8 corresponding protein targets as the Merck benchmark but with no overlap in the ligand series selection. This curation yielded 191 cocrystal structures with known binding affinities which were passed to the MT_SameES protocol for fast binding free energy estimation (see details in the Supporting Information). The resulting regression lines for each target-ligand series were obtained by scoring against these cocrystal structures with both the MT-AMBERFF19SB and
Figure 5. Correlation between predicted binding free energies and experimental values for all eight target sets using the 6 different free energy protocols (For simplicity, the results for the Eg5 target set using the alternative loop are not included.).
the GARF energy functions, and the MT free energy calculation results ($\Delta G_{\text{pred}}$) were refitted to the corresponding regression lines as the rescaled binding free energy values ($\Delta G_{\text{pred}}^*$) to reproduce the experimentally determined free energies magnitudes.

$$\Delta G_{\text{pred}}^* = a \times \Delta G_{\text{pred}} + b$$

(14)

Figure 1B illustrates the calculation procedure for all the target sets from the Merck benchmark, and this procedure yields results for the 9 target sets (including the 8 primary targets along with the EG5 target with the alternative loop) which are summarized in Tables 2, 3, and 4 (Detailed plots for each set can be found in Figure S5). Generally, all our free energy protocols achieved acceptable to good correlations and error performance against the whole benchmark. The two MOEdock-based protocols using MT-AMBERFF19SB and GARF MT energy functions exhibit very similar correlation coefficients against the whole benchmark, with Pearson’s $R$ as 0.551 and 0.554 and Kendall’s $\tau$ as 0.519 and 0.515 for the MOE-MT-amber and MOE-MT-garf protocols, respectively. The MD simulations bring significant improvement to the MT binding free energy estimation accuracy with Pearson’s $R$ range from 0.520 to 0.658, and Kendall’s $\tau$ ranged from 0.459 to 0.510. Meanwhile, except for the MOE-MT-amber protocol, the rescaled results exhibited overall errors with the RMSE values which range between 1.2 and 1.5 kcal/mol.

When considering each individual target subset, the MOEdock-based and simulation-based MT protocols generate medium to good correlations for the CDK8, c-Met, Eg5, and TNKS2 sets with Pearson’s $R$ values ranging from 0.394 to 0.852. The additional sampling provided by MD leads to significant improvement in the binding free energy estimation for all target subsets. In particular, PFKFB3 benefits most from this additional sampling. For example, when considering the GARF potential, the Pearson’s $R$ values improve from 0.056 for MOE-MT-garf to 0.840 for cMD-MT-garf and 0.741 for CHMC-cMD-MT-garf. However, HIF-2α shows uneven performance in which the cMD-MT-garf protocol obtains Pearson’s $R = 0.507$, while all of the other protocols’ Pearson’s $R$ values are less than 0.3. Against the 9 target sets, the cMD-MT-amber protocol generated high correlations (Pearson’s $R \geq 0.6$) for 3 sets and medium correlations ($0.4 \leq \text{Pearson’s } R \leq 0.6$) for 3 sets, and the cMD-MT-garf protocol generated high correlations (Pearson’s $R \geq 0.6$) for 3 sets and medium correlations ($0.4 \leq \text{Pearson’s } R \leq 0.6$) for 6 sets. When adding the CHMC Monte Carlo protocol, the CHMC-cMD-MT-amber protocol generated high correlations (Pearson’s $R \geq 0.6$) for 5 sets and medium correlations ($0.4 \leq \text{Pearson’s } R \leq 0.6$) for 1 set, and the CHMC-cMD-MT-garf protocol generated high correlations (Pearson’s $R \geq 0.6$) for 4 sets and medium correlations ($0.4 \leq \text{Pearson’s } R \leq 0.6$) for 3 set. This was in stark contrast to the dock-only based versions for this benchmark in that without thorough binding mode sampling, the docking-based protocols did not achieve as good of a ranking performance. The MOE-MT-amber protocol generated high correlations (Pearson’s $R \geq 0.6$) for 1 set and medium correlations ($0.4 \leq \text{Pearson’s } R \leq 0.6$) for 4 sets, and the MOE-MT-garf protocol generated high correlations (Pearson’s $R \geq 0.6$) for 1 set and medium correlations ($0.4 \leq \text{Pearson’s } R \leq 0.6$) for 3 sets.

As shown in Table 5, the rescaling procedure - which was performed against 8 separately curated sets of protein–ligand complexes - made significant impact to the protein–ligand binding free energy ranking performance when evaluating the
whole benchmark including all the target sets. This improvement is observed both in the correlation coefficients and in error reduction. This is especially true for ranking across different target sets. Against the whole benchmark, utilizing rescaling increased the Pearson’s R and Kendall’s τ values for all protocols except for the two MOE docking-based protocols, and all the simulation-based protocols had reduced RMSE values after introducing rescaling.

Without rescaling, the raw estimation results showed reasonable binding free energy ranking performance for the test cases within most target sets, while struggling to rank among different target sets. Relative binding free energy estimation between protein—ligand complexes has always been a challenge for both absolute and relative binding free energy methods (especially when there were major structural differences between complex molecules). Reaction-pathway dependent free energy methods rely heavily on the free energy map simulation of the sampling windows along purposely designed alchemical structural mutation pathways or protein—ligand dissociation/binding pathways. Significant structural differences between protein—ligand molecular systems require simulations with respect to high-dimensional collective variables (CVs) which heavily burden the convergence of molecular simulations. On the other hand, end-point free energy protocols usually depend on the generation of conformational probabilities for representative energy states or numerical energy characterization of significant conformations. A major structural difference often introduces large, distinct conformational changing features among different molecules which negatively impacts the conformational state sampling thoroughness for bound-state protein, free-state protein, ligand molecules, and ultimately the conformational energy calculation accuracy.

The MT method improves accuracy by using a ratio of the bound-state and free-state partition functions for the estimation of the binding free energies. This way common computational errors within both bound-state and free-state partition functions cancel out during the free energy calculation procedure. To better capture the significant binding modes, we can utilize docking procedures (which are often limited to rigid-protein structures) or - as in the case of the present work - different MD simulation methods performed on the bound-state protein—ligand complexes to cover more sampling space. For each individual target set, we clearly observe improvement using the simulation-based MT protocols compared to the docking-based MT protocols as summarized in Tables 2, 3 and 4. However, the new MT protocols applied in this study still lack the ranking capability for the protein—ligand complexes with distinct protein targets which suggests that extra simulation procedures are needed to capture the free energy baselines introduced by different proteins. The rescaling procedure is a useful tool for the absolute binding free energy protocols to improve the binding free energies ranking capabilities across different protein targets. While simulation-based protocols all had their prediction errors reduced, the two MOEdock-based protocols had increased errors after introducing the rescaling procedure. This suggests that with the same protein and different ligand structures, compared to dock-based protocols, simulation-based protocols yield similar conformational and interaction features within and between binding site protein residues versus the corresponding protein target found in the cocrystal structures from rescaling data sets. Therefore, the energy

Table 6. Averaged Computational Time Monitored for All Target Sets from the Merck Benchmark

| protein | MOE-MT | cMD-MT | CHMC-cMD-MT |
|---------|--------|--------|-------------|
| CDK8    | 10     | 1409   | 855         |
| c-Met   | 9      | 1082   | 810         |
| Eg5     | 12     | 1174   | 249         |
| Eg5 with an alternative loop | 29 | 1166 | 233 |
| HIF-2α  | 5      | 742    | 310         |
| PFKFB3  | 9      | 1452   | 786         |
| SHP-2   | 23     | 1684   | 907         |
| SYK     | 16     | 1143   | 532         |
| TNKS2   | 6      | 860    | 362         |

baseline generated by the MOE IFD protocol for each protein target significantly deviates from the energy baseline from the corresponding target sets used in the rescaling procedure.

When considering the computational time for performing all the free energy protocols against every protein—ligand complex from the Merck benchmark, the MD simulation was performed on a computer with 4 × NVIDIA Tesla V100 16GB (900 GB/s) GPUs and 128GB Memory, and the MOEdock was performed on a computer with an E5-2640 20-Core v4 (2.4 GHz) and 128GB Memory. The average computing time required for every protein—ligand complex per target set in the Merck benchmark is provided in Table 6. As the fastest protocol, the IFD MOEdock procedure spent 5 to 29 min on average (per protein—ligand complex) for all the target sets. This is contrasted with the cMD-MT protocol which required roughly 1190 min or 19.83 h on average to simulate a 250 ns trajectory for each of the protein—ligand complexes from this benchmark. Finally, for CHMC sampling which also included 50 ns short-term cMD simulations, these calculations cost 560.44 min or 9.34 h on average for sampling one trajectory for each protein—ligand complex from the benchmark. Because the multithread simulations in the CHMC-cMD protocol were performed parallely, we presented the averaged time for every trajectory as the total time for the sampling workflow for each test case in the corresponding target set, noting that it was not an apples-to-apples computational time comparison because the parallel computing required more computational resources; but, employing multiple GPU cards was a simple way to support the parallel computing efficiency, while it was more expensive to improve the computational speed for the single-trajectory simulations. Overall, the CHMC-cMD protocol achieved close or even better free energy estimation accuracy compared to the 250 ns cMD protocol but at less simulation cost.
tions, we also discovered that particular binding modes from docking could be significantly altered through the MD simulations leading to significant differences in the binding free energy estimation accuracy for some target sets, e.g., the PFKFB2 set from this benchmark.

Binding mode generation using docking methods is largely dependent on the input structure's conformation along with the binding cavity on the protein surface. The IFD method only modestly refines the conformation of the residues in the binding site which contains the fitted ligand structure. When trying to evaluate the binding free energy with a single complex pose, important contacts could be missing, and unstable pairwise contacts may be considered important for maintaining the protein–ligand binding modes. This computational trial showed that the simulation-based free energy protocols sometimes significantly altered the conformation of the protein–ligand binding modes compared to the original docking poses, and some significant atom-pairwise interactions recognized by the docking protocol soon disappeared in the MD trajectories. Herein, we illustrate two example cases which reveal that the MD procedure often develops more reasonable binding modes leading to better binding free energy predictions compared to using docking procedures alone (target: Eg5–ligand: CHEMBL1084431 and target: PFKFB3–ligand: ligand_41).

For the Eg5 – CHEMBL1084431 protein–ligand molecular model, Figure 6 depicts the geometric difference between a representative binding mode selected from the converged cMD trajectory (Figures 6A and 6C) and the best scored docking pose from the MOE IFD protocol (Figures 6B and 6D). Comparing Figures 6A and 6B, we see that the binding site residues wrap tighter around the ligand molecule during the MD simulation versus the IFD protocol. Figures 6C and 6D show that both IFD and MD recognize the hydrogen bond between the ligand's N1 amine and the Gly103 backbone carbonyl group along with the hydrogen bond between the ligand's N2 amine and the Glu102 backbone carbonyl group. However, only the MD trajectory shows that the side chain carboxyl group of Glu104 swings down to form another hydrogen bond with the ligand's N1 amine. By reviewing the fraction of hydrogen bonds over the last 200 ns MD trajectory as reported in Table 7, we further discover that the N1 amine group moves back and forth to form several alternating hydrogen bonds with the surrounding residues. Specifically, when the N1 group moves to the left, it forms two hydrogen bonds with the Gly103 backbone and Glu104 side chain; but when the N1 group moves to the other side of the pocket, it forms one hydrogen bond with the Glu102 side chain. As a result, though the N2 group forms a stable hydrogen bond with the Glu102 residue (89.3% of the time through the last 200 ns), the N1 group was quite "bouncy" at the entrance of the binding pocket. This suggests that the ligand molecule, CHEMBL1084431, might not bind as tightly at the Eg5 binding site as shown from the MOE docking result. Against this complex, the MOE-MT-
Amber protocol generated the binding free energy as $-15.02$ kcal/mol. Given the unstable hydrogen bonds formed between the ligand’s N1 group and the receptor, the cMD-MT-Amber free energy protocol generated a much lower binding affinity as $-11.14$ kcal/mol (which agrees better with the experimental value as $-9.8$ kcal/mol).

For the PFKFB3–ligand_41 complex, there were two major differences in the binding modes generated by the two sampling protocols. In the MOE IFD best-scored pose, we observe that two residues at the PFKFB3 binding site act like a pair of “clips” clamping the quinoxaline group of the ligand in the middle: the Arg15-Gly16 backbone amide approaches the ligand’s quinoxaline by forming a $\pi$–$\pi$ stacking interaction, while Val187 provides the side-chain isopropyl group to form a C–H–$\pi$ interaction against the other side of the quinoxaline group. Notably, through the 250 ns MD simulation, the ligand’s quinoxaline group still exhibits close contacts with the Arg15-Gly16 backbone amide. However, the Val187 residue moves away from the ligand binding region as the beta-sheet containing it twists during the simulation. Meanwhile, Val213 moves to contact the ligand’s quinoxaline group and form a new C–H–$\pi$ interaction which replaces the Val187 residue. Another binding-mode difference is observed at the ligand’s quinoxaline binding region. For this interaction, the MOE IFD protocol places the ligand’s pyrazine ring in a location to form two hydrogen bonds with both the Ser122 and the Asn133 side chains of the receptor, while the hydrogen bond between Ser122 and the ligand’s pyrazine nitrogen is not observed in the converged MD trajectory. Instead, MD simulation shows that the Ser122 and Lys17 side chains gradually drew each other closer and form a hydrogen bond network including the Ser122 and Lys17 side chains and the Arg18 backbone. Moreover, by studying the hydrogen bond’s fraction over the last 200 ns MD trajectory, we find that the hydrogen bond between the Asn133 side chain and the ligand’s pyrazine nitrogen is not stable through the simulation, and only 18.3% of the snapshots in the last 200 ns MD trajectory had the hydrogen bond formed between Asn133 and the ligand’s pyrazine nitrogen. For binding affinity estimation, the MOE-MT-Amber protocol determined a binding free energy of $-11.36$ kcal/mol, and the cMD-MT-Amber free energy protocol recognized the binding affinity as $-7.72$ kcal/mol. This latter estimation agrees better with the experimental value of $-7.39$ kcal/mol (see Figure 7).

### Table 7. Hydrogen Bonding Percentages for the Eg5–CHEMBL1084431 and PFKFB3–ligand_41 Complexes over Their Last 200 ns MD Trajectories

| acceptor          | donor      | frac/% |
|-------------------|------------|--------|
| GLU_102@O        | LIG@N2     | 89.3   |
| GLY_103@O        | LIG@N1     | 26.7   |
| GLU_102@OE2      | LIG@N1     | 14.5   |
| GLU_102@OE1      | LIG@N1     | 3.8    |
| GLU_104@OE2      | LIG@N1     | 9.6    |
| GLU_104@OE1      | LIG@N1     | 7.8    |
| LIG@N1           | ASN133@ND2 | 18.3   |
| LIG@N3           | ASN133@ND2 | 6.6    |

**Figure 7.** Illustration of protein–ligand binding modes for the PFKFB3–ligand_41 complex. (A) Representative binding mode selected from the converged cMD trajectory. During the MD simulation, Val213 approached the quinoxaline group of the ligand to form a C–H–$\pi$ interaction, while Val187 moved away from the ligand. (B) During the MD simulation, no hydrogen bonding interaction between SER122 and the ligand is observed, while the hydrogen bonding interaction between ASN133 and the ligand was unstable. (C) The MOE IFD protocol placed the ligand at the binding site with the quinoxaline group close to the Val187 side chain, while Val213 was far from the ligand. (D) The best scored docking pose recognized two hydrogen bonding interactions between the ligand and the SER122 and ASN133 side chains.
The IFD protocol is focused on optimizing the intermolecular interaction potential energy by locally altering the conformations of the ligand and the receptor residues at the interface, but it is generally incapable of relaxing the intramolecular potential stress given the conformational complexity of most protein receptors. During the protein—ligand binding mechanism, certain intraprotein interactions are preferred over some possible significant intermolecular interactions to achieve lower overall potential energy for the whole complex. Moreover, sometimes a stable binding mode could only be achieved with large protein conformational changes which can scarcely be obtained from a docking protocol without the support from large-scale simulation. These two examples clearly illustrate that some significant protein—ligand interactions formed from the docking protocol are unstable or not even recognized during the MD simulation, and some significant protein—ligand interactions can only be formed through significant protein motif movement.

**Free Energy Estimation Using Single Conformation vs Conformation Ensembles from Simulation Trajectories.** With a more comprehensive evaluation of the intra- and intermolecular interactions, force fields usually lead to different binding modes than the docking generations alone. Moreover, simulations against many test cases in this work generated multiple converged conformational clusters, suggesting multiple energy minima during the protein—ligand recognition. From the conformational energy point of view, significant energy minima, especially the global minimal conformation, usually dominated the partition function of a canonical ensemble, and thus, we further explored the binding free energy estimation accuracy by only using the energy-minimum snapshot from the simulation trajectories of the bound-state protein—ligand complexes.

We proposed two rules for selecting the energy-minimum snapshot. From the 250 ns conventional MD trajectories of every protein—ligand complex, the first protocol directly selected the energy-minimum snapshot according to the force field computation. The logic for choosing the energy-minimum snapshot was straightforward, but the wrong snapshot might be picked due to errors in energy functions. The second protocol first chose the largest conformational cluster from the multitrajectory CHMC sampling with 50 ns MD simulations of a protein—ligand complex and then selected the energy-minimum snapshot with respect to this cluster. The parallel simulations using CHMC with MD simulation protocols were expected to cover more extensive conformational space. Assuming the multitrajectory simulation converged after 50 ns local sampling, the largest conformational cluster had a higher opportunity to contain the “global” energy minimum among all the sampled conformations. The snapshot selected from the simulations of every protein—ligand complex was then treated with the MTScoreES protocol for binding free energy estimation.

In the following analysis, we used the term “snapshot-cMD-MT-amber” denoting the binding free energy estimation using the snapshot selected from the 250 ns MD simulation and treated with the MTScoreES protocol using the Amber energy function; the term “snapshot-cMD-MT-garf” denoting the binding free energy estimation using the snapshot selected from the CHMC parallel sampling followed by 50 ns conventional MD simulation and treated with the MTScoreES protocol using the GARF energy function.

Against the CASF benchmark, free energy estimation results using single snapshots were concluded in Table 8. It was not surprising that the prediction accuracy and ranking performance were not as good as using the conformational ensemble-based protocols (See the corresponding results in Table 1.). Moreover, the results illustrated that binding free energy estimation using more thorough sampling and converged conformational ensembles from the simulation outperformed the binding affinity prediction using single conformations or docking poses. It was also worth noting that snapshots selected from the largest conformational cluster generated better binding free energy prediction results compared to using the energy-minimum snapshot from the 250 ns single-trajectory simulation. This suggested that the high-probable conformation generated from the MD simulation may be a better representative conformational state than the energy minimum selected from the simulation trajectory.

Table 9 concluded the statistical results for the free energy estimation using single snapshots for the Merck benchmark. For the eight target sets, both the global minimal snapshots and the minimal snapshots from the largest conformational clusters had comparable binding affinity prediction accuracy. This suggested that both the single-trajectory and parallel simulations converged at similar conformational states.

When comparing to the conformational ensemble-based free energy estimation protocols, we discovered that the all protocols had comparable ranking capability against the CDK8, c-MET, Eg5, and TNKS2 sets. On the other hand, for the HIF-2α, PFKFB3, SHP-2, and SYK sets, the single-snapshot-based protocols generally had higher correlation coefficients (Pearson’s R and Kendall’s τ values) compared to the docking-based protocols but were outperformed by the conformational ensemble-based MD protocols (cMD-MT and CHMC-cMD-MT-protocols). It was likely that the protein—ligand complexes in the CDK8, c-MET, Eg5, and TNKS2 sets had dominantly significant global minimal conformational states and that binding free energy estimators using single conformations or conformational ensembles from docking and MD simulations all had generally acceptable accuracy. Against the HIF-2α, PFKFB3, SHP-2, and SYK sets, the MD simulation significantly altered the protein—ligand binding modes from the docking poses. Single snapshots from the MD simulations generated more reasonable binding affinities.

### Table 8. Statistical Results for Protein-Ligand Binding Free Energy Estimation Using Single Snapshots Treated with Different MT-Based Protocols

| MT protocols | Pearson’s R | Kendall’s τ | MAE (kcal/mol) |
|--------------|-------------|-------------|----------------|
| snapshot-cMD-MT-amber | 0.59 | 0.43 | 2.22 |
| snapshot-cMD-MT-garf | 0.63 | 0.45 | 1.86 |
| snapshot-CHMC-cMD-MT-amber | 0.63 | 0.46 | 2.32 |
| snapshot-CHMC-cMD-MT-garf | 0.65 | 0.47 | 1.78 |

https://doi.org/10.1021/acs.jcim.2c00278  
J. Chem. Inf. Model. 2022, 62, 5645–5665
Table 9. Statistical Results for All the Target Sets from the Merck Benchmark Using Single Snapshots Treated with Different Free Energy Protocols

| Ligand | Pearson’s $R$ | Kendall’s $τ$ | RMSE (kcal/mol) | Pearson’s $R$ | Kendall’s $τ$ | RMSE (kcal/mol) | Pearson’s $R$ | Kendall’s $τ$ | RMSE (kcal/mol) | Pearson’s $R$ | Kendall’s $τ$ | RMSE (kcal/mol) |
|--------|----------------|----------------|-----------------|----------------|----------------|-----------------|----------------|----------------|-----------------|----------------|----------------|----------------|
| CDK8   | 0.448          | 0.552          | 1.269           | 0.379          | 0.456          | 1.685           | 0.667          | 0.539          | 1.962           | 0.405          | 0.505          | 1.928           |
|        |                 |                |                 |                |                 |                 |                |                |                 |                |                 |                 |
| c-MET  | 0.715          | 0.724          | 1.685           | 0.667          | 0.596          | 1.685           | 0.667          | 0.596          | 1.685           | 0.405          | 0.464          | 1.928           |
|        |                 |                |                 |                |                 |                 |                |                |                 |                |                 |                 |
| Eg5    | 0.367          | 0.403          | 1.928           | 0.072          | 0.169          | 2.494           | 0.367          | 0.403          | 1.387           | 0.367          | 0.403          | 1.387           |
|        |                 |                |                 |                |                 |                 |                |                |                 |                |                 |                 |
| Eg5 with an alternative loop | | | | | | | | | | | | |
|RMSE (kcal/mol) | 2.024 | 2.180 | 1.816 | 0.121 | 0.183 | 0.113 | 0.270 | 0.354 | 0.212 | 1.176 | 0.429 | 1.054 |
|Kendall’s $τ$ | 0.196 | 0.270 | 0.140 | 0.030 | 0.270 | 0.140 | 0.196 | 0.270 | 0.140 | 0.030 | 0.270 | 0.140 |
|PFKFB3 | 0.414 | 0.596 | 0.395 | 0.150 | 0.357 | 0.056 | 0.176 | 1.624 | 1.054 | 1.249 | 0.307 | 0.312 |
|        |                 |                |                 |                |                 |                |                |                |                 |                |                 |                 |
| SHP-2  | 0.150          | 0.357          | 1.176           | 0.102          | 0.290          | 1.379           | 0.150          | 0.357          | 1.176           | 0.102          | 0.290          | 1.379           |
|        |                 |                |                 |                |                 |                |                |                |                 |                |                 |                 |
| SYK    | 0.250          | 0.252          | 2.535           | 0.156          | 0.122          | 2.535           | 0.250          | 0.252          | 2.535           | 0.156          | 0.122          | 2.535           |
|        |                 |                |                 |                |                 |                |                |                |                 |                |                 |                 |
| TNKS2  | 0.567          | 0.350          | 2.943           | 0.335          | 0.167          | 2.943           | 0.567          | 0.350          | 2.943           | 0.335          | 0.167          | 2.943           |
|        |                 |                |                 |                |                 |                |                |                |                 |                |                 |                 |
| total  | 0.509          | 0.470          | 2.171           | 0.384          | 0.341          | 2.171           | 0.509          | 0.470          | 2.171           | 0.384          | 0.341          | 2.171           |

Compared to the docking predictions, and the conformational ensembles from simulation further improved the binding affinity prediction accuracy.

**CONCLUSION**

In this study, we introduced two simulation-based conformational sampling protocols including (1) 250 ns single-trajectory Amber molecular dynamics and (2) CHMC Monte Carlo sampling followed by 50 ns multireplica Amber MD simulation and contrasted these with the previously published MOE induced-fit docking protocol. Each ensemble of protein–ligand complexes was generated and used to explore the free energy estimation accuracy and efficiency of the Movable Type absolute binding free energy method using different conformational sampling protocols. Compared to dock-based MT protocols, simulation-based MT protocols achieve significant improvement in binding free energy estimation accuracy against both the CASF-2016 and the Merck KGaA benchmarks. The CASF-2016 benchmark contains a wide variety of test cases with different structures for both ligands and proteins. The two simulation-based protocols using the GARF energy function obtained good prediction accuracy with Pearson’s $R \geq 0.7$, Kendall’s $τ = 0.52$, and an MAE < 2 kcal/mol, and the two simulation-based protocols using the MT-AMBERFF19SB force field had lower correlation coefficients with Pearson’s $R = 0.66$, Kendall’s $τ = 0.49$ and higher errors with an MAE = 2.10 kcal/mol. Against protein–ligand complexes with identical protein targets from the Merck benchmark, our simulation-based protocols obtained acceptable ($0.4 \leq$ Pearson’s $R \leq 0.6$) to good (Pearson’s $R \geq 0.6$) ranking performances for more than 2/3rd of all the target sets. With the help of free energy prediction rescaling using a separate set of protein–ligand cocrystal structures, Pearson’s $R$ values for all the protocols against the whole Merck benchmark were improved from 0.195–0.372 to 0.551–0.658 after rescaling, and RMSE values for all the simulation-based protocols were within 2 kcal/mol. Conversely, the RMSE values for dock-based protocols increased when introducing rescaling suggesting that the binding-site protein residues captured by the docking protocols deviated from the cocrystal structures in the corresponding rescaling set. As an absolute binding free energy method which relies on numerical energy performed on ensembles of conformational states, the MT protocol is heavily dependent on conformational sampling for protein–ligand complex systems. The present work clearly shows that utilizing MD protocols greatly improves binding free energy estimation accuracy. However, extensive simulation did not always result in better performance, and the MT-CHMC method, which employs multiple reasonable initial poses performed with parallel short-term simulations, reached similar estimation accuracy using less than half of the computational time. Since the method is compatible with many different conformational sampling methods, the present
work shows that MT is a flexible free energy tool for protein–ligand binding free energy estimation tasks.

**ASSOCIATED CONTENT**

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00278.

- Detailed explanation of Movable Type binding free energy estimation protocol (PDF)
- Free energy calculation results for CASF-2016 benchmark, rescale set, and Merck KGaA benchmark using cMD-MT-amber, cMD-MT-garf, CHMC- CMD-MT-amber, and CHMC-CMD-MT-garf protocols (PDF)

**AUTHOR INFORMATION**

### Corresponding Author

Zheng Zheng — School of Chemistry, Chemical Engineering and Life Science, Wuhan University of Technology, Wuhan 430070, PR China; QuantumBio Inc., State College, Pennsylvania 16801, United States; orcid.org/0000-0001-5221-3209; Email: laozhengzz@gmail.com

### Authors

- Wenlang Liu — School of Chemistry, Chemical Engineering and Life Science, Wuhan University of Technology, Wuhan 430070, PR China
- Zhenhao Liu — School of Chemistry, Chemical Engineering and Life Science, Wuhan University of Technology, Wuhan 430070, PR China
- Hao Liu — School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, PR China
- Lance M. Westerhoff — QuantumBio Inc., State College, Pennsylvania 16801, United States

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00278

### Notes

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

Z.Z. acknowledges support from the National Natural Science Foundation of China (Grant No. 21903061). L.M.W. acknowledges being supported in part by the National Institute of General Medical Sciences of the National Institutes of Health under Small Business Innovative Research (SBIR) Award R44GM134781. The authors also acknowledge the continued support of Michigan State University and MSU Technologies for licensing the MovableType technology to QuantumBio, Inc. We also thank Drs. Roger I. Martin, Oleg Borbulevych, Nupur Bansal, and Kenneth M. Merz, Jr. for their work and support in the prior study and the transfer and implementation of the MT technology in DivCon.

### REFERENCES

- (1) Schneider, G. Virtual screening: an endless staircase? Nat. Rev. Drug Discovery 2010, 9 (4), 273–276.
- (2) Slivoski, G.; Kothiwale, S.; Meiler, J.; Lowe Jr, E. W. Computational methods in drug discovery. Pharmacol. Rev. 2014, 66 (1), 334–395.
- (3) Dror, O.; Schniedman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Novel approach for efficient pharmacophore-based virtual screening: method and applications. J. Chem. Inf. Model 2009, 49 (10), 2333–2343.
- (4) Malmstrom, R. D.; Watowich, S. J. Using free energy of binding calculations to improve the accuracy of virtual screening predictions. J. Chem. Inf. Model 2011, 51 (7), 1648–1655.
- (5) Klebe, G. Applying thermodynamic profiling in lead finding and optimization. Nat. Rev. Drug Discovery 2015, 14 (2), 95–110.
- (6) Albanese, S. K.; Chodera, J. D.; Volkamer, A.; Keng, S.; Abel, R.; Wang, L. Is Structure-Based Drug Design Ready for Selectivity Optimization? J. Chem. Inf. Model 2020, 60 (12), 6211–6227.
- (7) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz Jr, K. M. The role of quantum mechanics in structure-based drug design. Drug Discovery Today 2007, 12 (17–18), 723–734.
- (8) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. Annu. Rev. Biophys. Biomol. Struct. 2003, 32, 335–373.
- (9) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 1982, 161 (2), 269–288.
- (10) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discovery 2004, 3 (11), 935–949.
- (11) Slezd, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. Curr. Opin. Struct. Biol. 2018, 48, 93–102.
- (12) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. J. Med. Chem. 2016, 59 (9), 4035–4061.
- (13) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. J. Chem. Inf. Model 2009, 49 (4), 1079–1093.
- (14) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. J. Chem. Inf. Model 2014, 54 (6), 1717–1736.
- (15) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. J. Chem. Inf. Model 2019, 59 (2), 895–913.
- (16) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat. Protoc. 2016, 11 (5), 905–919.
- (17) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 2009, 31 (2), 455–461.
- (18) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. Proteins 2003, 52 (4), 609–623.
- (19) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J. Med. Chem. 2006, 49 (21), 6177–6196.
- (20) Jain, A. N. Sulfex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J. Med. Chem. 2003, 46 (4), 499–511.
- (21) Clark, A. J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E. B.; Abel, R.; Friesner, R. A.; Berne, B. J. Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. J. Chem. Theory Comput. 2016, 12 (6), 2990–2998.
- (22) Merz, K. M., Jr. Limits of Free Energy Computation for Protein-Ligand Interactions. J. Chem. Theory Comput. 2010, 6, 1769–1776.
- (23) Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. Interdiscip. Sci. 2019, 11 (2), 320–328.
Williams, S. L.; Zhang, J. H.; LaMarche, M. J. Allosteric Inhibition of SHP2: Identification of a Potent, Selective, and Orally Efficacious Phosphatase Inhibitor. J. Med. Chem. 2016, 59 (17), 7773−7782.

(51) Chen, Y. N.; LaMarche, M. J.; Chan, H. M.; Fekkes, P.; Garcia-Fortanet, J.; Acker, M. G.; Antonakos, B.; Chen, C. H.; Chen, Z.; Cooke, V. G.; Dobson, J. R.; Deng, Z.; Fei, F.; Firestone, B.; Fodor, M.; Fridrich, C.; Gao, H.; Grunenfelder, D.; Hao, H. X.; Jacob, J.; Ho, S.; Hsiao, K.; Kang, Z. B.; Karki, R.; Kato, M.; Larrow, J.; La Bonte, L. R.; Lenoir, F.; Liu, G.; Liu, S.; Majumdar, D.; Meyer, M. J.; Palermo, M.; Perez, L.; Pu, M.; Price, E.; Quinn, C.; Shellya, S.; Shultz, M. D.; Slisz, J.; Venkatesan, K.; Wang, P.; Warmuth, M.; Williams, S.; Yang, G.; Yuan, J.; Zhang, J. H.; Zhu, P.; Ramsey, T.; Keen, N. J.; Sellers, W. R.; Stams, T.; Fortin, P. D. Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. Nature 2016, 535 (7610), 148−152.

(52) Currie, K. S.; Kropf, J. E.; Lee, T.; Blomgren, P.; Xu, J.; Zhao, Z.; Gallion, S.; Whitney, J. A.; Maclin, D.; Lansdon, E. B.; Maciejewski, P.; Rossi, A. M.; Rong, H.; Macaluso, J.; Barbosa, J.; Di Paolo, J. A.; Mitchell, S. A. Discovery of GS-9973, a selective and orally efficacious inhibitor of spleen tyrosine kinase. J. Med. Chem. 2014, 57 (9), 3856−3873.

(53) Buchstaller, H. P.; Anlauf, U.; Dorsch, D.; Kuhn, D.; Lehmann, M.; Leuthner, B.; Musil, D.; Radtki, D.; Ritzert, C.; Rohdich, F.; Schneider, R.; Esdar, C. Discovery and Optimization of 2-Arylquinazolin-4-ones into a Potent and Selective Tankyrase Inhibitor Modulating Wnt Pathway Activity. J. Med. Chem. 2019, 62 (17), 7897−7909.

(54) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J. Comput. Chem. 2000, 21 (2), 132−146.

(55) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J. Comput. Chem. 2002, 23 (16), 1623−1641.

(56) Zheng, Z.; Wang, T.; Li, P.; Merz, K. M., Jr. KCESA-Movable Type Implicit Solvation Model (KMTISM). J. Chem. Theory Comput. 2015, 11 (2), 667−682.