JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

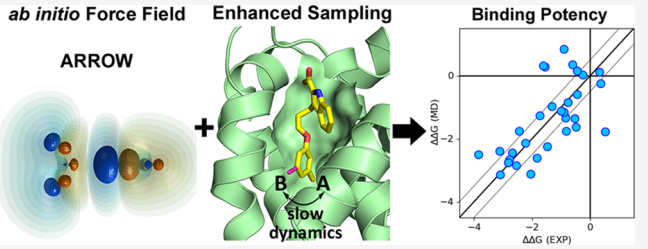
Protein–Ligand Binding Free-Energy Calculations with ARROW—A Purely First-Principles Parameterized Polarizable Force Field

Grzegorz Nawrocki,* Igor Leontyev, Serzhan Sakipov, Mikhail Darkhovskiy, Igor Kurnikov, Leonid Pereyaslavets, Ganesh Kamath, Ekaterina Voronina, Oleg Butin, Alexey Illarionov, Michael Olevanov, Alexander Kostikov, Ilya Ivahnenko, Dhilon S. Patel, Subramanian K. R. S. Sankaranarayanan, Maria G. Kurnikova, Christopher Lock, Gavin E. Crooks, Michael Levitt, Roger D. Kornberg, and Boris Fain*

Cite This: https://doi.org/10.1021/acs.jctc.2c00930 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Protein–ligand binding free-energy calculations using molecular dynamics (MD) simulations have emerged as a powerful tool for in silico drug design. Here, we present results obtained with the ARROW force field (FF)—a multipolar polarizable and physics-based model with all parameters fitted entirely to high-level ab initio quantum mechanical (QM) calculations. ARROW has already proven its ability to determine solvation free energy of arbitrary neutral compounds with unprecedented accuracy. The ARROW FF parameterization is now extended to include coverage of all amino acids including charged groups, allowing molecular simulations of a series of protein–ligand systems and prediction of their relative binding free energies. We ensure adequate sampling by applying a novel technique that is based on coupling the Hamiltonian Replica exchange (HREX) with a conformation reservoir generated via potential softening and nonequilibrium MD. ARROW provides predictions with near chemical accuracy (mean absolute error of ~0.5 kcal/mol) for two of the three protein systems studied here (MCL1 and Thrombin). The third protein system (CDK2) reveals the difficulty in accurately describing dimer interaction energies involving polar and charged species. Overall, for all of the three protein systems studied here, ARROW FF predicts relative binding free energies of ligands with a similar accuracy level as leading nonpolarizable force fields.



INTRODUCTION

Free-energy calculations of ligand binding to a protein can serve as a powerful tool for structure-based small-molecule drug design, especially at the stages of lead selection (hit-to-lead) and lead optimization, where ligands of high binding affinity are desired. A change of free energy, ΔG, upon binding quantitatively describes ligand–protein affinity. In silico calculations of protein–ligand binding ΔG have numerous advantages over expensive experimental approaches. The calculations can be performed in a fully automated manner; consequently, a large number of ligands can be evaluated and numerous drug candidates with diverse structures can be selected. The relative energy, $\Delta\Delta G$, i.e., calculation of ΔG of one ligand with respect to another, is usually sufficient to guide the process of ligand optimization and can be calculated more accurately than the absolute energy via alchemical transformations.[1,2]

In recent years, calculations of binding free energies of ligands in proteins, using all-atom force fields, have grown in importance. Effective software packages have been developed to perform such calculations.[3−5] All-atom force fields used in these calculations[6−8] typically combine parameters derived from quantum mechanics (QM) calculations and empirical parameters fitted to reproduce certain experimental observables. However, systematic studies performed on a large set of protein–ligand systems suggest that current methodologies may have reached a limit of about 1 kcal/mol mean absolute error (MAE) from the experiment.[9−12] There are two factors responsible for the low accuracy, insufficient quality of molecular force fields (FF), and poor conformational sampling.

Most of the currently available force fields are all-atom fixed-charge models, e.g., AMBER/GAFF,[6] CHARMM/CGenFF,[13] OPLS,[10,14,15] GROMOS,[16] and MMFF.[17] Although they are well established, highly refined, and computationally efficient, it is generally accepted that they are not sophisticated enough to describe complicated protein–ligand interactions with an accuracy required for drug design. Specifically, they do not

Received: September 9, 2022

allow an accurate description of electrostatic and exchange interactions in the proteins, or nonadditive effects of atomic interactions, e.g., electronic polarizability. The absence of electronic polarizability significantly limits the force field's ability to correctly describe highly heterogeneous environments such as protein active sites. For the same reason, these force fields are marked by low transferability of parameters.

A solution for these issues would be to develop more advanced models that are physics-based and explicitly represent nonadditive effects. Thanks to recent developments of computational hardware, especially graphics processing units (GPU), simulations of these models are currently perfectly feasible on the size and time scales required by protein−ligand systems. Additionally, more robust procedures for fitting analytical expressions to molecular potential surfaces and advanced analytical expressions themselves are available. Thus, a few polarizable force fields are under continuous development, e.g., AMOEBA[18] or CHARMM-Drude.[19] Nevertheless, they do not fully meet the expectations in the field of drug design, such as high accuracy, transferability, universality, or productivity. An increasing number of parameters in polarizable FF makes it harder to derive them unambiguously from the available experimental information. Consequently, applications of polarizable force fields to model complex systems, such as protein−ligand complexes, are relatively rarely reported in the literature.[20−23] Reported ligand-binding free-energy calculations with AMOEBA FF showed lower accuracy[20] for some systems than calculations with simpler nonpolarizable FF.

An appealing alternative to empirical force fields are force fields that rely purely on QM calculations, e.g., MMFF94,[17] MB-pol,[24] QMPFF,[25,26] QMFF,[27,28] QMDFF,[29] or QMPFF3.[26] Recently, these kinds of force fields have gained traction as accurate but computationally intensive QM calculations, such as those based on coupled-cluster (CCSDT) method, have become more accessible. QM calculations can provide detailed insights into energy interactions within and between the individual particles and also for the individual energy components based on suitable decomposition schemes.[30] Thus, QM can be used to parameterize advanced physics-based models based on their energy components, e.g., electrostatics, exchange, induction, or dispersion, can be fitted separately. Development of models with terms that have physical interpretation is essential for force field transferability, and high transferability is especially required from QM-based force fields since QM calculations are feasible only for small fragments of compounds.

In this paper, we present results obtained with the ARROW force field. It is an advanced physics-based model that includes multipolar electrostatics and anisotropic polarization. Additionally, its parameters are fitted exclusively to high-level QM data for a set of small compound monomers and dimers, without fitting to any experimental data. We already have shown that ARROW FF provides ΔG of solvation for arbitrary neutral molecules with unprecedented accuracy.[31] Here, we expand the coverage to all standard amino acids (neutral and charged) and limited ligand chemical phase space to make protein−ligand simulations feasible. We probe the accuracy of ARROW FF by calculating the relative binding free energies ($\Delta\Delta G$) for a series of ligands in MCL1, Thrombin, and CDK2 proteins for which experimental results are known. These systems are also related to real drug design projects and tend to be benchmark studies for various research groups for testing their molecular models.

An equally vital component of any ΔG prediction is thorough sampling of configurational space of a simulated system. Successful sampling is especially important for protein−ligand systems that often exist in multiple binding states. Numerous enhanced sampling techniques have been developed to make more states accessible during the molecular dynamics simulations, e.g., umbrella sampling,[32] TREX (Temperature Replica Exchange), HREX (Hamiltonian Replica Exchange),[33] REST1, and REST2 (Replica Exchange with Solute Tempering).[34,35] To ensure adequate conformational sampling, we developed and applied an enhanced sampling technique, a modified HREX coupled to a conformation reservoir generated through softening of the molecular potential, and a nonequilibrium (NEQ) MD.

To successfully disseminate the computational methodology presented in this paper to applications in the pharmaceutical industry, we have developed a user-friendly software package that facilitates ligand parameterization, system setup, running simulations, and their analysis. A key module of the package is ARBALEST,[36] an MD simulation program that supports the ARROW force field. ARBALEST is capable of running simulations on computer clusters with multiple CPUs and GPUs. It allows a user to perform free-energy calculations and use various enhanced sampling techniques including those described here.

THEORY

Generation of Conformation Reservoir Using Nonequilibrium MD. Recently, nonequilibrium MD-based techniques were used by several research groups for $\Delta\Delta G$ calculations[37] and enhanced conformational sampling of ligand-binding complexes.[38] Free energies for alchemical transitions were calculated from work values computed from multiple nonequilibrium MD runs as the Hamiltonian of the system gradually changes from one state to another in one or both directions.[39,40] Enhanced sampling on a rugged potential energy landscape of protein−ligand complexes was attempted by Gill et al.[38] using Nonequilibrium Candidate Monte Carlo (NCMC) moves. NCMC moves consist of nonequilibrium MD runs with the Hamiltonian of the system changing from nonsoftened to a softened state, with an addition of a regular MD stretch in the softened state of the Hamiltonian, followed by a reverse NEQ MD to the nonsoftened state of the Hamiltonian. An enhanced sampling is achieved due to fast interconversions of torsional states of a ligand in MD simulations with "softened" Hamiltonian. NEQ MD work calculations give proper acceptance probabilities for such MC moves that preserve the Boltzmann distribution of molecular system geometries.

In this paper, we use the methodology that can be considered as a modification of the NCMC sampling procedure.[38] Instead of complex MC moves described above, we are performing nonequilibrium runs from snapshots of equilibrated MD obtained with the softened Hamiltonian, changing the system Hamiltonian from the softened state to the regular (nonsoftened) state. Molecular system geometries at the end of the NEQ MD runs that pass a Metropolis-like acceptance criterion are used to prepare conformation reservoirs for HREX $\Delta\Delta G$ calculations.

First, a sufficiently long MD trajectory with the "softened" Hamiltonian that samples the ligand and protein conforma-

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

tions is generated. The "softened" Hamiltonian is constructed to reduce the potential barriers between the local minima and increase the inter-minima transition rates. For the chosen parameters, 10 ns long MD runs were sufficient to get an equilibrated ensemble of configurations in the "softened" Hamiltonian. To generate a Boltzmann-distributed ensemble of configurations in the original physical (nonsoftened) Hamiltonian, we run a set of nonequilibrium MD simulations starting from a set of geometries in the ensemble of the softened Hamiltonian and filtering the final conformations using a criterion that is based on nonequilibrium work. During the course of a nonequilibrium MD run, the molecular potential quickly changes from the "softened" to "nonsoftened" state. The work for the process is computed as the integral

$$W = \int \frac{\mathrm{d}H}{\mathrm{d}\lambda}\,\mathrm{d}\lambda \tag{1}$$

A Metropolis test is performed for the end point of the NEQ MD trajectory that determines whether to add the end system configuration to the reservoir or not

$$\text{accept if } \exp\left(-\frac{W - \Delta G_{s\to ns}}{kT}\right) > \xi \tag{2}$$

where ξ is a random number from [0,1] interval and $\Delta G_{s\to ns}$ is the free energy of the transition between softened and nonsoftened Hamiltonians. $\Delta G_{s\to ns}$ is computed iteratively.

The first approximation of $\Delta G_{s\to ns}$ is obtained using the Jarzynski equality

$$\exp\left(-\frac{\Delta G}{kT}\right) = \left\langle \exp\left(-\frac{W}{kT}\right)\right\rangle \tag{3}$$

where $\langle\ \rangle$ denotes averaging over all of the nonequilibrium MD runs, W is the work computed for a nonequilibrium run, thus

$$\Delta G_{s\to ns} = -kT\,\ln\left(-\left\langle \exp\left(-\frac{W}{kT}\right)\right\rangle\right) \tag{4}$$

It is expected that an approximation to $\Delta G_{s\to ns}$ obtained with the Jarzynski equality (eq 4) is not very accurate and a better estimate for $\Delta G_{s\to ns}$ can be obtained using the Maximum-Likelihood method[41] based on bidirectional nonequilibrium runs, the Bennett acceptance ratio method and the Crooks fluctuation theorem

$$\frac{P_{s\to ns}(W)}{P_{ns\to s}(W)} = \exp\left(\frac{W - \Delta G}{kT}\right) \tag{5}$$

In the Maximum-Likelihood method, $\Delta G_{s\to ns}$ is found by solving the equation

$$\sum_{i=1}^{n_F} \frac{1}{1 + \exp\left(\frac{M + W_i - \Delta G}{kT}\right)} - \sum_{j=1}^{n_R} \frac{1}{1 + \exp\left(\frac{M + W_j - \Delta G}{kT}\right)} = 0 \tag{6}$$

where $M = kT\,\ln\left(\frac{n_F}{n_R}\right)$ and n_F and n_R are the numbers of forward and reverse nonequilibrium MD runs, respectively.

We use values of $\Delta G_{s\to ns}$ obtained in eq 4 to initially filter the end point conformations of NEQ MD runs using the criteria in eq 2. Then, we use these configurations as starting points to run NEQ MD from nonsoftened to the softened Hamiltonian state. Work values obtained in MD runs in both directions are combined, solving eq 6, and the Bennett acceptance ratio (BAR) approach[42] is used to obtain an improved estimate of $\Delta G_{s\to ns}$ that we plug into eq 2 to obtain an improved equilibrium conformation distribution for the nonsoftened Hamiltonian. The Maximum-Likelihood equation (eq 6) can be solved again to correct values of $\Delta G_{s\to ns}$ based on an updated set of accepted conformations of nonsoftened Hamiltonian. We found that a few iterations were sufficient to converge $\Delta G_{s\to ns}$ to 0.1−0.2 kcal/mol accuracy so that the computed distribution of accepted configurations of nonsoftened Hamiltonian did not significantly change on further iterations.

METHODS

Quantum Mechanics. In our work, we use a variety of quantum mechanical data as a benchmark for energies and conformations. QM calculations were performed for the monomer model compounds at the MP2/aug-cc-pVQZ level and dimers of the ligand model compounds with amino acid fragments and water were computed with the silver standard, i.e., MP2/CBS, calculated with Helgaker cubic extrapolation from aug-cc-pVTZ->aug-cc-pVQZ as well as post-MP2 correction (i.e., plus CCSD(T)/aug-cc-pVDZ−MP2/aug-cc-pVDZ). More details on QM can be found in the Supporting Information and in our previous publication.[31]

Force Field. In the ARROW force field, the nonbonded interactions are composed of electrostatic, exchange-repulsion, and dispersion terms. The electrostatic and exchange-repulsion terms are multipolar with inclusion of charges, dipoles, and quadrupoles, and their radial dependence is a Slater-like exponential so that they well describe charge penetration effects. The dispersion term is conventionally represented by spherical terms (C6 and C8), and a Tang−Toennies-damped interaction. Many-body effects are modeled by anisotropic-induced dipoles interacting with the electrostatic and exchange-repulsion terms, as well as with one another. They are iterated to self-consistent field convergence on every nonbonded step. Additional description of the ARROW force field can be found in the SI. For a detailed description, including functional forms, the reader is referred to our previous works.[31,43]

Parameterization. Proteins and ligands were split into chemical functional groups. Their intermolecular parameters were determined by agreement with QM values of dimer and multimer energies, electrostatic potentials, multipole moments of monomers, polarization tensor, and interaction of fragments with point charges. To aid transferability, we also attempted to match the individual FF energy components to their corresponding QM counterparts, in addition to reproducing the total energy. The typical size of the fragments was not bigger than 10 heavy atoms, e.g., phenol. Larger molecules were built by joining together smaller fragments. We assume that all interactions except electrostatic stay the same (e.g., like in GAFF or AMOEBA) and we refine multipoles on the boundary atoms (e.g., boundary atoms in biphenyl when two benzene rings are merged) to have the best fit to the electrostatic potential around the merged place using the RESP[44] procedure that is applied to charges, dipoles, and multipoles. For this, we perform QM calculations of lower quality on joined molecular pieces of two fragments. Typically, we include fragments where boundary atoms are either hydrogens attached to carbons or carbons due to their typically more neutral charges in comparison to other more

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

electronegative elements, e.g., oxygen, nitrogen, chlorine, etc. The details of nonbonded parameterization have been described in our previous publication.[31]

Molecular Dynamics. MD simulations were performed with the ARBALEST simulation package using multiple CPUs with OpenMP and MPI libraries and NVIDIA graphics processing units (GPU) with CUDA library.[45] For long-range electrostatic interactions, Particle Mesh Ewald (PME)[46,47] was used. Dispersion and PME direct sum electrostatic interactions were cutoff at 9 Å distance. A multiple time step algorithm was used to integrate the equations of motion.[48] The system temperature was kept at 298 K using the Nosé−Hoover thermostat.[49] Pressure was maintained at 1 atm using the Berendsen barostat.[50]

Systems Setup. The following structures were used to set up three protein−ligand systems for MD simulations: MCL1 (PDB: 4hw2), Thrombin (PDB:2zc9), and CDK2 (PDB:1h1q). Missing residues and side chains in Thrombin were modeled using the Swiss-Model server.[51] Protonation states of protein residues were determined using PropKa.[52] Each complex was centered and aligned along its principal axes in a rectangular simulation box. The size of the box was adjusted to leave at least 5 Å distance between the protein and the box edges. The systems were solvated using tools from the GROMACS package.[53] Some water molecules inserted into the binding pocket were manually removed.

Systems Equilibration. The solvated protein−ligand systems were equilibrated in two steps. In the first step, all heavy atoms of protein, ligand, and crystallographic water were positionally restrained (k = 2.5 kcal/mol/Å^2). The potential energy of the system was minimized and the system equilibrated for 0.5 ns in the NVT ensemble. In the second step, only the Cα atoms that were farther away than 7 Å from the ligand were restrained. These restraints were maintained in all of the following production and reservoir generation simulations. The energy of the system was minimized again and the system was equilibrated for another 2 ns in the NPT ensemble.

Free-Energy Calculations. An alchemical transformation method was used for calculations of the free-energy change, $\Delta G_{R\to T}$, associated with mutation of the reference ligand, R, to the target ligand, T. In this method, the Hamiltonian of the reference ligand, H_R, is incrementally transformed to the Hamiltonian of the target ligand, H_T, using a chain of replicas with intermediate hybrid Hamiltonian states, governed by a scalar parameter λ changing from 0 to 1, where 0 corresponds to H_R and 1 to H_T. The exact coupling relation is described by eq S1 and S2.

The transformations were performed in the protein and solvent to determine $\Delta G_{R\to T}^{protein}$ and $\Delta G_{R\to T}^{solvent}$, respectively. Free-energy differences, ΔG, associated with the alchemical transformations were computed using the Bennett Acceptance Ratio (BAR)[42,54] and Thermodynamic Integration (TI) methods.[54,55] Finally, the relative binding free energy, $\Delta\Delta G_{R\to T}$, was determined as a difference between $\Delta G_{R\to T}^{protein}$ and $\Delta G_{R\to T}^{solvent}$.

For asymmetrical ligands for which two values were determined, $\Delta\Delta G_A$ and $\Delta\Delta G_B$, depending on which site, A or B, the target ligand was modeled into, the following formulas to calculate the combined $\Delta\Delta G_{A/B}$ were used

$$\exp\left(\frac{-\Delta\Delta G_{A/B}}{kT}\right) = \frac{\exp\left(\frac{-\Delta\Delta G_A}{kT}\right) + \exp\left(\frac{-\Delta\Delta G_B}{kT}\right)}{2} \quad (7)$$

$$\Delta\Delta G_{A/B} = -kT \ln \frac{\exp\left(\frac{-\Delta\Delta G_A}{kT}\right) + \exp\left(\frac{-\Delta\Delta G_B}{kT}\right)}{2} \quad (8)$$

Enhanced Sampling. *HREX with Potential Softening and Conformation Reservoir.* Sampling of the conformational space of replicas used for alchemical transformation was enhanced by the Hamiltonian replica exchange (HREX)[56] method. In this method, conformations of neighboring replicas periodically exchange, if they fulfill certain energy conditions, increasing the overall sampling. In our simulations, the exchanges are attempted every 120 s. Using real time for the wall time, instead of simulation time, allowed us to efficiently use a cluster of diverse GPUs. A typical alchemical transformation in proteins went through 800 exchange cycles, and 400 exchange cycles in water, roughly corresponding to 2−4 ns.

To efficiently sample conformations of ligands in the protein binding pockets, we reduced the energetic barriers between the potential minima by "softening" selected protein−ligand and ligand−ligand interactions. The softening was introduced into the HREX chain either directly or indirectly via a pre-prepared conformation reservoir. In the direct approach, the softening was gradually turned on from the terminal replicas (λ = 0.0 and λ = 1.0) toward the middle replica (λ = 0.5). As a result, ligands close to the middle were able to sample conformations more efficiently and propagate them toward the terminal nonsoftened replicas through HREX. In the indirect approach, the softening was used to generate a reservoir of enhanced conformations for the reference ligand. Then, the reservoir was coupled to a corresponding replica (λ = 0.0) from which the conformations propagated toward the target ligands (λ = 1.0) through HREX.

Generation of the conformation reservoir consists of two steps. In the first step, a long simulation with softened interactions was performed. In the second step, the softened ensemble was converted to a nonsoftened ensemble, i.e., desired reservoir. We explored two methods for the second step——HREX and NEQ MD. In the HREX approach, the softened ensemble was alchemically transformed to the nonsoftened ensemble by a set of intermediate λ-states (similarly as for the ligand mutation). In the NEQ approach, the Hamiltonian of the softened ensemble quickly changes from the softened to the nonsoftened state. Work and ΔG calculated during this process serve to filter a generated nonsoftened ensemble (see the following paragraph for details). Both approaches allow generation of a Boltzmann-distributed ensemble that can periodically insert random conformations to the corresponding replica (here, the reference replica, λ = 0) of mutation HREX.

Conformation Reservoir Generation Using Nonequilibrium MD. In our NEQ MD runs, the Hamiltonian of the molecular system linearly changes from the softened to the nonsoftened state during the time interval T, governed by a coupling parameter λ. For the systems studied, it was found that T = 10 ps provides a good tradeoff between the computational costs of the simulations and the accuracy of the results. For each of the NEQ MD runs, work values were computed as explained in the Theory section. The starting conformations of NEQ MD runs were drawn from the

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

trajectory generated in MD with a softened Hamiltonian as described above. The following workflow (see Figure 1) was used to generate a conformation reservoir:

(0) Generate the softened trajectory (MD trajectory of the system with "softened" Hamiltonian, having reduced potential barriers between relevant local potential minima of the ligand−protein complex).
(1) Take equally spaced conformations from the softened MD trajectory.
(2) Run NEQ MD starting from the chosen conformations changing the Hamiltonian from the softened to the nonsoftened state (forward NEQ MD runs).
(3) Compute work values for forward NEQ MD simulations, and compute ΔG between softened and nonsoftened Hamiltonian states using Jarzynski equality.
(4) Filter the NEQ simulations based on computed work and ΔG values using a Metropolis algorithm as described in the Theory section.
(5) Run NEQ MD from the nonsoftened to the softened Hamiltonian state (reverse NEQ MD) starting from the end conformations of filtered forward NEQ MD runs.
(6) Use forward and reverse NEQ MD results to compute ΔG with the bidirectional method.
(7) Go to 4.
(8) Repeat steps 4−7 500 times. We need to run reverse NEQ MD only for those configurations that were not filtered in the previous cycles.
(9) Average ΔG values obtained in cycles 4−7.
(10) Obtain a final set of filtered MD conformations from forward NEQ runs using the averaged value of ΔG between softened and nonsoftened Hamiltonian states (Figure 1).

The methods are described in more detail in the Supporting Information.

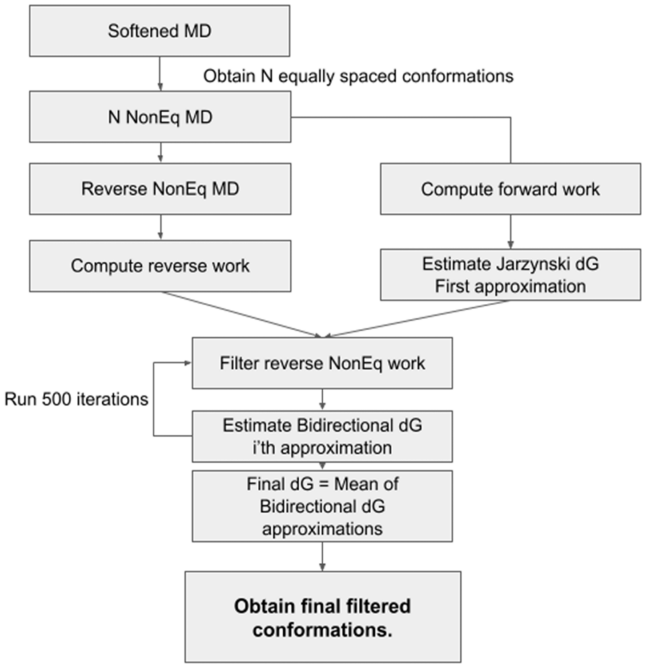


Figure 1. Generation of conformation reservoir using nonequilibrium MD.

RESULTS AND DISCUSSION

Relative Binding Free-Energy Predictions. ARROW FF has shown its ability to predict solvation free energy of arbitrary small neutral molecules with unprecedented accuracy (MAE: 0.2−0.3 kcal/mol).[21] Here, we probe its ability to predict protein−ligand relative binding free energy. Our test set consists of three proteins——MCL1, Thrombin, and CDK2 (Figure S1)——each with a series of binding ligands (Figures S2−S4). The complexes proved to be stable during 10 ns long MD simulations with an average RMSD of Cα atoms from the X-ray structures of 1.3, 1.2, and 1.9 Å for MCL1, Thrombin, and CDK2, respectively (Figure S5). Such a deviation is on a similar level to that of other force fields.[19] Nonetheless, the following ligand-binding simulations were performed with peripheral Cα atoms positionally restrained to avoid the effects of potential slow conformational changes. The $\Delta\Delta G$ predictions obtained with ARROW FF are shown against experimental values in Figure 2. The exact values, correlation coefficients, slopes, and errors can also be found in Tables S1−S3.

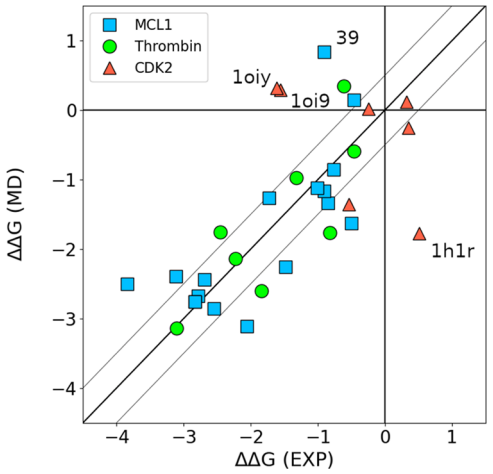


Figure 2. Parity plot comparing the relative binding free energies $\Delta\Delta G$ for ligand mutations in MCL1, Thrombin, and CDK2 as predicted by ARROW FF and the experiment. Results with ARROW FF were calculated with HREX and conformation reservoir generated via potential softening and NEQ MD. Selected ARROW $\Delta\Delta G$ values with the largest deviation from the experiment (MAE > 1.5 kcal/mol) are marked with labels - ligands 1h1r, 1oi9, and 1oiy bound to CDK2, and ligand 39 bound to MCL1. The thin gray lines are +/− 0.5 kcal/mol from the diagonal.

Overall, ARROW FF predicts $\Delta\Delta G$'s well (MAE: 0.7) and at a similar accuracy level as leading nonpolarizable force fields OPLS[10] (MAE: 0.6), GAFF[9,11] (MAE: 0.8), and CGenFF (MAE: 0.8)[12] (see Figure S6 for comparison). Notably, three $\Delta\Delta G$'s with the largest deviation from the experiment (~2 kcal/mol) come from mutations in the same protein—CDK2, i.e., mutations of 1h1r, 1oi9, and 1oiy ligands. We analyzed these simulations in detail, looking for putatively incorrectly described protein−ligand interactions and found 1oi9 being the most evident case. Namely, the X-ray structure of 1oi9 in CDK2 (PDB: 1oi9) indicates that a hydroxyl group of the ligand forms a hydrogen bond with the carboxylate group of ASP87. However, such a hydrogen bonding interaction is not observed during the simulations. Lack of a strong O−H···O hydrogen bond well explains the binding being underestimated

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

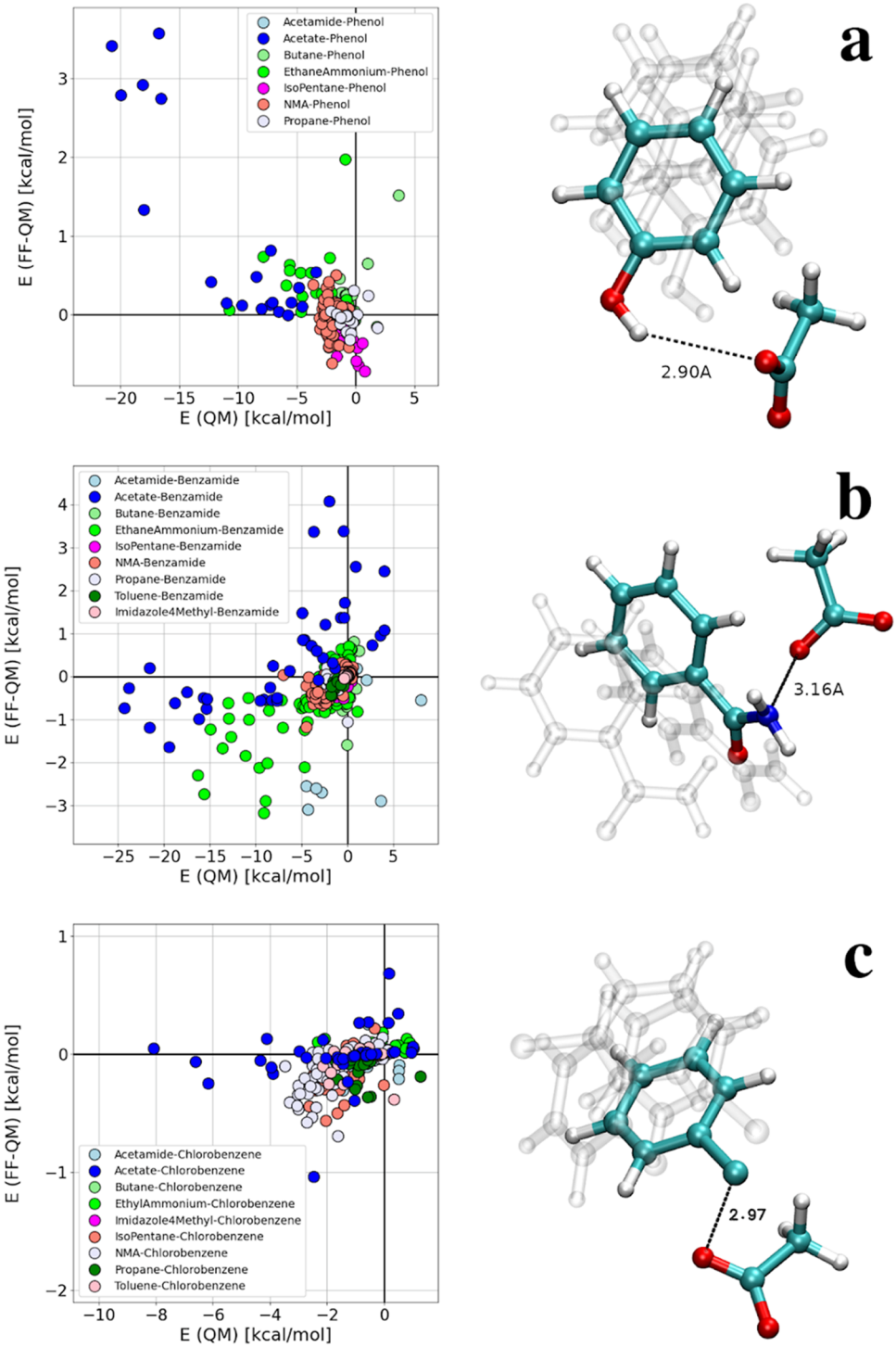


Figure 3. Difference in the interaction energy determined with ARROW FF and QM for amino acid fragments of CDK2 and ligands: (a) 1oi9 (phenol), (b) 1oiy (benzamide), and (c) 1h1r (chlorobenzene). Configurations with the largest discrepancy are shown on the right (opaque) along with a few others (transparent).

by ∼2 kcal/mol. Additionally, we found that in simulations with the GAFF force field, this hydrogen bond is present and persists over the entire simulation, ultimately producing $\Delta\Delta G$ with a much smaller deviation from the experiment. To check if the ARROW FF misrepresents this or any other interaction with the 1oi9 ligand, we extracted dimers of protein−ligand fragments from the GAFF simulation and calculated their interaction energy using ARROW FF and QM ("silver standard"). The difference between the two energies, i.e., FF-QM, can be seen in Figure 3a. Indeed, the largest inconsistency is found for a pair of phenol (fragment of 1oi9) and acetate (fragment of ASP87).

Similar FF-QM calculations were performed for ligand 1oiy. Although a hydrogen bond between an amide group of the ligand and a carboxylate group of ASP87 periodically forms during the simulation with ARROW FF, that is consistent with the X-ray structure (PDB: 1oiy), the FF-QM indicates significant discrepancy (Figure 3b). $\Delta\Delta G$ of the third

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

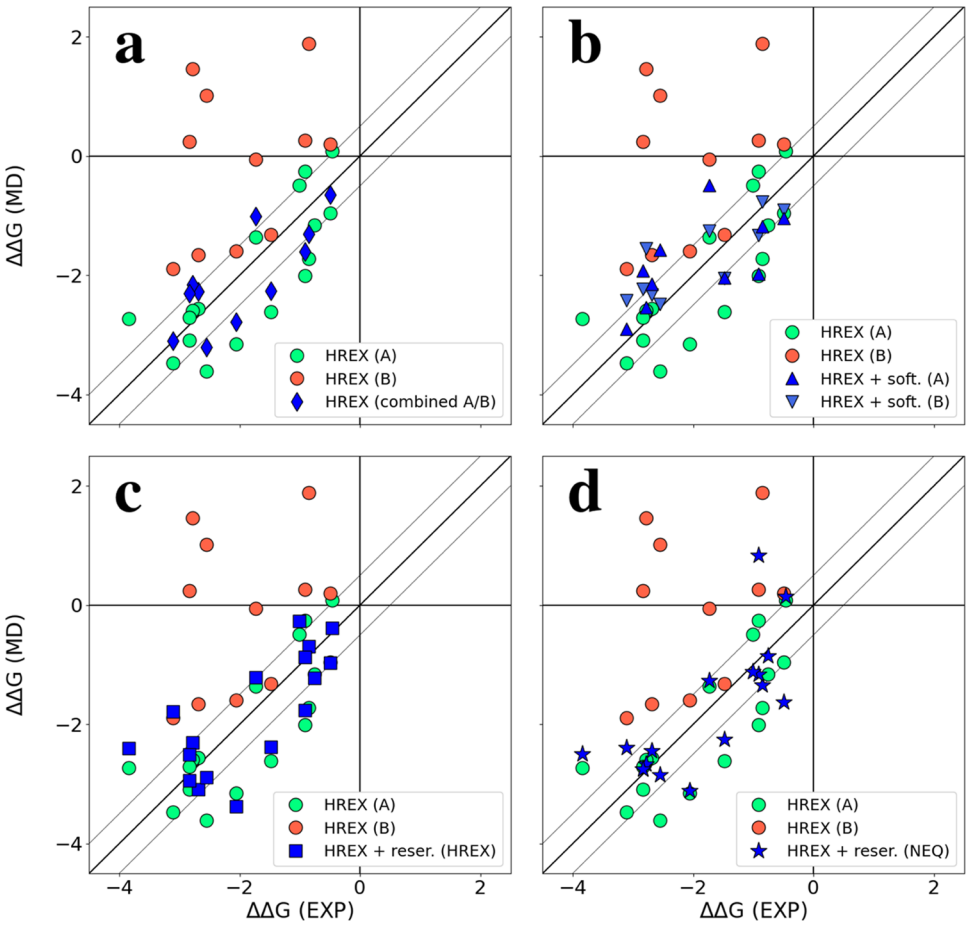


Figure 4. Comparison of $\Delta\Delta G$ as determined with ARROW force field and the experiment for MCL1. Green and red markers correspond to values obtained with HREX and a target ligand starting at A and B sides, respectively. Blue markers correspond to (a) combined A/B sides, (b) HREX with softening from A and B sides, (c) HREX with reservoir (from HREX), and (d) HREX with reservoir (from NEQ). Thin gray lines are +/− 0.5 kcal/mol from the diagonal.

questionable ligand, 1h1r, as opposed to 1oi9 and 1oiy, is overestimated with respect to the experiment. Nevertheless, interactions with ASP87 are likely to be the key in this case too, since the X-ray (PDB: 1h1r) indicates that the chlorine of the ligand and the acetate group of ASP87 are in close proximity (~3 Å). FF-QM calculations confirm this: finding acetate−chlorobenzene dimers showing the largest discrepancy (Figure 3c). These observations suggest that ARROW FF might not reproduce the interactions that involve charged groups sufficiently well.

It has been shown that for accurate free-energy (ΔG) calculations, nuclear quantum effect (NQE) should be taken into account.[43] However, we found that this effect mostly cancels out in our relative free-energy ($\Delta\Delta G$) calculations, where ΔG determined in water is subtracted from ΔG determined in a protein. We repeated two of our calculations of ligands binding to the CDK2 protein with PIMD = 4 (Path Integral Molecular Dynamics), which models NQE. As can be seen in Table S4, although corresponding ΔG values are reduced due to NQE, the final $\Delta\Delta G$ values are not very different. For this reason, as well as the high computational cost of using PIMD, we neglected NQE in our present calculations.

Conformational Sampling. In addition to force field accuracy, adequate conformational sampling is another factor that determines the validity of binding free-energy calculations. Missing or inadequate sampling of strong or weak binding states can result in underestimation or overestimation of the binding free energy, respectively. This question is even more compelling in the case of polarizable models that can significantly raise the bar for the currently available enhanced sampling techniques. Having this in mind, we paid particular attention to extensively sample the conformations in our protein−ligand systems.

For all of the three protein systems studied here, we choose ligands with the simplest benzyl group as a reference. Thus, mutations to the target ligands mostly relied on growing an additional group from the benzyl site. If the group grows asymmetrically, i.e., in *ortho* or *meta* position, then, there are two alternative sites for it (see Figure S7). We call them sites A and B. It is important to mention that in our simulations, any ligand, even the reference, does not flip from site A to B, or vice versa, when it is in the binding pocket. It is unclear if such a flip is possible in reality, or a ligand needs to leave and reenter the packet to do so. In either case, it is unknown apriori how these two sites contribute to the experimental $\Delta\Delta G$. In the case of 1h1r ligand in CDK2, both the orientations of the chlorobenzene group in *meta* position were resolved from the X-ray experiment (Figure S9). This suggests that each of them should be taken into account in the binding free-energy calculations.

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

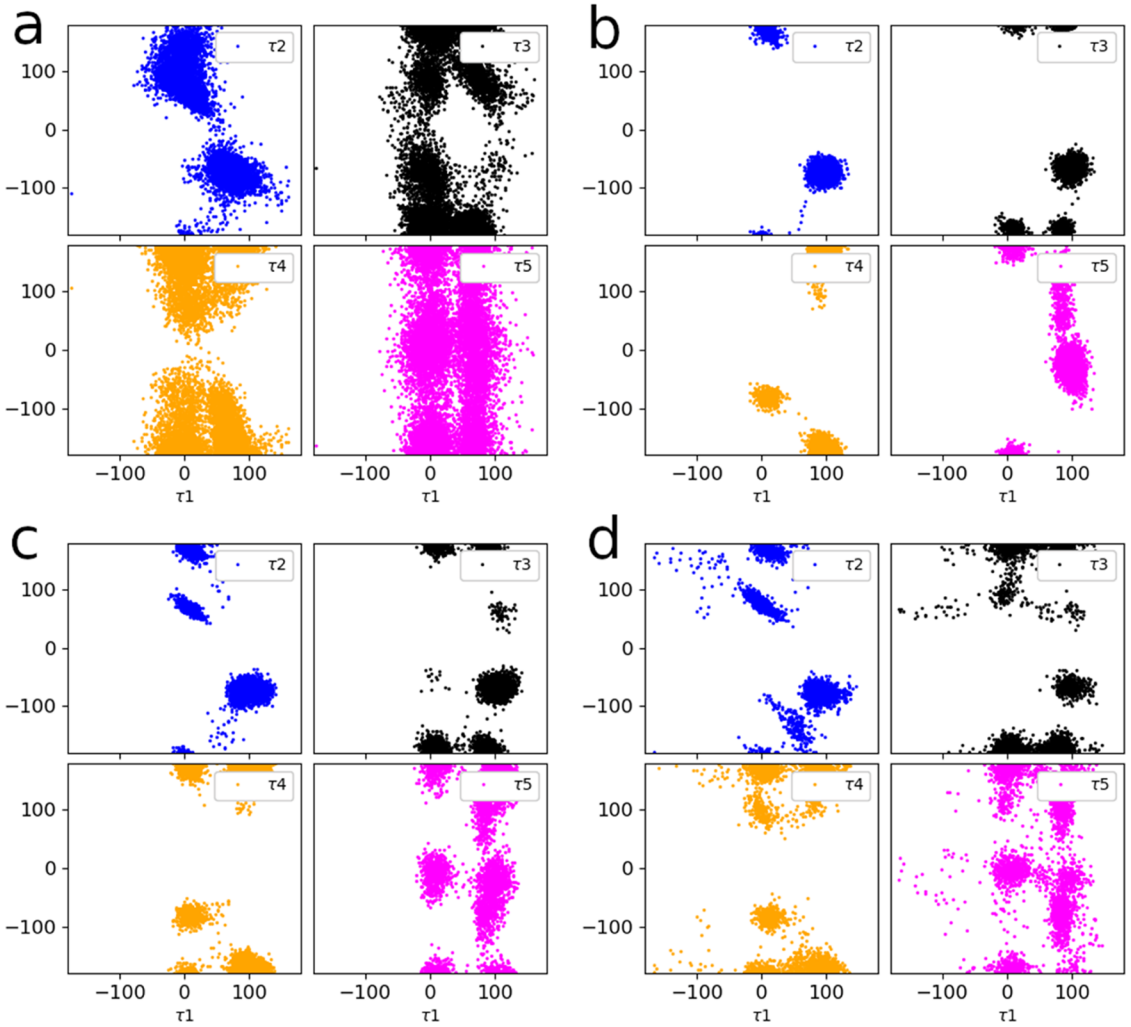


Figure 5. Pairwise distribution of rotatable torsions of ligand 27 in MCL1 from (a) softened potential MD, (b) nonsoftened potential MD, (c) nonsoftened potential MD with a reservoir (generated with HREX), and (d) nonsoftened potential MD with a reservoir (generated with nonequilibrium protocol).

Here, we discuss this sampling problem, and how we deal with it, in detail for the case of MCL1. Indeed, for asymmetrical ligands, we obtained two different sets of $\Delta\Delta G$ values (see Figure 4a) depending on which site, A or B, the ligands started the simulations from (MAE: 0.61/1.38 kcal/mol). Although the obtained ΔG's can be combined according to eq 8, it is in question if there are any more missing states. Furthermore, analyzing the simulations, we noticed that certain torsional angles along the ligand backbone undergo rare transitions. We found that the X-ray structures (PDB: 4hw2, 4hw3) also contain their alternative states (Figure S10). Thus, to sample all of the possible states extensively, including flipping of the benzyl ring, we softened all of the backbone torsions of the ligand (Figure S8a) and also the protein−benzyl interactions. In our first approach, we applied the softening directly to the mutation HREX chain (maximum for the middle replica, $\lambda = 0.5$). Although convenient, to keep a sufficient exchange rate between the replicas, we had to increase the number of λ-states from 11 to 21. As expected, the obtained $\Delta\Delta G$ values with this method were mostly found between the A and B states determined with the regular HREX (Figure 4b) (MAE: 0.66/0.49). Nevertheless, they still depend on which site the simulation started from, which is an indication of the convergence issue. Moreover, as the softening applies to the hybrid ligand ($\lambda = 0.5$), it might be suspected to not perform as effectively for different mutations.

Utilizing our second approach, we deconvolute the softening from the mutation. We run a single long simulation with potential softening of the reference ligand 27 in MCL1. We made sure that all of the torsions and the benzyl ring flipped between the different states multiple times (Figures 5a, S14, and S15). To convert the softened ensemble to the nonsoftened ensemble, i.e., reservoir, we used two different methods—HREX and NEQ. We found that both methods produced similar conformational ensembles (compare Figure 5c,d). What is more, we found that the reservoirs contain conformations otherwise not sampled (compare, e.g., Figure 5d,b). When the reservoirs were attached to the replica of reference ligand 27 ($\lambda = 0.0$), the conformations efficiently propagated along the mutation HREX chain enhancing sampling also of the target ligands (see Figure S13 with an example of replica exchanges). $\Delta\Delta G$'s obtained with both the reservoirs were found consistent with each other as well (compare Figure 4c,d) and either reduced the discrepancy with the experiment (MAE: 0.56/0.59). The only significant difference was noticed for ligand 39 (see Figure 1) whose

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

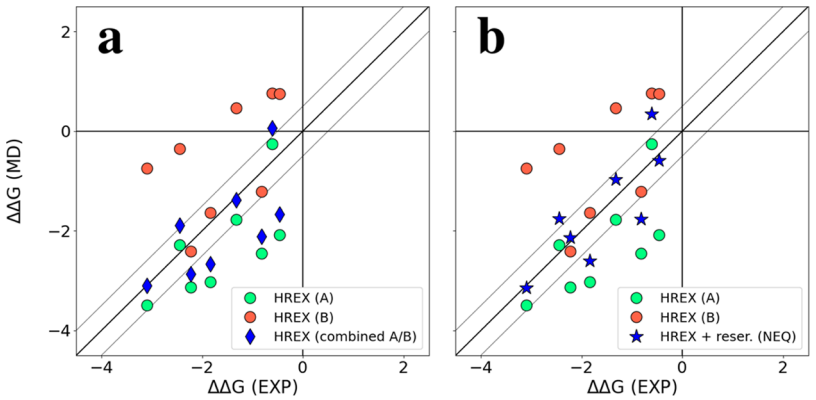


Figure 6. Comparison of $\Delta\Delta G$ as determined with the ARROW force field and experiment (Ki)[57] for Thrombin. Green and red markers correspond to values obtained with HREX and a target ligand starting at A and B sides. Blue markers correspond to (a) combined A/B sides and (b) HREX with reservoir (from NEQ). The thin gray lines are +/− 0.5 kcal/mol from the diagonal.

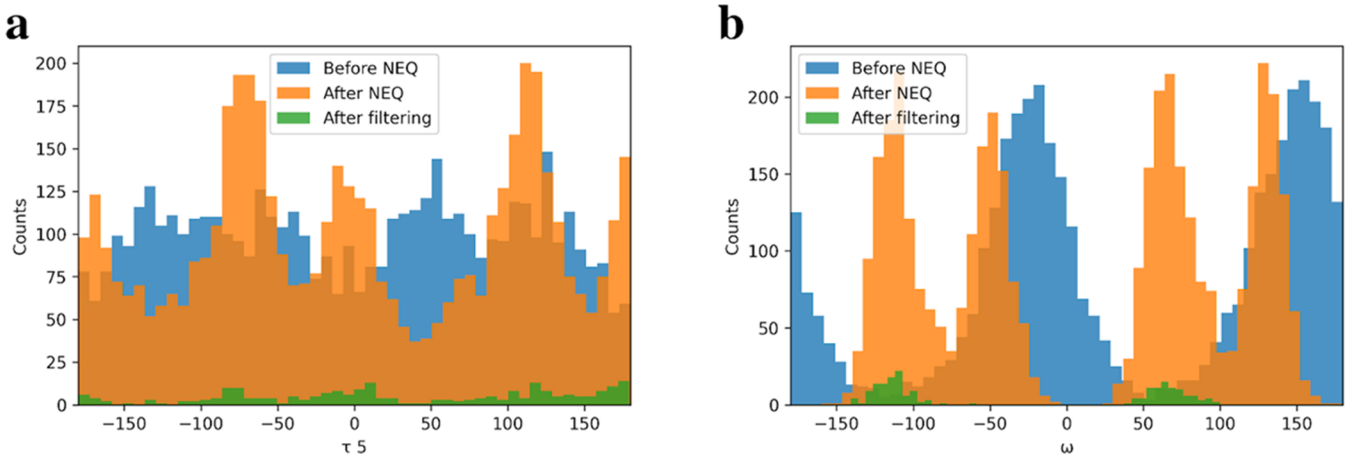


Figure 7. Distributions of benzyl ring (a) torsion $\tau 5$ for ligand 27 bound to MCL1 and (b) torsion ω of ligand 5 bound to Thrombin before and after NEQ MD runs for conformer reservoir generation.

sampling is particularly challenging because of a large phenyl group. With the NEQ-generated reservoir, ligand 39 was found to partially leave the binding pocket. Nonetheless, because of the high computational parallelizability, the NEQ method was chosen to generate conformational reservoirs for the other systems studied here.

Figure 6a shows $\Delta\Delta G$ values determined for a series of ligands binding Thrombin in two alternative orientations, A and B. A-orientation is clearly more preferable than B-orientation and "combined" $\Delta\Delta G$ (A/B) values computed with eq 8 values are very close for A-side $\Delta\Delta G$'s and agree well with the Ki experiment (fluorescence labeling) (MAE: 0.66). Nevertheless, coupling of HREX with a conformation reservoir generated using nonequilibrium MD makes the predictions even more accurate (Figure 6b, MAE: 0.50).

Nonequilibrium MD. To generate a Boltzmann-distributed reservoir of conformation with regular, i.e., nonsoftened, Hamiltonian, conformations from the trajectory with "softened" protein−ligand interactions were used as starting points for nonequilibrium MD runs. Figure 7a shows the distribution of the torsion angle $\tau 5$ (see Figure S8a) that describes the orientation of the benzyl ring of ligand 27 in MCL1—before NEQ MD runs (the softened MD trajectory), at the end of 10 ps NEQ MD runs, and after filtering based on the computed work as described in the Theory section. Figure 7b shows the same distributions for the torsion angle ω (see Figure S8b) describing the orientation of the benzyl ring of ligand 5 in Thrombin. One can see how NEQ MD runs and filtering change the observed distributions of benzyl torsions. Especially, it can be clearly seen for ligand 5 of Thrombin (Figure 7b). For the regular nonsoftened Hamiltonian, the distribution of the benzyl torsion angle ω has narrow peaks around −120 and 60°. Distribution of ω angle for softened Hamiltonian MD before NEQ MD runs is wide with probability maxima around −20 and 160°. After NEQ MD runs, the distribution of ω angle shows four narrow peaks at −120, −60, 60, and 130°. NEQ work-based filtering removes configurations at −60 and 130° so filtered configurations have the correct Boltzmann distribution corresponding to the nonsoftened Hamiltonian of the system.

As outlined in the Methods section, bidirectional ΔG was computed via the iterative procedure. The value ΔG convergence is defined here as ΔG (bidirectional). In two systems studied, free energies computed via the bidirectional approach and Jarzynski equality differ. For MCL1 ligand 27: ΔG (Jarzynski) = −34.20 kcal/mol, ΔG (bidirectional) = −32.26 kcal/mol. For Thrombin ligand 5: ΔG (Jarzynski) = −2.20 kcal/mol, ΔG (bidirectional) = −1.83 kcal/mol. It is known that Jarzynski equality expression is strongly affected by tails of the work distributions often resulting in too negative computed ΔG values. The bidirectional approach is not prone

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

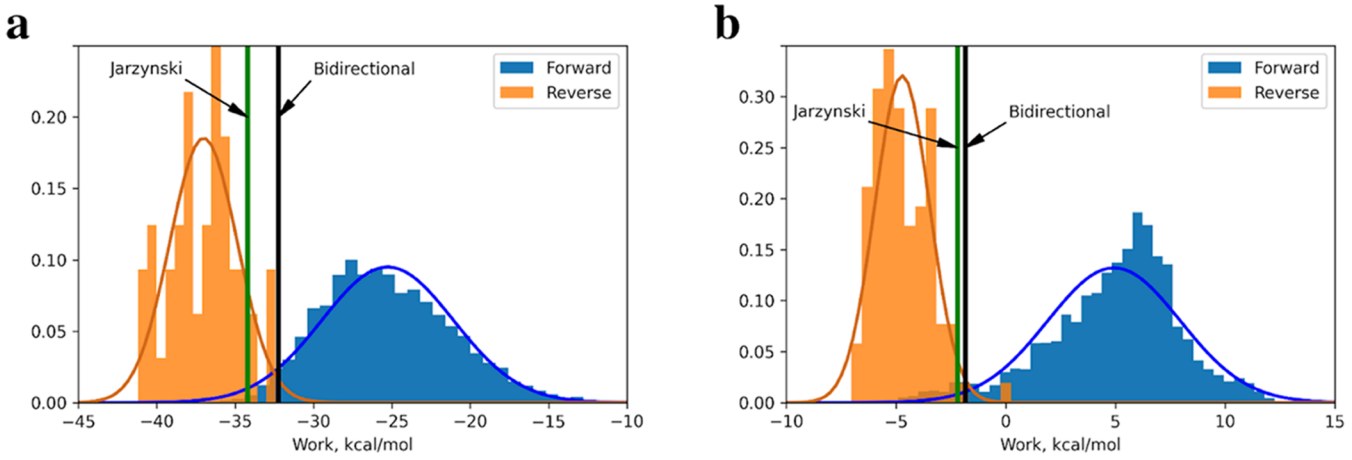


Figure 8. Forward and reverse work distribution for (a) MCL1 ligand 27 and (b) Thrombin ligand 5. ΔG values between softened and nonsoftened Hamiltonian states of the system computed with one-directional (Jarzynski) and bidirectional (Crooks theorem-based) approaches are shown with green and black vertical lines.

to these problems and provides more robust estimates of ΔG between the softened and nonsoftened states.

Figure 8 shows distributions of work values computed for forward and reverse NEQ MD runs. The distribution of reverse work is sparse since the reverse NEQ MD starts from end conformations of forward NEQ runs that passed the Metropolis criteria (see Theory Section). The acceptance ratio for the MCL1 ligand 27 system was 4.8% and that for Thrombin ligand 5 was 4.7%. Blue and orange curves represent Gaussian curves that fit distributions of forward and reverse work values, respectively. The crossing point of these curves may be used as another estimate of the free-energy difference between softened and nonsoftened states of the molecular system and is close to the estimate obtained by the bidirectional approach. Green and black vertical lines correspond to the ΔG values computed via Jarzynski and bidirectional approach.

CONCLUSIONS

Based on the three protein−ligand systems, we have shown that ARROW FF is able to predict relative binding free energy with almost chemical accuracy and is currently on par with leading all-atom fixed-charge force fields. We identified that the largest discrepancy with the experimental results is associated with binding interactions that involve charged groups. Analogous QM simulations demonstrated that these kinds of interactions are not well represented by the model and will be addressed in a later publication.

Despite current limitations, ARROW proves its potential in the field of drug design. As our model is physics-based and relies purely on QM calculations, the sources of error can be narrowed down to particular energy terms and refined separately. Additionally, because our technology does not require any experimental data, the force field refinement can be performed in a systematic manner. Although, in general, this is an endless process, we believe that there is a particular complexity and accuracy of the model that needs to be reached for successful drug design. We think that making a force field faithful to high-level ab initio calculations is a step in that direction.

Since effective sampling is a common challenge and usually cannot be completely deconvoluted from the force field accuracy, here, we paid particular attention to sufficiently sample protein−ligand conformations. We have shown that a conformation reservoir generated through potential softening in a nonequilibrium process is an efficient way to extend conformational space of a ligand in the protein binding pocket.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00930.

Details of the methods, comparison of the enhanced sampling techniques, comment on alternative experimental results for Thrombin, structures of proteins and ligands, tables with values of the binding free energy, comparison with other force fields, and RMSD of the proteins (PDF)

AUTHOR INFORMATION

Corresponding Authors

Grzegorz Nawrocki − *InterX Inc., Berkeley, California 94710, United States*; orcid.org/0000-0002-7115-5857; Email: grzegorz.nawrocki@interxinc.com

Boris Fain − *InterX Inc., Berkeley, California 94710, United States*; Email: boris.fain@interxinc.com

Authors

Igor Leontyev − *InterX Inc., Berkeley, California 94710, United States*

Serzhan Sakipov − *InterX Inc., Berkeley, California 94710, United States*

Mikhail Darkhovskiy − *InterX Inc., Berkeley, California 94710, United States*

Igor Kurnikov − *InterX Inc., Berkeley, California 94710, United States*

Leonid Pereyaslavets − *InterX Inc., Berkeley, California 94710, United States*

Ganesh Kamath − *InterX Inc., Berkeley, California 94710, United States*

Ekaterina Voronina − *InterX Inc., Berkeley, California 94710, United States; Faculty of Physics, Lomonosov Moscow State University, Moscow 119991, Russia*

Oleg Butin − *InterX Inc., Berkeley, California 94710, United States*

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Alexey Illarionov − *InterX Inc., Berkeley, California 94710, United States*
Michael Olevanov − *Faculty of Physics, Lomonosov Moscow State University, Moscow 119991, Russia*
Alexander Kostikov − *InterX Inc., Berkeley, California 94710, United States*
Ilya Ivahnenko − *InterX Inc., Berkeley, California 94710, United States*
Dhilon S. Patel − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*
Subramanian K. R. S. Sankaranarayanan − *Center for Nanoscale Materials, Argonne National Lab, Lemont, Illinois 60439, United States; Department of Mechanical and Industrial Engineering, University of Illinois, Chicago, Illinois 60607, United States*
Maria G. Kurnikova − *Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0002-8010-8374
Christopher Lock − *InterX Inc., Berkeley, California 94710, United States; Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, California 94304, United States;* orcid.org/0000-0002-4087-5894
Gavin E. Crooks − *InterX Inc., Berkeley, California 94710, United States*
Michael Levitt − *Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, United States*
Roger D. Kornberg − *Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00930

Notes

The authors declare no competing financial interest.
InterX Inc. is a subsidiary of NeoTX Holdings Ltd., Rehovot, Israel 7670202.

■ ACKNOWLEDGMENTS

The authors thank InterX Inc. for their generous support. The authors also thank Sean Greenslade for his help with cluster maintenance and Hulda Chen for useful comments. Work performed at the Center for Nanoscale Materials, a U.S. Department of Energy Office of Science User Facility, was supported by the U.S. DOE, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

■ REFERENCES

(1) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911−2937.
(2) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex. *Science* **1987**, *235*, 574−576.
(3) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(4) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(5) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(6) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(7) MacKerell, A. D.; Jr; Banavali, N.; Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2000**, *56*, 257−265.
(8) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.
(9) Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M., Jr. Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, *59*, 3128−3135.
(10) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291−4300.
(11) He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein−Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5*, 4611−4619.
(12) Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L. Automated, Accurate, and Scalable Relative Protein−Ligand Binding Free-Energy Calculations Using Lambda Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 7895−7914.
(13) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31*, 671−690.
(14) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.
(15) 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. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.
(16) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(17) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(18) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(19) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D., Jr. Force Field for Peptides and Proteins Based on

the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9*, 5430−5449.
(20) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of Protein−ligand Binding Free Energy by Using a Polarizable Potential. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290−6295.
(21) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential. *J. Comput. Chem.* **2009**, *30*, 1701−1711.
(22) Harger, M.; Lee, J.-H.; Walker, B.; Taliaferro, J. M.; Edupuganti, R.; Dalby, K. N.; Ren, P. Computational Insights into the Binding of IN17 Inhibitors to MELK. *J. Mol. Model.* **2019**, *25*, No. 151.
(23) Khoruzhii, O.; Donchev, A. G.; Galkin, N.; Illarionov, A.; Olevanov, M.; Ozrin, V.; Queen, C.; Tarasov, V. Application of a Polarizable Force Field to Calculations of Relative Protein−ligand Binding Affinities. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10378−10383.
(24) Medders, G. R.; Babin, V.; Paesani, F. Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties. *J. Chem. Theory Comput.* **2014**, *10*, 2906−2910.
(25) Donchev, A. G.; Galkin, N. G.; Pereyaslavets, L. B.; Tarasov, V. I. Quantum Mechanical Polarizable Force Field (QMPFF3): Refinement and Validation of the Dispersion Interaction for Aromatic Carbon. *J. Chem. Phys.* **2006**, *125*, No. 244107.
(26) Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.; Pereyaslavets, L. B.; Tarasov, V. I. Assessment of Performance of the General Purpose Polarizable Force Field QMPFF3 in Condensed Phase. *J. Comput. Chem.* **2008**, *29*, 1242−1249.
(27) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. Derivation of Class II Force Fields. VIII. Derivation of a General Quantum Mechanical Force Field for Organic Compounds. *J. Comput. Chem.* **2001**, *22*, 1782−1800.
(28) Hagler, A. T. Quantum Derivative Fitting and Biomolecular Force Fields: Functional Form, Coupling Terms, Charge Flux, Nonbond Anharmonicity, and Individual Dihedral Potentials. *J. Chem. Theory Comput.* **2015**, *11*, 5555−5572.
(29) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.
(30) Naseem-Khan, S.; Gresh, N.; Misquitta, A. J.; Piquemal, J.-P. Assessment of SAPT and Supermolecular EDA Approaches for the Development of Separable and Polarizable Force Fields. *J. Chem. Theory Comput.* **2021**, *17*, 2759−2774.
(31) Pereyaslavets, L.; Kamath, G.; Butin, O.; Illarionov, A.; Olevanov, M.; Kurnikov, I.; Sakipov, S.; Leontyev, I.; Voronina, E.; Gannon, T.; Nawrocki, G.; Darkhovskiy, M.; Ivahnenko, I.; Kostikov, A.; Scaranto, J.; Kurnikova, M. G.; Banik, S.; Chan, H.; Sternberg, M. G.; Sankaranarayanan, S. K. R. S.; Crawford, B.; Potoff, J.; Levitt, M.; Kornberg, R. D.; Fain, B. Accurate Determination of Solvation Free Energies of Neutral Organic Compounds from First Principles. *Nat. Commun.* **2022**, *13*, No. 414.
(32) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(33) Sugita, Y.; Okamoto, Y. Replica-Exchange Multicanonical Algorithm and Multicanonical Replica-Exchange Method for Simulating Systems with Rough Energy Landscape. *Chem. Phys. Lett.* **2000**, *329*, 261−270.
(34) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749−13754.
(35) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(36) InterX Inc., Berkeley, CA, USA. Arbalest - MD Simulation Program. 2022.
(37) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L. Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. *Chem. Sci.* **2020**, *11*, 1140−1152.
(38) Gill, S. C.; Lim, N. M.; Grinaway, P. B.; Rustenburg, A. S.; Fass, J.; Ross, G. A.; Chodera, J. D.; Mobley, D. L. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. *J. Phys. Chem. B* **2018**, *122*, 5579−5598.
(39) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(40) Crooks, G. E. Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Phys. Rev. E* **1999**, *60*, 2721−2726.
(41) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003**, *91*, No. 140601.
(42) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(43) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, R. D.; Fain, B. On the Importance of Accounting for Nuclear Quantum Effects in Ab Initio Calibrated Force Fields in Biological Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 8878−8882.
(44) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem. A* **1993**, *97*, 10269−10280.
(45) Vingelmann, P.; Fitzek, F. H.; Nvidia CUDA, Release: 10.2. 89 *NVIDIA* https://developer.nvidia.com/cuda-toolkit2020.
(46) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(47) Giese, T. J.; Panteva, M. T.; Chen, H.; York, D. M. Multipolar Ewald Methods, 1: Theory, Accuracy, and Performance. *J. Chem. Theory Comput.* **2015**, *11*, 436−450.
(48) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990−2001.
(49) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(51) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.
(52) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(53) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(54) Shirts, M. R.; Pande, V. S. Comparison of Efficiency and Bias of Free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. *J. Chem. Phys.* **2005**, *122*, No. 144107.
(55) Straatsma, T. P.; McCammon, J. A. Multiconfiguration Thermodynamic Integration. *J. Chem. Phys.* **1991**, *95*, 1175−1188.
(56) Woods, C. J.; Essex, J. W.; King, M. A. Enhanced Configurational Sampling in Binding Free-Energy Calculations. *J. Phys. Chem. B* **2003**, *107*, 13711−13718.

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(57) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. More than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the s1 Pocket of Thrombin. *J. Mol. Biol.* **2009**, *390*, 56−69.

https://doi.org/10.1021/acs.jctc.2c00930
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Supporting Information

Protein-ligand binding free energy calculations with ARROW - a purely first principle parameterized polarizable force field

Grzegorz Nawrocki[1*], Igor Leontyev[1], Serzhan Sakipov[1], Mikhail Darkhovskiy[1], Igor Kurnikov[1], Leonid Pereyaslavets[1], Ganesh Kamath[1], Ekaterina Voronina[1,2], Oleg Butin[1], Alexey Illarionov[1], Michael Olevanov[2], Alexander Kostikov[1], Ilya Ivahnenko[1], Yevhen K. Cherniavskyi[1], Dhilon S. Patel[3], Subramanian K. R. S. Sankaranarayanan[4,5], Maria G. Kurnikova[3], Christopher Lock[1,6], Gavin E. Crooks[1], Michael Levitt[7], Roger D. Kornberg[7], Boris Fain[1*]

[1] InterX Inc., Berkeley, California 94710, USA, a subsidiary of NeoTX Holdings Ltd., Rehovot, Israel 7670202
[2] Faculty of Physics, Lomonosov Moscow State University, Moscow 119991, Russia
[3] Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[4] Center for Nanoscale Materials, Argonne National Lab, Lemont, IL 60439, USA
[5] Department of Mechanical and Industrial Engineering, University of Illinois, Chicago, IL 60607, USA
[6] Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, CA 94304, USA
[7] Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
* Corresponding authors: boris.fain@interxinc.com, grzegorz.nawrocki@interxinc.com

METHODS

Quantum Mechanical details

In our work we use a variety of quantum mechanical data. For monomer fragments used for intermolecular parameterization we use the following: an electrostatic potential map with a large number of points on many vdw-distances from the molecule (thousands per fragment), dipoles, quadrupoles, and a polarization tensor as well as energies of interaction (hundreds per heavy atom in the fragment) with a small probe charge (typically 0.1 proton charge). All of these calculations are performed at the MP2/aug-cc-pVQZ level. Such monomer fragments cover the chemical space of proteins and simulated ligands. For merging several fragments we calculate electrostatic potential on a less expensive level of MP2/aug-cc-pVTZ(d/p).

For dimer energy we use high quality QM data because small inaccuracies from these can translate into large errors when typical fragments of interactions are repeated, e.g. N-methyl acetamide with water. We use total energies calculated with the silver standard *i.e.* MP2/CBS, calculated with Helgaker cubic extrapolation from aug-cc-pVTZ->aug-cc-pVQZ as well as

post-MP2 correction (i.e. plus CCSD(T)/aug-cc-pVDZ - MP2/aug-cc-pVDZ). To improve transferability and facilitate optimization we use DFT-SAPT decomposition with dHF correction at aug-cc-pVTZ level with PBE0 functional. We employ four parts of DFT-SAPT decomposition which have corresponding manifestation in our FF : ES (electrostatic E1pol), EX (exchange-repulsion E1exch), IND (induction, E2ind + E2ind-exch + δHF), DS (dispersion, E2disp+E2disp-exch + Esilver-standard - DFT-SAPTtotal_energy). The dispersion term accumulates all disagreements between total energy described by “silver standard” and DFT-SAPT+δHF energies. Further details on QM calculations for the dimers can be found in the Supplementary methods in our previous publication [1]. These dimers include interaction between protein fragments, between protein fragments and ligand fragments, and between protein and ligand fragments and water.

The diversity of ligand chemical space requires special accuracy in bonded space. Every unique typified bond, angle, bond-angle and out-of-plane perturbation of small fragments of ligand were performed from the ground energy state of the ligand. Correct torsional perturbations play a key role in a configurational space of ligand molecules. We sample torsion energies with rigid rotational perturbations as well as “relaxed” perturbation where coordinates of all atoms are optimized except for ones in a particular fixed torsion of interest. In rare cases we have performed two dimensional rigid and relaxed potential energy scans. In cases where the ligand was larger than 18 heavy atoms, it was split into smaller pieces and hydrogens were added in the place of covalent bonds removed. All QM calculations for bonded perturbation were performed at MP2/aug-cc-pVTZ level.

Force field description and parameterization strategy

We are developing a force field that is based on the best available high accuracy QM data and a complex physics-based functional form. In our development we rely on previous discoveries in force fields. Our model - the ARROW Force Field or ARROW FF - has the following features: The non-bonded interactions are composed of electrostatic, exchange-repulsion and dispersion terms. The electrostatic and exchange-repulsion terms are multipolar with inclusion of charges, dipoles and quadrupoles, and their radial dependence is a Slater-like exponential so that these are better able to describe charge penetration. The dispersion component is conventionally represented by spherical terms (C6 and C8), and a Tang-Toennies-damped interaction. Many-body effects are modeled by anisotropic atomic induced dipoles interacting with the electrostatic and exchange-repulsion terms and with each other and iterated to self-consistent field (SCF) convergence on every non-bonded step. The intermolecular parameters of ARROW FF are determined by agreements with QM values of dimer and multimer energies, electrostatic potentials, multipole moments of monomers, polarization tensor and interaction of fragments with point charges. To aid transferability, we also attempt to match the individual FF energy components to their corresponding QM counterparts, in addition to reproducing the total energy. We use the functional form of the bonded interactions that is taken from MMFF94, with force constants and equilibrium values fitted to QM energies. The functional form details can be found in our previous works [1,2] which in turn are based on the QMPFF3 force field functional form [3,4].

We assign force field types based on our knowledge of different chemical functional groups. This is somewhat correlated with the MMFF94 approach. Every atom that is different under topology from another has its own type in every chemical functional group. Various aromatic or heterogeneous rings are considered as independent chemical functional groups. Sometimes force field types in different functional groups are merged by their chemical similarities, e.g. carbonyl carbon and oxygen in acetamide and N-methyl acetamide.

We split protein and ligands into chemical functional groups and parameterize them using monomer, dimer quantum mechanical data as described above. Typical size of the fragments is not bigger than 7 heavy atoms, e.g. phenol. Bigger molecules are built joining together smaller fragments and applying special rules that modify electrostatic parameters of "joint" atoms at the interface between fragments. The details of non-bonded parameterization have been described in supplementary information to our previous publication [1]. The interaction between neutral fragments typically has MAE on the order of ~0.3 kcal/mol (comparable to our previous work, where one of the two fragments was a water molecule), while the error is bigger (MAE about ~1 kcal/mol or worse) for charged-neutral fragments interactions. There are many factors that explain the origin of errors for charged groups.

First of all, the interaction energies of the charge with the neutral species (dimers) is large (e.g. about -20 kcal/mol), if not the largest for the entire set of interactions present in the protein-ligand complexes. These interaction energies are also large in comparison to a typical hydrogen bond (e.g. water-water) interaction which is about -5 kcal/mol. To further complicate the issue, the minimum and its corresponding potential energy surface for these charged-polar interactions is much narrower and harder to describe with the same accuracy as neutral compounds using the functional form. The second reason is that despite the functional form being sophisticated, it is unable to provide an equally accurate description of interactions for all the species so the scale of errors tends to be larger; the errors being roughly proportionally bigger. While our force field is able to predict neutral-neutral hydrogen bonded interactions with a MAE of about ~ 0.3 kcal/mol, our charge-neutral interactions tend to be proportionally worse and the MAE for such interactions is about ~1 kcal/mol or bigger, see e.g. **Figure 3**. Our functional form has physical terms that are proportional to the overlap integral *i.e.* exchange-repulsion and part of induction. It was shown that this also covers energy of charge transfer which is proportional to overlap integral. We believe that the complexity and magnitude of interactions as well as limitations of the functional form to the applied force field is the main reason for imperfect description of the potential energy surface, and not specific interactions such as "charge transfer". The errors are larger for every part of the interactions, e.g. electrostatics, penetration energy, exchange-repulsion, dispersion, and induction, as these occur at a bigger overlap due to stronger interactions.

Alchemical ligand transformations

The alchemical ligand transformation method was used for calculations of relative binding free energies (RBFE). In this method the Hamiltonian of the reference ligand ($\lambda=0$) is

incrementally transformed to the Hamiltonian of the target ligand using a chain of intermediate hybrid Hamiltonian states (λ-states), governed by a scalar parameter λ.

During the alchemical transformation, atoms of the hybrid ligand molecule with a common topology, referred to as the COMMON part, are mutated directly from one to another. Within the COMMON part all the interactions (bonded and non-bonded) are coupled between the A (λ=0) and B (λ=1) states linearly:

$$U_{\lambda} = (1 - \lambda)U_{A}^{Com}(r) + \lambda U_{B}^{Com}(r) \qquad \text{(Eq. S1)}$$

Topologically distinct groups of the hybrid molecule are transformed to corresponding DUMMY parts. Bonded interactions of the DUMMY atoms with the COMMON part atoms are coupled linearly based on **Eq. S1**. Non-bonded interactions of the DUMMY atoms with the COMMON part or with atoms of any other DUMMY part are switched between the "real" and "dummy" states with a nonlinear soft-cored function:

$$U_{\lambda} = (1 - \lambda)^{k}U_{A}^{Dum}(r + \lambda r_{sc}n) + \lambda^{k}U_{B}^{Dum}(r + (1 - \lambda)r_{sc}n) \qquad \text{(Eq. S2)}$$

where r_{sc} is the soft-coring radius parameter, k is the scaling factor power and n is the unit vector of the vector r. In the alchemical pathway, we decouple both the electrostatic and van-der-Waals interactions simultaneously. The soft-core **Eq. S2** helps to avoid singularities and instabilities in intermediate lambda states. The soft-coring radius r_{sc}=1.5 Å and the scaling factor power k=2 provided smooth shape of the dG/dλ profiles, lower statistical errors and no noticeable sampling-trapping artifacts for all studied compounds.

As λ changes from 0 to 1, some ligand-protein and ligand-ligand interactions are gradually turned on and others are turned off. The mixed single and dual topologies approach [5] was used to specify alchemical ligand Hamiltonian transformations. The COMMON part of the hybrid molecule for most of the compounds was defined as the maximum common substructure (MCS) between the reference and target compounds while the rest was defined as a DUMMY part. Typically we used 11 replicas spanning λ from 0.0 to 1.0 with an interval of 0.1. For some mutations it was beneficial to add two extra states: λ=0.05 and 0.95 totaling to 13 λ-states to provide a finer description of the alchemical transformation.

In the cases where a rotation around the bond next to the DUMMY-COMMON linker bond was important for conformation sampling, the COMMON part was reduced by one heavy atom in order to provide scaling down of the high barrier torsion potential in the course of the mutation, and achieve thermal accessibility of the important rotameric states in the HREX simulations.

Here, we give the details about how the bonded interactions involving the dummy atoms are treated in free energy perturbation (FEP) to avoid singularities and instabilities. It is known

that if the dummy atom does not have any bonded interactions with the rest of the molecule (COMMON part) and moves freely in the whole simulation volume it makes the sampling very difficult. Thus, in the molecule end state ARBALEST keeps the dummy atoms bonded with the rest of the molecule. However, if there is more than one bonded stretch or bonded angle or bonded dihedral angle interactions between a dummy atom and the rest of the molecule, the distributions sampled for the hybrid molecule (molecule including the dummy atoms) will be different from the molecule without the dummy atoms. So ARBALEST keeps only one bonded stretch, bonded angle, and bonded dihedral angle interactions involving a dummy atom whereas all the other bonded interactions are scaled to 0 at the end state. In this way, the contributions of the dummy atoms to the free energies in the binding complex and in pure solvent will be identical and cancel out in the relative binding energy providing its independence from the choice of dummy atoms.

For most systems, this method works well, but for some systems with certain symmetry, it will cause serious problems in the FEP simulation. To avoid possible instabilities and singularities caused by the missing bonded interactions involving dummy atoms, the remaining dihedral potential in the end state is transformed to the single minimum function that keeps dummy atoms at the initial state position. This ensures that in the end state the dummy atom will keep its position as in the initial state and does not switch to the position of other atoms of the COMMON part that results in clashes and instabilities. Although the distributions sampled for the hybrid molecule might be a bit different from the distributions for the molecule without the dummy atoms, the error introduced in this treatment is negligible because the fluctuations of bond angle and transformed dihedral are sufficiently small for the sets of ligands studied in this article. In addition, the error in the relative binding affinity becomes even smaller due to cancellation of errors between the free energies of the protein-ligand complex and ligands in the solvent.

For each λ-state an MD simulation of the molecular system is performed. Alchemical transformations were performed independently in protein and water environments. The relative binding free energy was calculated as a transformation free energy in a protein minus the transformation free energy in water. Statistical sampling of the molecular system conformations was enhanced by Hamiltonian replica exchange (HREX)[6] between the neighboring λ-states. Our typical alchemical transformation simulation went though 1,200 exchange cycles. Replica exchanges were attempted every 120 seconds in wall time, thus, the actual simulation time of each λ-replica varied. This strategy allowed us to efficiently use a cluster of diverse GPU's.

Enhanced sampling

HREX with potential softening

In order to compute relative free energies of ligand binding we performed simulations of alchemical transformation of ligands in protein and in water with conformational sampling enhanced by Hamiltonian Replica Exchange in the space of λ-replicas (HREX) [6]. The molecular Hamiltonian in the middle λ points is "softened" so the potential barriers between the

conformations of the ligand and surrounding protein residues are lowered and rates of transitions between the local potential minima are greatly increased. Conformations generated in the middle λ-replicas are propagated to the end points of the transition as a result of HREX, thus, after certain simulation time Boltzmann equilibrium distribution is established for the ligand and protein conformations and for all λ points of the transition. This approach is similar to the REST2 method.[7]

Disadvantages of this in-the-middle potential softening approach compared to regular λ-HREX simulations include the increase of the number of intermediate λ-states needed to ensure a sufficient acceptance rate of replica exchanges. The greater number of atoms involved in the softening of the Hamiltonian and greater the change in the Hamiltonian the more replicas are needed in alchemical HREX simulations. The scale of changes of dG/dλ values as a function of λ is also increased with an increased potential softening at middle λ values. This results in a larger statistical uncertainty of the computed ΔG values and longer MD simulation time needed for ΔG convergence. The larger number of λ-states also means a slower propagation of system conformations due to replica exchange from the middle to the end λ-states again resulting in a slower ΔG convergence.

HREX with conformation reservoir

Alternatively, the statistical sampling of protein-ligand geometries during the alchemical transformation was enhanced by attachment of a reservoir of conformations of the simulated system to the λ-HREX chain. The reservoirs were prepared beforehand and contained conformations from local potential minima of the ligand-protein distributed with Boltzmann probabilities. Random conformations from the reservoir were inserted every two exchange cycles to the λ=0 replica and were allowed to propagate along the chain of λ-replicas using HREX. The reservoir could be generated only once since the same reference ligand was used for a full series of mutations. Conformation reservoirs were prepared in two steps. In the first step a widely sampled conformational ensemble was generated in a single MD simulation by softening particular protein-ligand and ligand ligand interactions. In the second step the "softened" ensemble was converted to the "non-softened" ensemble using either HREX or non-equilibrium approach.

MD with softened potential

In order to efficiently sample conformations of ligands in the protein binding pockets we reduced the energetic barriers between potential minima by "softening" particular ligand-ligand and ligand-protein interactions. For ligand 27 in MCL1 torsion potentials along the ligand backbone (τ1-5, see **Figure S8a**) were scaled by a factor of 0.3. The non-bonded interactions within the ligand were scaled by a factor of 0.1 and soft-cored using the radius of 0.5 Å. Additionally, to allow the benzyl group to rotate, non-bonded interactions between this group and the protein were scaled by a factor of 0.1 and soft-cored using the radius of 1.5 Å. A molecular dynamics simulation with a potential softening of ligand 27 in MCL1 was run for 10 ns. 500 conformations were extracted from this trajectory using a 20 ps interval which further served as a softened ensemble.

Conformation reservoir generation using HREX

HREX was applied to generate a non-softened reservoir of conformations from the softened ensemble defined above. 17 replicas with the Hamiltonian spanning the softened (λ=0) to the non-softened (λ=1) potential were used. Random conformations from the softened ensemble were inserted into the softened replica (λ=1) every two exchange cycles, from where they had a chance to propagate toward the non-softened replica (λ=0) if favorable. 3,200 exchange cycles were performed. After skipping the first 1 ns as an equilibration the conformations sampled by the non-softened replica (λ=0) were extracted every 10 ps to generate a reservoir for the free energy calculations. The reservoir for mutations in MCL1 contained 683 geometries of ligand 27 - MCL1 complex, covering the accessible conformational space.

RESULTS

Enhanced sampling techniques comparison

We have shown that three enhanced sampling techniques, i.e. in-the-middle softening, HREX reservoir, and non-equilibrium reservoir, improve ΔΔG predictions for MCL1. Each of these allows sampling of ligand conformations otherwise omitted in the regular MD. The advantage of the in-the-middle softening method is its simplicity, since the enhanced sampling is a part of the mutation λ-HREX chain. However, it is also a weakness of this method, since maximum softening applies to the hybrid system (λ=0.5, mix of half grown reference and target ligands), thus, optimal softening parameters might differ between mutations. Additionally, ΔΔG convergence requires more λ-replicas and longer MD runs. As we have shown, 17/21 replicas simulated over roughly 3-6 ns (1,200 exchange attempts) do not guarantee stable results.

Application of the other two enhanced sampling techniques is more complex since it requires an additional step - generation of a conformation reservoir. For either HREX or non-equilibrium reservoir a softened trajectory needs to be generated beforehand. It can be considered as a “bottleneck” of these methods since selection and adjustment of the softening parameters is tricky. The softening should favor the same, or at least similar, important states as the regular Hamiltonian. On one hand, it should be sufficient to allow an easy transition between these states. On the other hand, softening should not be exaggerated, as a large phase space of available states can make the generation of a Boltzmann distributed ensemble impossible. Then, the softened ensemble needs to be converted to the non-softened ensemble, i.e. the reservoir, by either λ-HREX chain or non-equilibrium process. We have shown for MCL1 that conformation reservoirs produced with the both approaches are similar and the ΔΔG values obtained with them are close. However, the advantage of the non-equilibrium over the HREX method is that it can make use of a large number of diverse computational resources. In an extreme case, all the non-equilibrium processes can be run independently at once, thus, the overall performance depends only on the performance of a single process. In our study, generation of a reservoir that consists of 230 conformations of ligand 27 in MCL1 required

running of up to 3,900 non-equilibrium runs, 10 ps each (total MD length is ~40ns). In contrast, the HREX method needs parallel computing resources limited to the number of replicas, all run long enough to reach the overall equilibrium. In our study, 17 replicas of ligand 27 in MCL1 were run over 9.3-19.0 ns with 3,200 exchange attempts (total MD length is ~250ns). Although reservoir generation requires additional effort, it needs to be performed only once, for the reference ligand, as in our study, and new conformations can propagate efficiently to any target ligand along the mutation λ-HREX chain. Without in-the-middle softening only 11 λ-states were needed in the alchemical transitions for efficient ΔG integration and fast replica exchange. 4-6 ns MD runs for λ-states were used for converged results in HREX with reservoir calculations (total MD length ~50 ns per mutation), while computational cost for in-the-middle softening calculations were about doubled per mutation (~ 100 ns) and ΔΔG was still not fully converged.

Alternative experimental results for Thrombin

Figure S11 shows the computed ΔΔG values in Thrombin against alternative experimental results from isothermal titration calorimetry (ITC). Both series correlate well (r=0.96), however, the slope of the linear correlation is far from unity (slope=1.90). The correlation of computed results to the other set of binding ΔΔG experimental data reported in the same paper based on the fluorescence measurements of ligand substitution in the binding pocket (K_i) is somewhat worse (r=0.81), while the slope of linear fit of theoretical ΔΔG values to experimental ones is closer to unity (slope=0.96) (**Figure 5**). It is likely that experimental ITC data had a systematic error that resulted in a smaller linear correlation slope between experimental and theoretically computed ΔΔG values.

We found that stability of the ligand bound conformations and ΔΔG values of mutations in Thrombin are very sensitive to torsional parameters of the ligands. When ligand torsional parameters were fit to QM calculations the ligand was not completely stable in the protein pocket during long MD simulations. Therefore, in ΔΔG calculations we applied positional restraints to the part of the ligand after the C-N bond in the amide group of the ligand (see **Figure S8b**).

Enhanced sampling for CDK2

ΔΔG calculations with ARROW FF were performed for seven ligands bound to CDK2 protein using HREX simulations (no additional softening and reservoirs) starting from two DUMMY orientations, A and B, of the benzene ring of 1q1h ligand. Some “combined” A/B ΔΔG values were markedly different from the experimental values (**Figure S12a**). The attachment of the conformation reservoir obtained by NEQ MD runs did not change the computed ΔΔG significantly (**Figure S12b**). Thus, we conclude that insufficient sampling is an unlikely reason for deviation of theory from experiment in this case.

REFERENCES

(1) Pereyaslavets, L.; Kamath, G.; Butin, O.; Illarionov, A.; Olevanov, M.; Kurnikov, I.; Sakipov, S.; Leontyev, I.; Voronina, E.; Gannon, T.; Nawrocki, G.; Darkhovskiy, M.; Ivahnenko, I.;

Kostikov, A.; Scaranto, J.; Kurnikova, M. G.; Banik, S.; Chan, H.; Sternberg, M. G.; Sankaranarayanan, S. K. R. S.; Crawford, B.; Potoff, J.; Levitt, M.; Kornberg, R. D.; Fain, B. Accurate Determination of Solvation Free Energies of Neutral Organic Compounds from First Principles. *Nat. Commun.* **2022**, *13* (1), 414.

(2) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, R. D.; Fain, B. On the Importance of Accounting for Nuclear Quantum Effects in Ab Initio Calibrated Force Fields in Biological Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (36), 8878–8882.

(3) Donchev, A. G.; Galkin, N. G.; Pereyaslavets, L. B.; Tarasov, V. I. Quantum Mechanical Polarizable Force Field (QMPFF3): Refinement and Validation of the Dispersion Interaction for Aromatic Carbon. *J. Chem. Phys.* **2006**, *125* (24), 244107.

(4) Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.; Pereyaslavets, L. B.; Tarasov, V. I. Assessment of Performance of the General Purpose Polarizable Force Field QMPFF3 in Condensed Phase. *J. Comput. Chem.* **2008**, *29* (8), 1242–1249.

(5) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *J. Phys. Chem.* **1994**, *98* (5), 1487–1493.

(6) Woods, C. J.; Essex, J. W.; King, M. A. Enhanced Configurational Sampling in Binding Free-Energy Calculations. *J. Phys. Chem. B* **2003**, *107* (49), 13711–13718.

(7) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115* (30), 9431–9438.

(8) He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein--Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5* (9), 4611–4619.

(9) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17* (7), 4291–4300.

(10) Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L. Automated, Accurate, and Scalable Relative Protein–Ligand Binding Free-Energy Calculations Using Lambda Dynamics. *J. Chem. Theory Comput.* **2020**, *16* (12), 7895–7914.

(11) Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M., Jr. Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, *59* (7), 3128–3135.

(12) He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5* (9), 4611–4619.

(13) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. More than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the s1 Pocket of Thrombin. *J. Mol. Biol.* **2009**, *390* (1), 56–69.

mutation	ΔΔG (exp)	ΔΔG (MD) [A]	ΔΔG (MD) [B]	ΔΔG (MD) [A/B]	ΔΔG (MD) [res. HREX]	ΔΔG (MD) [res. NEQ)	ΔΔG (MD) [soft. A]	ΔΔG (MD) [soft. B]	ΔΔG GAFF[a]	ΔΔG OPLS4[b]	ΔΔG CGenFF[d]
27->28	-0.50	-0.96	0.20	-0.65	-0.97	-1.63	-1.04	-0.90	-0.54	0.30	-0.04
27->30	-1.73	-1.36	-0.06	-1.01	-1.21	-1.26	-0.49	-1.25	-0.74	-0.14	-0.60
27->35	-2.69	-2.56	-1.66	-2.27	-3.09	-2.44	-2.15	-2.33	-2.66	-1.42	-1.42
27->38	-0.85	-1.72	1.89	-1.31	-0.69	-1.34	-1.18	-0.77	-2.30	-2.28	-1.67
27->43	-0.91	-2.01	0.26	-1.60	-1.76	-1.16	-1.98	-1.33	-1.84	-0.41	-0.82
27->46	-1.48	-2.61	-1.32	-2.26	-2.38	-2.25	-2.04	-2.05	-2.55	-1.23	0.88
27->52	-3.11	-3.47	-1.89	-3.10	-1.79	-2.39	-2.90	-2.42	-2.63	-2.90	-2.05
27->36	-2.06	-3.15	-1.59	-2.78	-3.38	-3.11	-	-	-2.96	-1.95	-1.31
27->44	-2.55	-3.61	1.01	-3.21	-2.89	-2.85	-1.57	-2.48	-3.3	-2.40	-2.22
27->42	-2.78	-2.59	1.46	-2.16	-2.30	-2.67	-2.54	-1.55	-1.46	-1.80	-2.07
27->45	-2.83	-2.71	0.24	-2.30	-2.51	-2.75	-1.92	-2.23	-3.21	-1.37	-2.02
27->41	-1.01	-0.49	-0.49	-0.49	-0.27	-1.12	-0.49	-0.49	-2.34	-1.50	-
27->32	-0.46	0.08	0.08	0.08	-0.38	0.15	0.08	0.08	-0.82	-0.65	-0.67
27->33	-0.76	-1.16	-1.16	-1.16	-1.22	-0.85	-1.16	-1.16	-1.49	-1.31	-0.54
27->37	-2.83	-3.09	-3.09	-3.09	-2.94	-2.75	-3.09	-3.09	-3.3	-2.44	-3.25
27->39	-0.91	-0.26	-0.26	-0.26	-0.87	0.84	-0.26	-0.26	-2.61	-1.74	-2.03
27->53	-3.84	-2.73	-2.73	-2.73	-2.40	-2.50	-2.73	-2.73	-2.89	-3.69	-4.39
r	1.00	0.80	0.44	0.83	0.74	0.75	0.79	0.86	0.60	0.70	0.72
slope	1.00	0.74	0.33	0.83	0.79	0.72	0.88	1.00	0.52	0.67	0.85
MAE	0.00	0.61	1.38	0.54	0.59	0.56	0.66	0.49	0.82	0.66	0.77

Table S1. Relative binding free energy of ligands to MCL1. Values duplicated for symmetrical ligands are shown in gray. [a] [8], [b] [9], [d] [10]

mutation	ΔΔG (exp, ITC)	ΔΔG (exp, Ki)	ΔΔG (MD)[A]	ΔΔG (MD)[B]	ΔΔG (MD)[A/B]	ΔΔG (MD) [res. NEQ]	ΔΔG GAFF[c]	ΔΔG OPLS4[b]
5 -> 1a	0.10	-0.61	-0.26	0.76	0.06	0.35	-1.20	-0.47
5 -> 3a	-0.74	-1.32	-1.77	0.46	-1.38	-0.97	-2.24	-1.31
5 -> 1b	-0.88	-2.45	-2.28	-0.35	-1.89	-1.75	-2.63	-1.61
5 -> 6a	-1.60	-3.10	-3.49	-0.75	-3.10	-3.14	-2.82	-3.35
5 -> 6e	-1.33	-1.84	-3.03	-1.64	-2.67	-2.60	-2.56	-1.98
5 -> 7a	-0.64	-0.82	-2.45	-1.21	-2.11	-1.76	-1.97	-1.77
5 -> 3b	-0.28	-0.46	-2.08	0.75	-1.67	-0.59	-2.07	-1.57
5 -> 6b	-1.31	-2.23	-3.13	-2.41	-2.87	-2.14	-1.91	-2.27
		***r*:**	0.92	0.74	0.92	0.96	0.77	0.88
	ITC	**slope:**	1.62	1.52	1.63	1.90	0.69	1.27
		MAE:	1.47	0.78	1.12	0.80	-1.34	0.96
		***r*:**	0.71	0.52	0.71	0.81	0.73	0.77
	Ki	**slope:**	0.75	0.64	0.75	0.96	0.39	0.66
		MAE:	0.83	1.20	0.66	0.50	-0.57	0.44

Table S2. Relative binding free energy of ligands to Thrombin. [c] [11], [b] [9]

mutation	ΔΔG (exp)	ΔΔG (MD)[A]	ΔΔG (MD)[B]	ΔΔG (MD)[A/B]	ΔΔG (MD) [res. NEQ]	ΔΔG GAFF[a]	ΔΔG OPLS4[b]	ΔΔG CGenFF[d]
1h1q->1h1r	0.51	-1.67	-0.59	-1.52	-1.77	-1.32	-0.79	-0.44
1h1q->1oi9	-1.56	-0.20	-0.20	-0.20	0.29	-1.46	-1.83	-1.06
1h1q->1oiy	-1.61	-0.27	-0.27	-0.27	0.32	-1.17	-2.04	-1.65
1h1q->20	-0.54	-2.18	-0.57	-1.95	-1.35	-1.02	-1.10	-1.01
1h1q->21	0.35	-0.08	-0.51	-0.23	-0.25	-0.78	-0.06	-1.36
1h1q->22	0.32	0.65	0.11	0.43	0.12	-0.93	-0.28	-1.17
1h1q->26	-0.25	-0.05	-0.05	-0.05	0.02	-0.92	-0.88	-1.56
r	1.00	-0.04	-0.14	-0.09	-0.50	-0.79	0.93	0.54
slope	1.00	-0.04	-0.45	-0.09	-0.53	-0.04	0.77	0.19
MAE	0.00	1.07	0.73	1.00	1.13	0.85	0.60	0.92

Table S3. Relative binding free energy of ligands to CDK2. Values duplicated for symmetrical ligands are shown in gray. [a] [8], [b] [9]

mutation	ΔΔG (exp)	PIMD	ΔG (MD) [A] protein	ΔG (MD) [B] water	ΔΔG (MD)
1h1q->1oi9	-1.56	0	-3.62	-3.42	-0.20
1h1q->1oi9	-1.56	4	-3.23	-2.97	-0.26
1h1q->1oiy	-1.61	0	-13.21	-12.94	-0.27
1h1q->1oiy	-1.61	4	-12.65	-12.39	-0.25

Table S4. Relative binding free energy of two ligands to CDK2 determined with PIMD=4, taking nuclear quantum effect into account, and without it (same as in Table. S3).

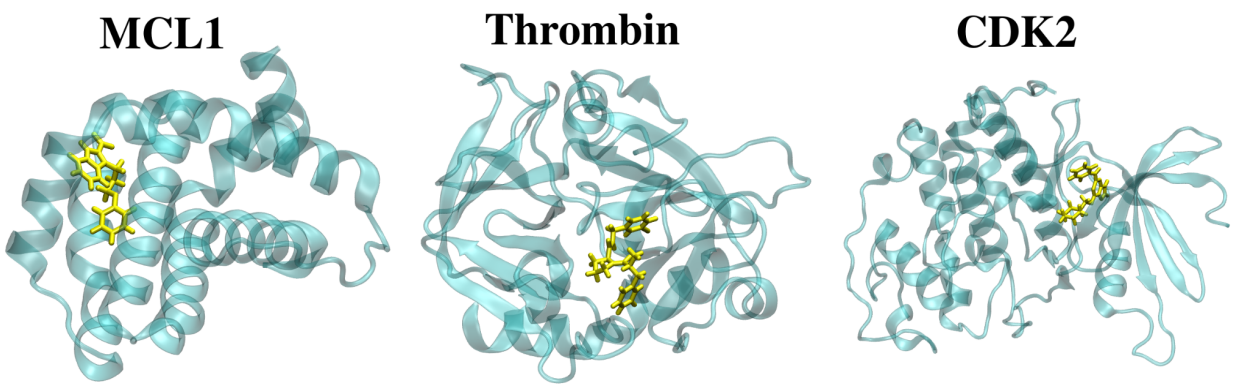


Figure S1. Proteins with reference ligands: MCL1 with ligand 27, Thrombin with ligand 5, and CDK2 with ligand 1h1q.

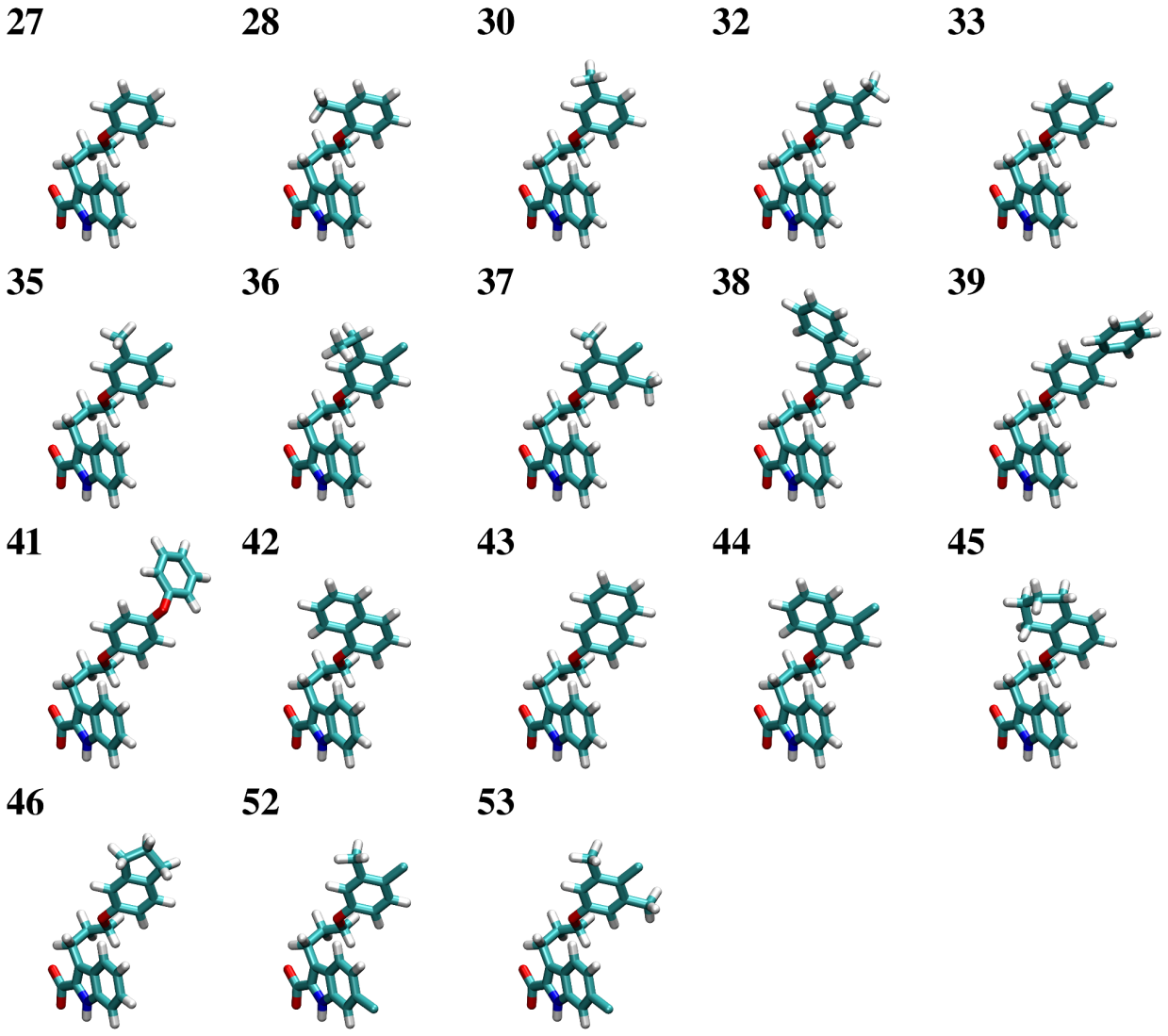


Figure S2. Ligands for MCL1.

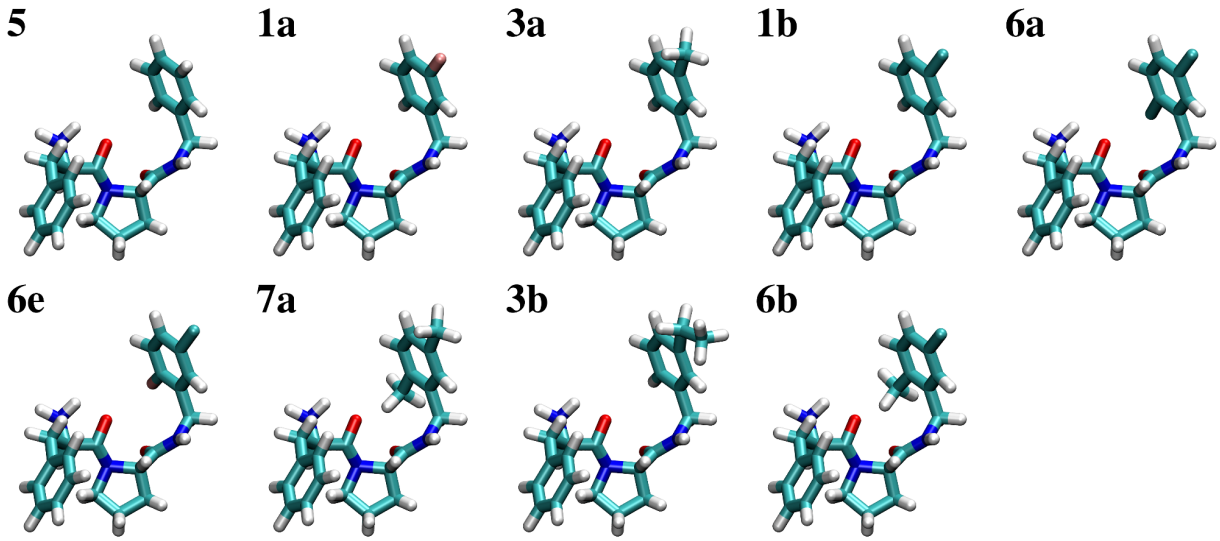


Figure S3. Ligands for Thrombin.

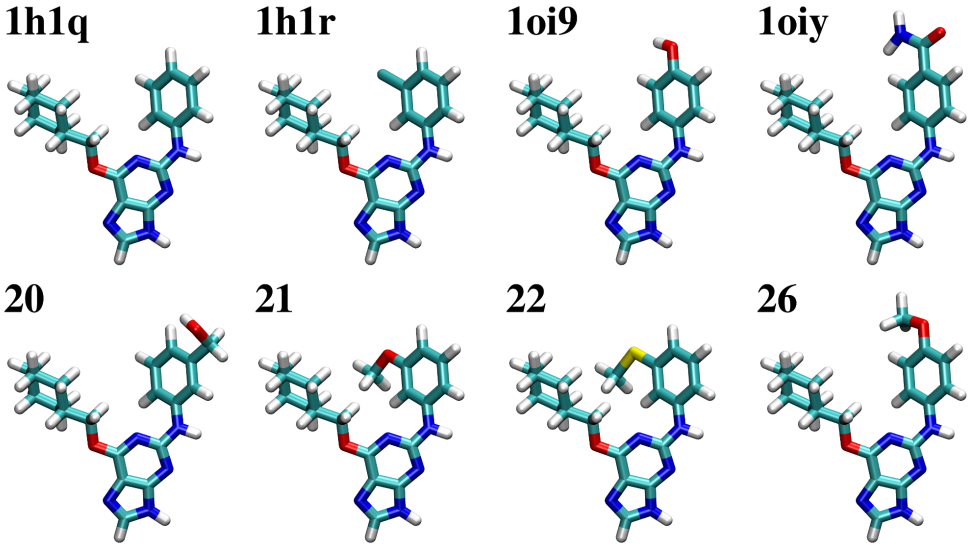


Figure S4. Ligands for CDK2.

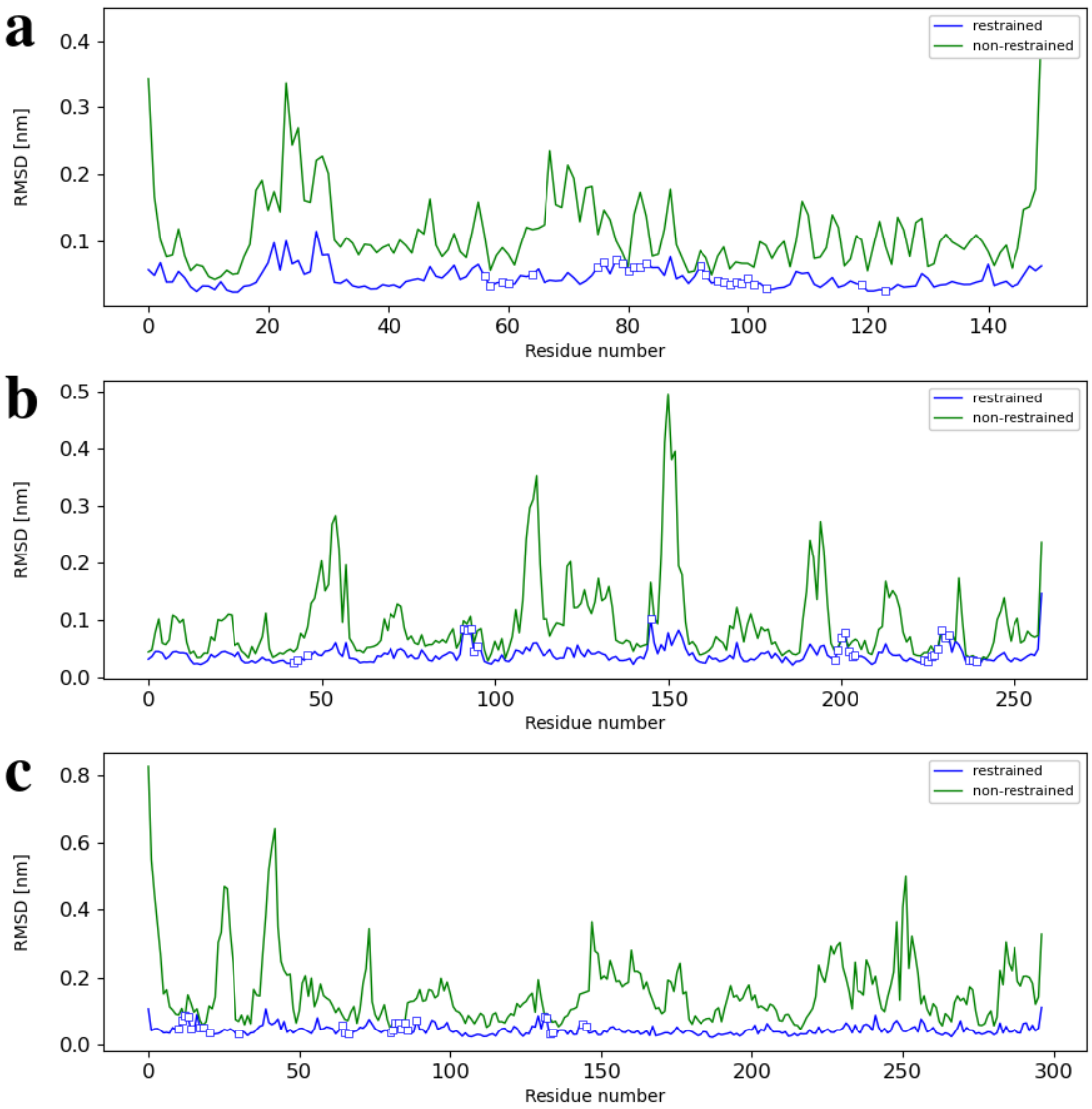


Figure S5. Root mean squared displacement (RMSD) of C_α atoms from the X-ray structures for **a)** MCL1 with ligand 27, **b)** Thrombin with ligand 6a, and **c)** CDK2 with ligand 1h1q. The RMSD was calculated from restrained and non-restrained simulations. In the restrained simulations, the C_α atoms of the binding pocket that were not restrained are shown by empty squares.

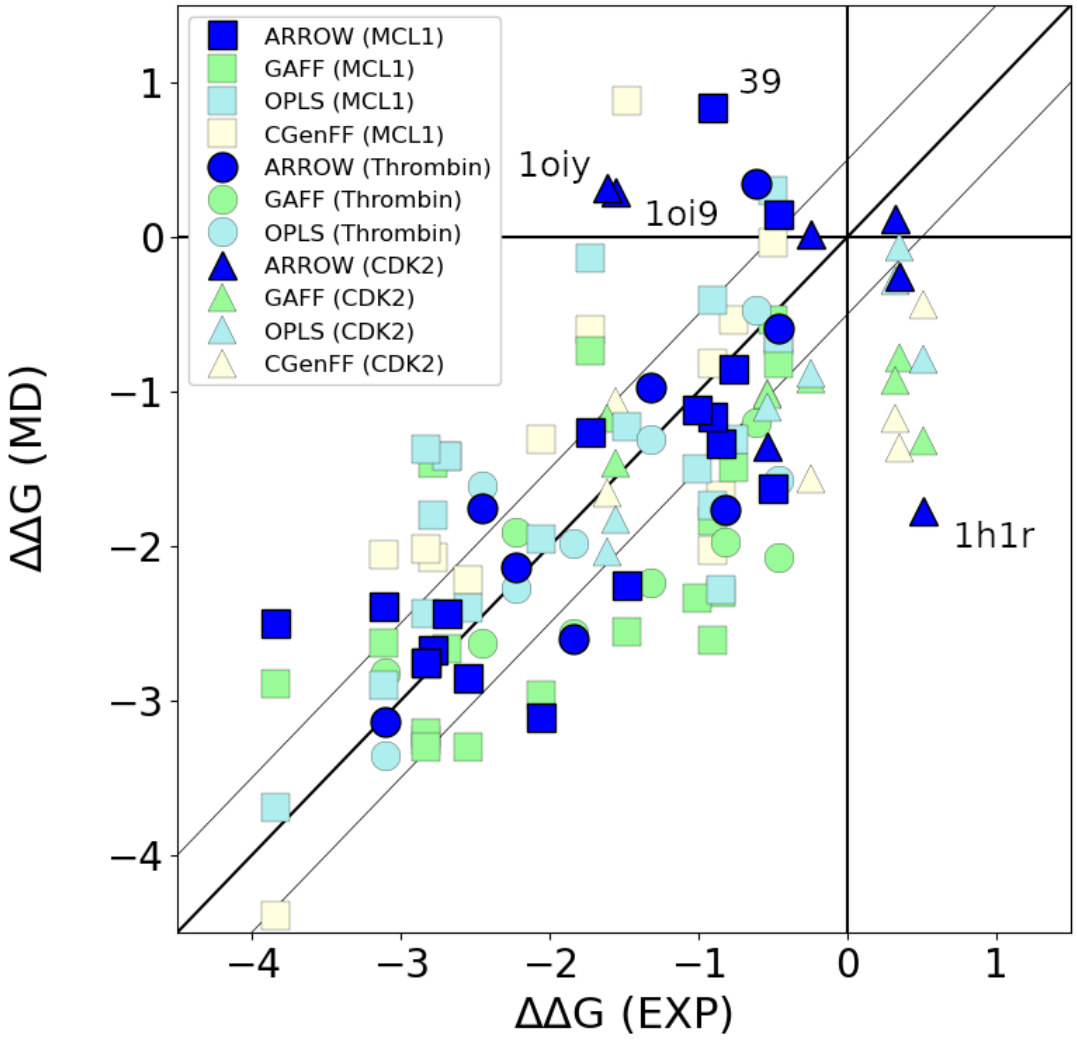


Figure S6. A parity plot comparing the relative binding free energies ΔΔG for ligand mutations in MCL1, Thrombin, and CDK2 as predicted by ARROW FF and experiment. Also shown for comparison are the predictions of GAFF [11,12], OPLS [9], and CGenFF [10] force fields. Results with ARROW were calculated with HREX and conformation reservoir generated *via* potential softening and NEQ MD. Selected ARROW ΔΔG values with the largest deviation from the experiment (MAE > 1.5 kcal/mol) are marked with labels - ligands 1h1r, 1oi9, and 1oiy in CDK2, and ligand 39 in MCL1. The thin gray lines are +/- 0.5 kcal/mol from the diagonal.

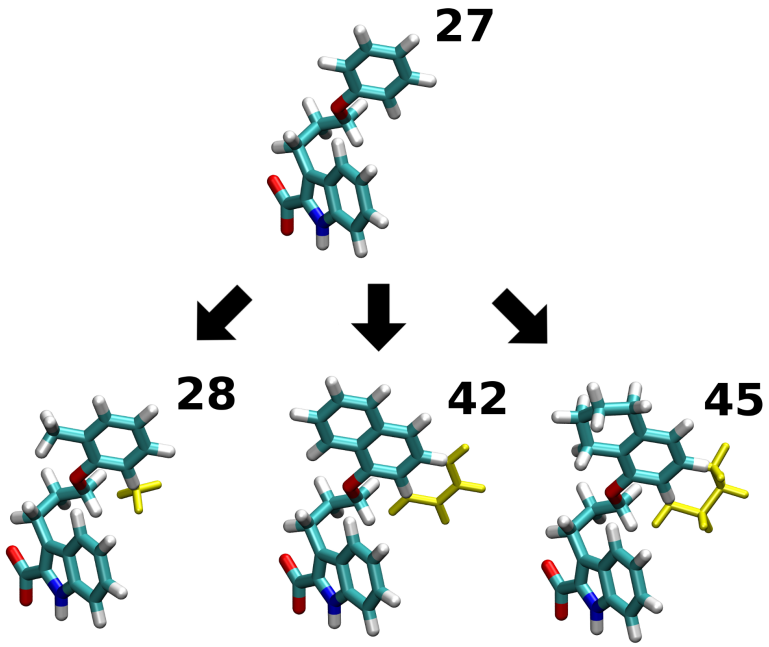


Figure S7. Examples of asymmetrical ligand mutations in MCL1. Alternative initial orientations of the mutated sites are shown in yellow.

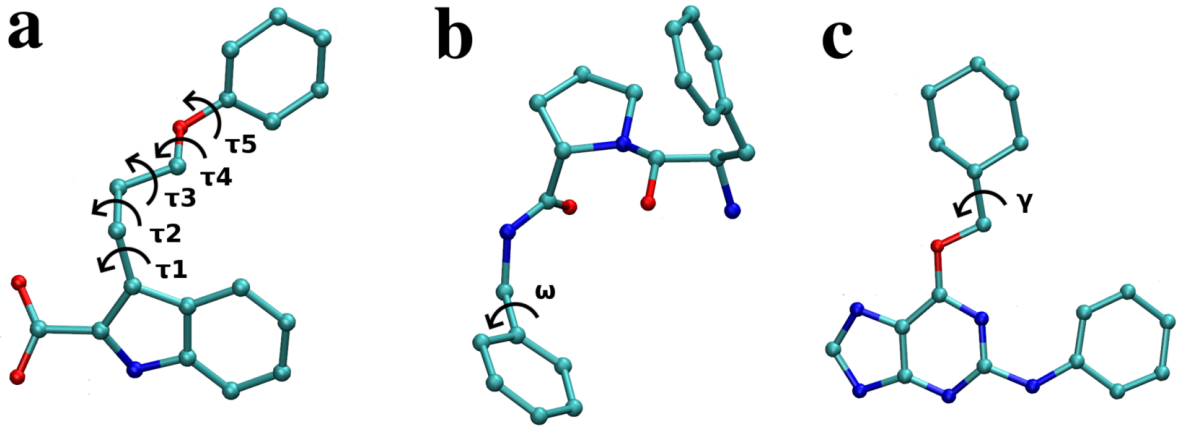


Figure S8. Ligand **a)** 27 of MCL1, **b)** 5 of Thrombin, and **c)** 1h1q of CDK2. Hydrogen atoms are not shown. Rotatable torsions softened to enhance sampling are shown with labels - $\tau 1, 2, 3, 4, 5$ for ligand 27, ω for ligand 5, and γ for 1h1q.

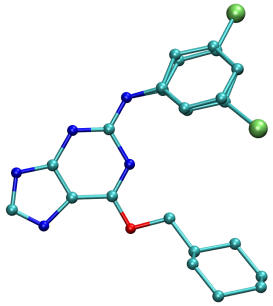


Figure S9. Alternative conformations of ligand 1h1r (chlorobenzene site) in CDK2 from X-ray structure (PDB: 1h1r, chain A).

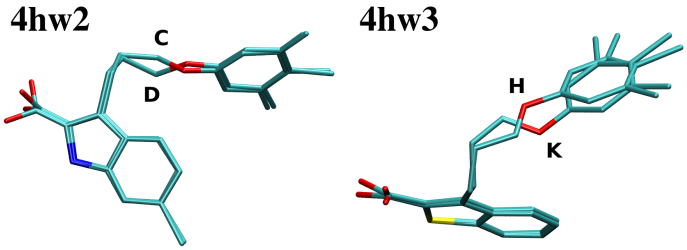


Figure S10. Selected conformations of ligands 53 and 60 in MCL1 from X-ray structures (PDB: 4hw2 and 4hw3, respectively) superimposed on each other. The conformations are labeled by their chain ID's.

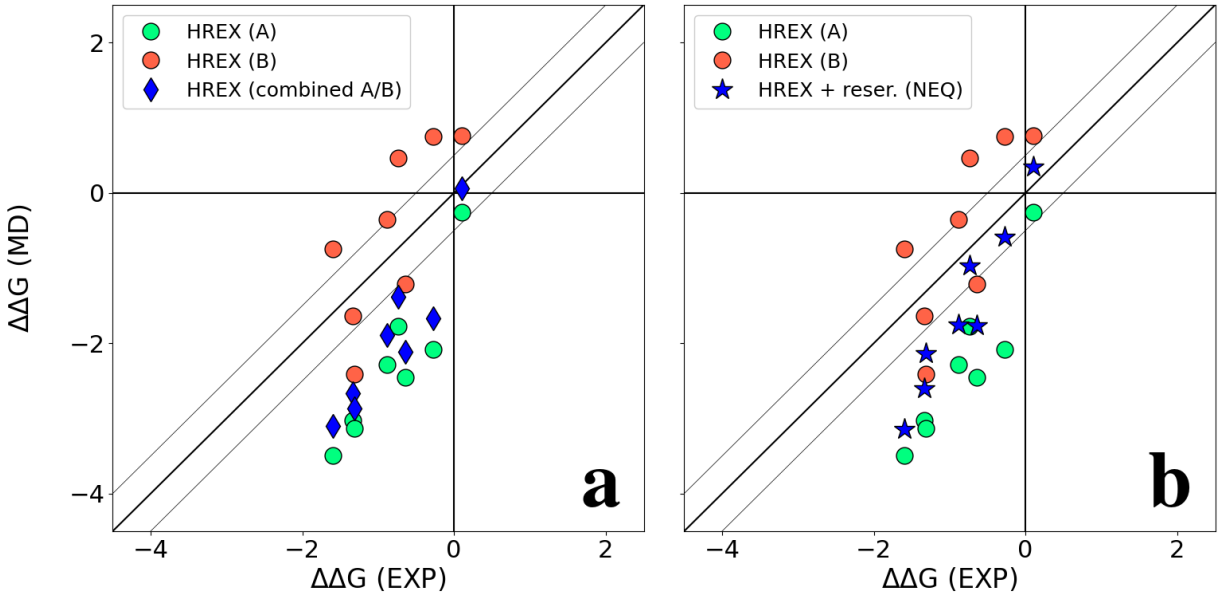


Figure S11. Comparison of ΔΔG as determined with ARROW force field and ITC experiment [13] for Thrombin. Green and red markers correspond to values obtained with HREX and a target ligand starting at A and B sides, respectively. Blue markers correspond to **a)** combined A/B sides, **b)** HREX with reservoir (from NEQ). Thin gray lines are +/- 0.5 kcal/mol from the diagonal.

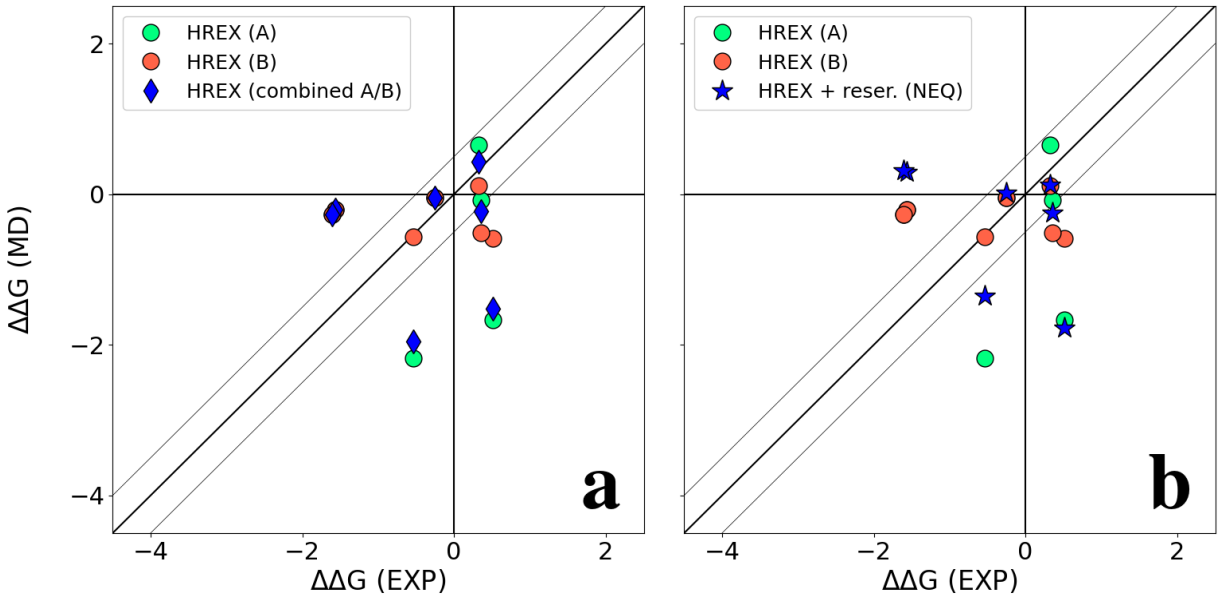


Figure S12. Comparison of ΔΔG as determined with ARROW force field and experiment for CDK2. Green and red markers correspond to values obtained with HREX and a target ligand starting at A and B sides, respectively. Blue markers correspond to **a)** combined A/B sides, **b)** HREX with reservoir (from NEQ). Thin gray lines are +/- 0.5 kcal/mol from the diagonal.

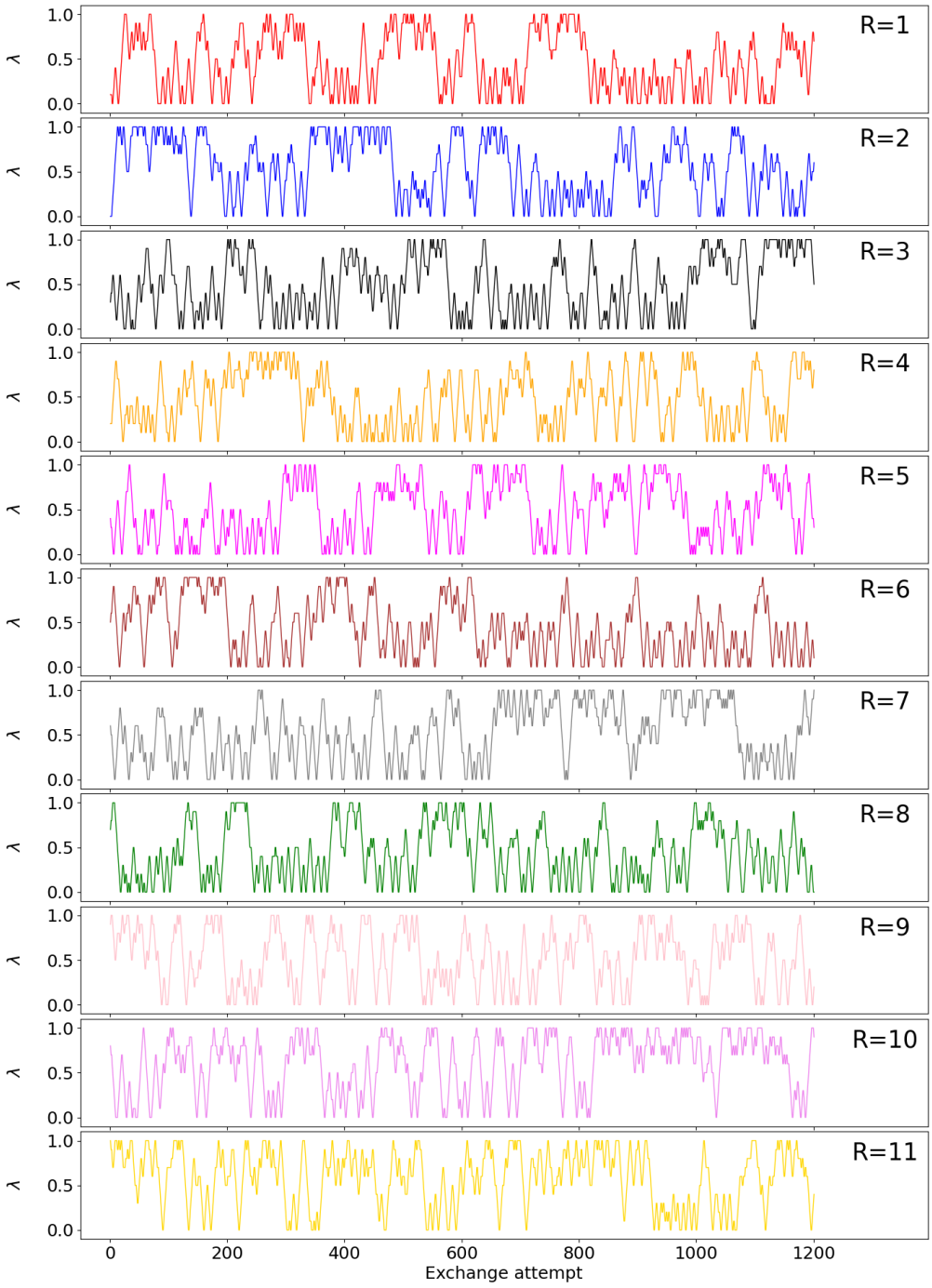


Figure S13. Dynamics of 11 system replicas in the space of Hamiltonians spanning from λ=0.0 (ligand 27) to λ=1.0 (ligand 30) in MCL1 protein during HREX simulation. A conformation reservoir (from NEQ) was attached to λ=0.0 replica. The neighboring replicas are exchanging with the following rates: 0.0↔0.1: 0.85, 0.1↔0.2: 0.89, 0.2↔0.3: 0.90, 0.3↔0.4: 0.77, 0.4↔0.5: 0.67, 0.5↔0.6: 0.66, 0.6↔0.7: 0.65, 0.7↔0.8: 0.69, 0.8↔0.9: 0.64, 0.9↔1.0: 0.59.

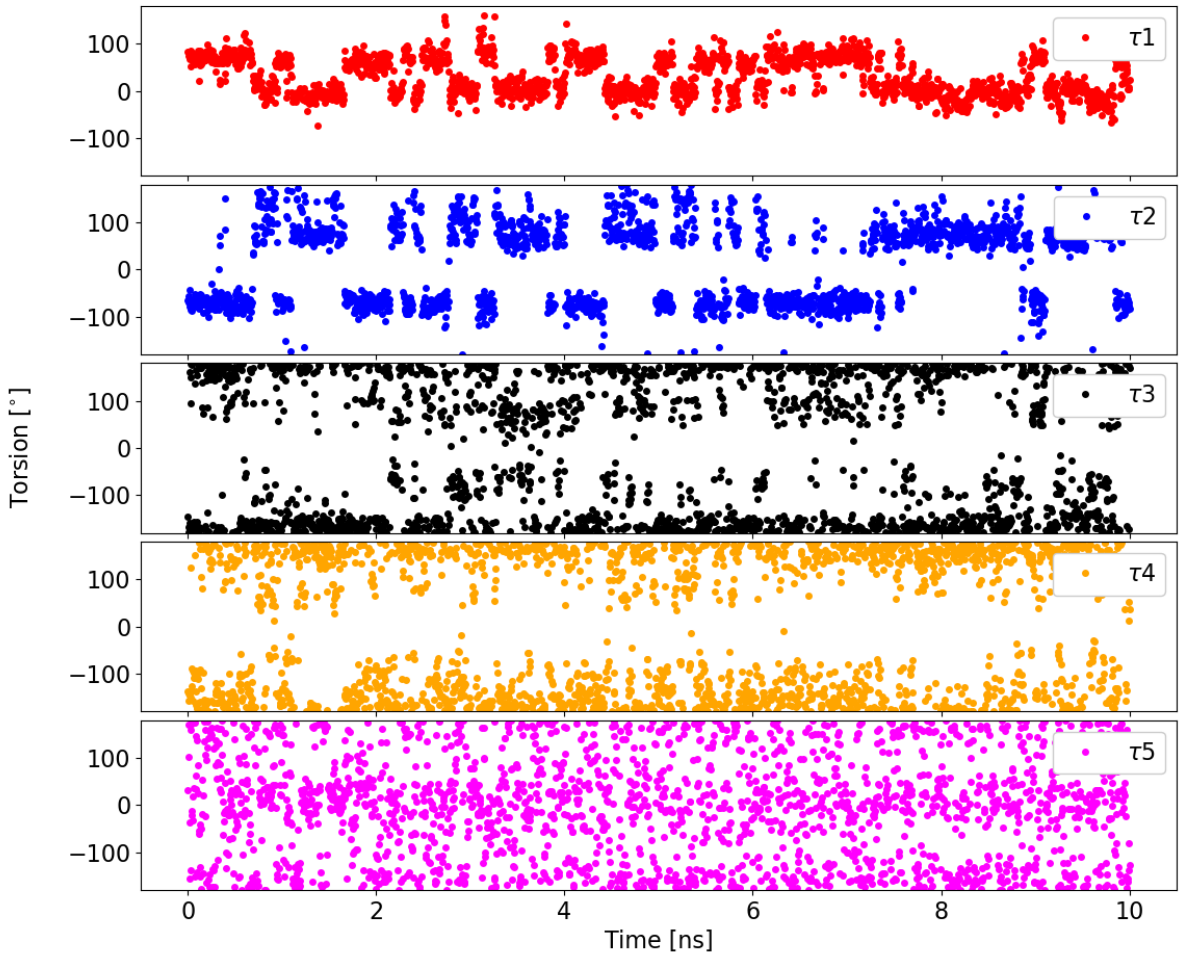


Figure S14. Dynamics of rotable torsions of ligand 27 in MCL1 protein during 10 ns MD with softened potential.

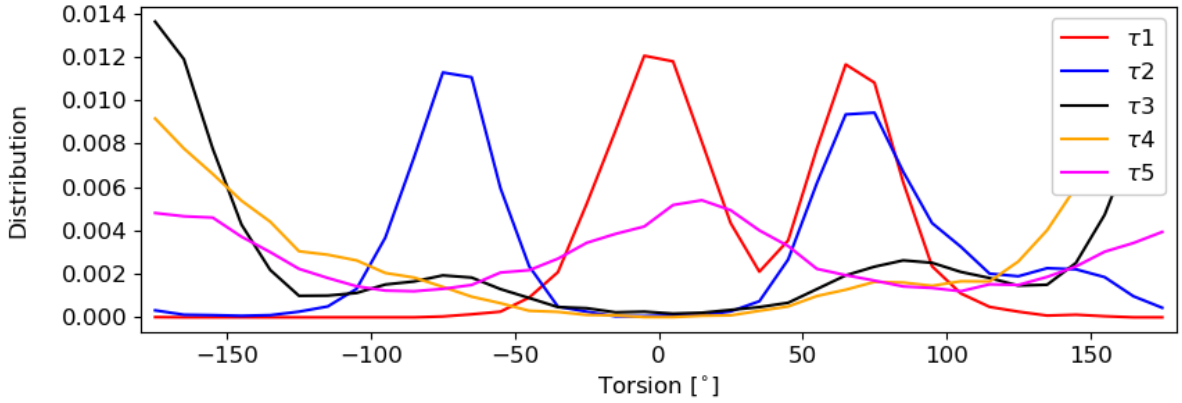


Figure S15. Distribution of rotable torsions of ligand 27 in MCL1 protein during 10 ns MD with softened potential.