HHS Public Access

Author manuscript

*J Comput Aided Mol Des*. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

*J Comput Aided Mol Des*. 2018 January ; 32(1): 89–102. doi:10.1007/s10822-017-0050-5.

# CDOCKER and $\lambda$-dynamics for prospective prediction in D3R Grand Challenge 2

**Xinqiang Ding**[1],
Department of Computational Medicine & Bioinformatics, Univeristy of Michigan Ann Arbor, MI 48109

**Ryan L. Hayes**[1],
Department of Chemistry, Biophysics Program, University of Michigan Ann Arbor, MI 48109

**Jonah Z. Vilseck**[1],
Department of Chemistry, Biophysics Program, University of Michigan Ann Arbor, MI 48109

**Murchtricia K. Charles**, and
Department of Computational Medicine & Bioinformatics, Univeristy of Michigan Ann Arbor, MI 48109

**Charles L. Brooks III**
Department of Computational Medicine & Bioinformatics, Univeristy of Michigan Ann Arbor, MI 48109Department of Chemistry, Biophysics Program, University of Michigan Ann Arbor, MI 48109

## Abstract

The opportunity to prospectively predict ligand bound poses and free energies of binding to the Farnesoid X Receptor in the D3R Grand Challenge 2 provided a useful exercise to evaluate CHARMM based docking (CDOCKER) and $\lambda$-dynamics methodologies for use in "real-world" applications in computer aided drug design. In addition to measuring their current performance, several recent methodological developments have been analyzed retrospectively to highlight best procedural practices in future applications. For pose prediction with CDOCKER, when the protein structure used for rigid receptor docking was close to the crystallographic holo structure, reliable poses were obtained. Benzimidazoles, with a known holo receptor structure, were successfully docked with an average RMSD of 0.97 Å. Other non-benzimidazole ligands displayed less accuracy largely because the receptor structures we chose for docking were too different from the experimental holo structures. However, retrospective analysis has shown that when these ligands were re-docked into their holo structures, the average RMSD dropped to 1.18 Å for all ligands. When sulfonamides and spiros were docked with the apo structure, which agrees more with their holo structure than the structures we chose, 5 out of 6 ligands were correctly docked. These docking results emphasize the need for flexible receptor docking approaches. For $\lambda$-dynamics techniques, including multisite $\lambda$-dynamics (MS$\lambda$D), reasonable agreement with experiment was observed for the 33 ligands investigated; root mean square errors of 2.08 and 1.67 kcal/mol were obtained for free energy sets 1 and 2, respectively. Retrospectively, soft-core potentials, adaptive landscape flattening, and biasing potential replica exchange (BP-REX) algorithms were critical to

[1]These authors contributed equally for this work.

model large substituent perturbations with sufficient precision and within restrictive timeframes, such as was required with participation in Grand Challenge 2. These developments, their associated benefits, and proposed procedures for their use in future applications are discussed.

**Keywords**

D3R; Drug Design Data Resources; Protein-ligand Docking; Free Energy Calculation; CDOCKER; lambda dynamics

## 1 Introduction

The prediction of protein-ligand binding poses and binding affinities with accuracies sufficient to guide experimental drug discovery is an ongoing challenge for the field of computer aided drug design. A plethora of computational methodologies exists from which insights into target specific molecular recognition can be gained [5,6, 26,27,47], including docking and virtual screening approaches [34,29,30], QSAR and informatics[5], and rigorous all-atom free energy and molecular dynamics simulations [26,2,13,1]. Yet community-wide blinded challenges, including the SAMPL [51], CSAR [11], and D3R Grand Challenges [19], continue to teach us that additional improvements are needed. For example, docking results from CSAR 2014 and D3R Grand Challenge 2015 exercises suggest that scoring or rank-ordering ligand affinities is more challenging than pose prediction, and, as observed in previous challenges, scoring results did not improve even when crystal structures of bound ligands were supplied (often termed “Stage 2” scoring or rank-ordering). Extrinsic factors in the docking setup, system preparation, and pose selection based on user “chemical intuition” were further observed to cause differences in predictive accuracies [11,19]. For binding free energy predictions, errors have typically ranged 1–2 kcal/mol for even the most successful methods, and explicit solvent, alchemical free energy simulations have tended to provide greater reliability over other free energy approaches [51,19].These blinded challenges have been useful exercises to evaluate method performance in a prospective manner, and facilitate community wide comparisons across many methodologies [51,11,19]. We note that there is a lack of community sharing of the underlying force field or scoring functions between participants. We suggest that significantly more progress would come if all participants had the opportunity to utilize their approaches with the force fields and scoring functions used by others.

In this D3R challenge, referred to as Grand Challenge 2, participants were invited to investigate a number of different ligands bound to the Farnesoid X Receptor (FXR). Pharmaceutically, FXR is a target of interest to treat a number of metabolic and vascular diseases, including dyslipidemia, atherosclerosis, and diabetes [17,37,36,3], but, computationally, FXR is challenging to address. For example, it has known conformational flexibility in the binding site, and binds a variety of ligands with differing size, shape, chemical moieties, and binding poses [17,37,36,3]. As part of this challenge, the ligands to investigate included known benzimidazole and isoxazole based ligands [17, 37,36,3], as well as two previously unpublished series of sulfonamide and spiro containing compounds and a handful of miscellaneous molecules. In total, 102 ligands with experimental affinities

($IC_{50}$s) in the high picomolar to micromolar range were provided for investigation, including 10 inactive molecules (Fig. 1). Stage 1 of Grand Challenge 2 focused on pose prediction and ordinal ranking of ligand affinity. Upon completion, 36 crystal structures were released typifying the binding mode for each ligand series. Stage 2 then repeated docking and scoring of all 102 ligands. Alchemical free energy calculations were welcomed at either stage of the challenge.

Our own objective in participating in Grand Challenge 2 was to evaluate the performance of two CHARMM based simulation technologies [10,9,25]: (i) molecular dynamics based docking (CDOCKER) [49,18] and (ii) $\lambda$-dynamics free energy methodologies [33,32,4,22,15]. Introduced at the turn of the century, CDOCKER was developed to deliver improved docking efficiency and accuracy via the exploration of ligand conformational space with molecular dynamics simulated annealing and soft-core potentials. Using a rigid, grid-based potential for the protein, and an all-atom representation of the ligands, a docking success rate of ~70% was achieved. Accuracy further improved by 6–10% when docked structures were subjected to a full force field minimization as the final step to docking [49]. More recent improvements to CDOCKER include flexible side chain sampling concurrent with ligand docking [18]; however, due to time restraints, these new advances were not fully explored in this challenge. For $\lambda$-dynamics, a number of improvements have been made within the past 5–7 years. Notably, the ability to model many substituent perturbations at multiple sites around a ligand core, termed multisite $\lambda$-dynamics (MS$\lambda$D), has allowed much larger chemical spaces to be explored than previously possible [32,4,22]. More recently, algorithmic advances have introduced biasing-potential replica exchange to accelerate $\lambda$-space sampling, new soft-core potentials to facilitate small to large substituent perturbations, and new biasing potentials to flatten the alchemical free energy landscape and improve end-point sampling [4,22]. Many of these improvements are still in the developmental phase. The opportunity to explore their usefulness in a “real-world” application afforded by Grand Challenge 2 proved a useful exercise to evaluate the current status of their progress and encourage on-going developments.

In this publication, we briefly review the approaches taken for pose prediction and free energy calculation. We do not discuss affinity scoring because we used the scoring function in Autodock Vina[41], which is not developed in our lab. We discuss our approach in solving the problems posed by Grand Challenge 2 and the submitted results. For $\lambda$-dynamics, results obtained after the submission deadline with methodological developments happening concurrently with participation are also discussed. For both CDOCKER and $\lambda$-dynamics, this report includes a retrospective analysis of method developments and subsequent results to highlight areas of improvement for future applications.

## 2 Methods

### 2.1 Ligand Parameterization

For simplicity, ligand numbering will be referred to by number only (e.g. FXR_10 will be referred to as 10). Smile strings supplied by D3R were converted into three dimensional ligands using MarvinSketch [35]. Ligand protonation states were assigned manually such that carboxylic acids were deprotonated and the amine in ligand 2 was protonated. Prior to

force field parameterization, each ligand was geometrically optimized using PM6 semi-empirical calculations in Gaussian [39,20]. Ligands were then parameterized using ParamChem [44,45], a molecular typing engine corresponding to the CHARMM general force field [43–45].

### 2.2 Protein-ligand Docking

In this challenge, the CHARMM-based CDOCKER [49] method was used for docking. CDOCKER is a protein-ligand docking program written in CHARMM script that makes use of the diverse functionalities available in CHARMM, such as the CHARMM force field, atom manipulation, energy minimization, and molecular dynamics engine [9]. To accelerate energy and force calculations, CDOCKER uses a grid-based potential for representing non-bonded interactions between protein and ligand, i.e., rigid receptor docking, which can be pre-calculated based on the structure of the protein. The docking protocol described in a prior article was followed in this challenge [49]. Each stage of the docking protocol, including generation of grid potentials for non-bonded interactions between protein and ligand, generation of initial conformations of ligands, molecular dynamics (MD) based simulated annealing and minimization, and ranking of docked poses, is detailed in the following sections.

**Generating Grid Potentials for FXR Non-bonded Interactions—**To accelerate the calculation of energies and forces in simulated annealing and energy minimization applications described below, a grid potential was generated around the FXR binding pocket for both van der Waals and electrostatic interactions. The size of the grid was equal to the longest dimension of the native ligand plus 10 Å. The center of the grid is placed at the geometric center of the ligand. The grid spacing was 0.5 Å. To facilitate the MD-based simulated annealing algorithm to search for the lowest energy pose, the following soft-core potential was used to calculate the grid potential:

$$E_{ij}(r_{ij}) = E_{\max} - a \cdot r_{ij}^{b} \quad \text{if} |E_{ij}^{*}| > \frac{|E_{\max}|}{2} \qquad (1)$$

where $E_{ij}^{*}$ is the regular non-bonded interaction energy, without soft-core potentials. $E_{\max}$, $a$, and $b$ are constants. The value of $E_{\max}$ controls the level of softening for the soft-core potential. The values of $a$ and $b$ are determined by the conditions that the energy and force at the cutoff distance are continuous. Two grid potentials, referred to as soft-grid-1 and soft-grid-2, with different levels of softening, i.e., different values of $E_{\max}$, were calculated and used in the following simulated annealing stages. The detailed values of $E_{\max}$ are shown in Table 1.[49]

**Generation of Initial Ligand Conformations—**Prior to performing MD-based simulated annealing to dock the ligands to FXR, initial low energy conformations of each ligand in vacuum were computed. With simulated annealing, ligands were heated up from 300 K to 1000 K in $10^4$ MD steps with a step size of 1 fs and then cooled back down from 1000 K to 100 K in $2 \times 10^4$ steps with a step size of 1 fs. Simulated annealing was repeated

200 times, identifying 200 initial conformations per ligand. Final ligand conformations were subsequently optimized using steepest descent energy minimization for 100 steps.

**Molecular Dynamics based Simulated Annealing and Minimization—**MD-based simulated annealing was next used to search for the lowest energy bound pose for each ligand. Initial starting coordinates were generated by random rotations and translations of each of the 200 initial ligand conformations generated in the above step until the interaction energy was smaller than 3000 kcal/mol. Simulated annealing then commenced in two phases: a heating and a cooling phase. In the heating phase, the soft-grid-1 grid potential was used for calculating the non-bonded interaction energies between ligands and FXR, and MD was run with temperature increases from 300 K to 700 K in 4.5 ps. The cooling phase comprised three stages. In the first stage, the soft-grid-1 grid potential was used and the temperature in the MD simulations was decreased from 700 K to 300 K over 21.0 ps. In the second and third stages, the soft-grid-2 grid potential was used. In the second stage, the temperature was decreased from 500 K to 300 K in 10.5 ps, and in the third stage, the temperature was decreased from 400 K to 50 K in 4.5 ps. The MD step size used in all simulated annealing stages was 1.5 fs. Finally, the grid potential was deleted and the terminal ligand conformations were minimized within the all-atom protein environment. In the minimization, the van der Waals and electrostatic interactions were switched off between 8 Å and 10 Å using potential switching and force switching, respectively. A distance dependent dielectric constant of 3 was used for electrostatic interactions. The above MD-based simulated annealing and minimization protocol is based on the CDOCKER paper. [49]

**Ranking Resulting Poses—**Each initial conformation of a ligand was repeatedly docked 21 times using MD-based simulated annealing and minimization. Overall, the docking protocol generated 4200 poses for each ligand. These resulting poses were ranked based on the sum of intra-ligand energy and protein-ligand interaction energies. Five poses were selected from the lowest energy poses for each ligand that were also structurally distinct, with the root mean square deviations (RMSDs) of at least 1 Å between all selected poses. For benzimidazole ligands, only the lowest scored pose was submitted for evaluation. For all other ligands, all 5 selected poses were submitted. These selected poses were ranked using AutoDock Vina [41], and the lowest score for each ligand was submitted as the binding affinity score.

### 2.3 Free Energy Methods

Grand Challenge 2 specified two sets of ligands for computing relative binding free energies. The first set was composed of 15 sulfonamides, while the second set was composed of 18 spiros. Multisite $\lambda$-Dynamics [33,32] was used to predict relative binding free energies within these two sets of compounds. MS$\lambda$D is a free energy technique which has undergone a renaissance in recent years. This free energy challenge provided an opportunity to test many new MS$\lambda$D developments.

In standard free energy methods, like FEP [52] and TI [40], $\lambda$ is an alchemical parameter that tunes the potential between two ligands. Simulations are then run at closely spaced values of $\lambda$ to compute relative free energies between the two ligands in a series of discrete

steps. In MS$\lambda$D, $\lambda$ is treated as a dynamic variable with velocity and mass that moves under the forces due to the potential. $\lambda$ may be generalized to a multidimensional $\lambda$ space to consider transitions between several different ligands within a single simulation, resulting in greater computational efficiency [33]. In addition, $\lambda$ variables may be associated with perturbations at multiple sites [32]. With these generalizations, the potential energy function becomes

$$V(\{x\},\{\lambda\})=V(x_0,x_{si})+\sum_{s}^{M}\sum_{i}^{N_s}\lambda_{si}(V(x_0,x_{si})+V(x_{si},x_{si}))+\sum_{s}^{M}\sum_{t>s}^{M}\sum_{i}^{N_s}\sum_{j}^{N_t}\lambda_{si}\lambda_{tj}V(x_{si},x_{tj})+V_{\mathrm{Bias}}(\{\lambda\}) \quad (2)$$

where $M$ is the number of substitution sites, $N_s$ is the number of competing substitutions at site $s$, $\lambda_{si}$ is the alchemical scaling factor for substituent $i$ at site $s$, $V(x_0, x_0)$ is the interaction of the environment atoms among themselves, $V(x_0, x_{si}) + V(x_{si}, x_{si})$ is the interaction of the atoms $x_{si}$ in site $s$, substituent $i$ with the environment and themselves, $V(x_{si}, x_{tj})$ is the interaction between atoms from two different sites, and $V_{\mathrm{Bias}}$ is a biasing potential used to enhance sampling. Relative free energies are determined from the free energy difference between two thermodynamic half cycles, in this case within bound and solvated simulations of the ligands. Within each half cycle, the free energy may be estimated from the populations of each physical ligand according to $\Delta G = -k_B T \ln P$. Populations are determined by counting the number of frames for which the $\lambda$'s of a particular ligand are greater than a cutoff of 0.99 at all sites.

Recent advances in MS$\lambda$D include the development of implicit constraints, biasing potential replica exchange (BP-REX), adaptive landscape flattening, the introduction of soft-core potentials, and porting CHARMM MS$\lambda$D simulations to GPUs. Constraints are required to ensure that the $\lambda$ variables at a specific site remain between zero and one, and sum to unity. With implicit constraints, these conditions are automatically satisfied, with the added bonus that the implicit constraints may be tuned to focus sampling on the physically relevant endpoints [31]. In the implicit constraints, a new set of variables, $\theta_{si}$, become the alchemical dynamical variables, and these $\theta$ values are mapped back to $\lambda_{si}$ values by the relation

$$\lambda_{si}=\frac{\exp(c\sin(\theta_{si}))}{\sum_{j}^{N_s}\exp(c\sin(\theta_{sj}))} \quad (3)$$

BP-REX uses replica exchange to swap biasing potentials among simulations, to drive the system between physical states, and has been shown to substantially accelerate sampling [4]. Adaptive landscape flattening adds biasing potentials to remove barriers in alchemical space to speed transitions between substituents [22]. In particular, deep endpoint traps occur for large perturbations, and biasing potentials that flatten these traps speed sampling by several orders of magnitude and thus allow MS$\lambda$D to explore perturbations like the ones performed

in this work. Beyond the cited biasing potentials [22], which included linear, quadratic, and endpoint terms, an additional biasing potential of the form

$$V_{\text{Bias}} = \sum_{s}^{M} \sum_{i}^{N_s} \sum_{j \neq i}^{N_s} \chi_{si,sj} \lambda_{sj} \exp(-\lambda_{si}/\sigma) \quad (4)$$

with $\sigma = 0.18$ was also used. The introduction of soft-core interactions has led to much more robust and self-consistent free energy estimates [22], so soft cores were used universally in this work. Finally, MS$\lambda$D has been introduced into the domdec [25] module of CHARMM [10,9] and ported to Graphic Processing Units (GPUs). This very recent development meant we were also assessing code correctness during D3R, but this advance enabled substantial simulation speedups.

**Simulation Setup—**MD-based MS$\lambda$D simulations employed similar procedures as previously reported [4]. Simulations were run in CHARMM [10,9] using the domdec package [25]; retrospective BP-REX simulations were run using a modified version of CHARMM still under development within the CHARMM community. The CHARMM36 force field for proteins [7,8] was used with TIP3P water [28] and ligand parameters were obtained using MATCH [50] and the CGenFF force field [43]. Protein simulations were started from crystal structures 1hqmf with ligand 17 and 1kjyp with ligand 12 provided by D3R at the end of stage 1 of the challenge. Titratable residues in the proteins were assigned their typical protonation states corresponding to physiological pH. Specifically, aspartic and glutamic acids were deprotonated, and lysine and arginine were protonated. Histidines in free energy set 1 were arbitrarily protonated on the $\delta$N. Histidines in free energy set 2 were protonated on the N, except for His298, which was protonated on the $\delta$N, as guided by the Reduce program [48] and the Molprobity webserver [12]. Protein simulations were run in 100 mM NaCl with at least a 10 Å buffer of water on each side of the protein-ligand complex in a cubic cell with periodic boundary conditions. Solvent simulations were similarly run in pure water with 10 Å between the ligands and the periodic box edge. Simulations were initially run in the NVT ensemble at 298 K using Langevin dynamics with a friction coefficient of 10 $ps^{-1}$ and later also run in the NPT ensemble at 1 atm and 298 K using a Langevin pressure piston and a Hoover thermostat. MD time steps of 1.6–2.0 fs were used for all simulations and SHAKE was used to constrain all hydrogen-heavy atom bond lengths [42]. Non-bonded cutoffs were set to 12 Å and were force switched to zero from 10 to 12 Å. Particle mesh Ewald (PME) [14,16] methods were not used to treat long-range electrostatic interactions in these calculations. Unless otherwise noted, dihedral potentials were not scaled by $\lambda$ in order to ensure reasonable geometries even when ligands were switched off. CHARMM NOE restraints were used to harmonically restrain analogous atoms in different substituents to the position of that atom in the first substituent. This increases phase space overlap and improves sampling. In FE1, the first heavy atoms of the substituents were restrained together. Since MS$\lambda$D in CHARMM has not implemented core swaps yet, FE1_4 required a more complex treatment: atoms in the thiophene and R1 phenyl rings were shared by the two ligands, while the bicycle and R2 phenyl rings were duplicated

to represent both ligands. Duplicated portions were held together by 6 restraints. In FE2 the 1 positions of all rings and the 4 positions of six-membered rings were restrained together.

In order to calculate relative free energies with MSλD, each free energy set was divided into subgroups of similar ligands. Free energy set 1 (FE1) was divided into FE1_1 (17, 49, 46, 101, 98, 96, 102), FE1_2 (17, 91, 47, 99, 95, 48, 100), FE1_3 (17 and 45), and FE1_4 (91 and 93). Free energy set 2 (FE2) was divided into FE2_1 (10, 12, 76, 84, 74, 85, 88), FE2_2 (12, 89, 82, 78, 77, 83, 81), and FE2_3 (10, 12, 38, 41, 73, 75, 79). These groups are depicted in Figures 2 and 3. Biasing potentials were iteratively determined for each FE1 and FE2 subgroup with adaptive landscape flattening [22] over a total of 30–35 ns of sampling. Next, 100 ps of equilibration was performed. Finally, 5 independent production runs of 20 ns each were performed using different random velocity seeds, for a total of 100 ns of production sampling. Standard deviations were calculated over the 5 duplicate production simulations for each binding free energy calculation. Relative free energies were computed with respect to reference compounds 17 and 10 for FE1 and FE2, respectively.

# 3 Results

## 3.1 Docking Results

In the challenge, the FXR apo crystal structure was provided, but 26 other holo structures of FXR were found in the PDB database with different ligands bound, including benzimidazoles. After clustering the 26 holo structures using Bio3D[38], two representative structures were chosen for use in our docking protocol: 3OKI [37] and 3DCT [3]. Because the bound ligand in 3OKI is a benzimidazole compound, this structure was exclusively used for docking the benzimidazole ligands of the challenge. All other ligands were docked to both 3OKI and 3DCT structures. We used an RMSD cutoff of 2 Å to decide if the docked pose was correct.

The docking results are shown in Fig. 4. For all the benzimidazole ligands except ligand 13, the top-1 ranked docking poses have a RMSD less than 2 Å with respect to the native poses. For all the other ligands, except ligand 5, the top-1 ranked docking poses all had an RMSD greater than 2 Å. Errors in most of the miscellaneous ligands were high among all participants, due to their significantly varied binding poses from other compounds, so it is not surprising that CDOCKER failed to identify these poses. Retrospectively, it is also not surprising that the benzimidazole ligands were docked correctly, because the benzimidazole holo structure of FXR was available in the PDB database and was used. To test whether the correct pose for other ligands could be identified with the correct FXR holo structure corresponding to each ligand, we re-docked all 35 ligands into their corresponding D3R supplied native holo structures of FXR. The results are shown in Fig. 5 (A), where all 35 ligands except ligand 18 were re-docked correctly and the average RMSD for the top-1 re-docked ligands is 1.18 Å. Both the prospective and retrospective re-docking results show that CDOCKER is quite reliable in re-docking, but emphasize the importance of choosing a reasonable receptor structure model for docking if flexible receptor docking is not employed[18].

The reliability of CDOCKER on re-docking suggests that one possible reason for the poor results when docking non-benzimidazole ligands is that the FXR structures used for docking were too different from the crystallographic holo structures. Fig. 6 shows the RMSD between FXR apo and holo structures. The holo structures of the benzimidazole ligands are all similar to each other but quite different from the holo structures of all the other ligands, except ligand 5, as well as the FXR apo structure. Compared with the 3OKI structure, the 3DCT structure is more closely related to the holo structures of the non-benzimidazole ligands. In our original docking protocol, we docked all non-benzimidazole ligands with both 3OKI and 3DCT. Our submitted 5 poses were the lowest 5 energy poses among all poses docked with both 3OKI and 3DCT structures. One potential artifact of this approach is that for non-benzimidazole ligands the poses docked with 3OKI could be incorrectly favored over those docked to 3DCT because of a lower calculated energy. For example, for 10, the lowest energy docked pose was from the 3OKI structure, which has a RMSD of 6.11 Å; however, the other 4 submitted poses were poses docked with 3DCT and of these 4 poses, the lowest RMSD was only 2.75 Å. Similarly, all 5 submitted poses for 15 and 4 out of 5 submitted poses for 17 were poses docked with 3OKI, but for which the RMSDs were greater than 4.0 Å (Fig. 4 (B)).

Compared with both 3OKI and 3DCT structures, the apo structure agrees more closely to the holo structures of non-benzimidazole ligands. Therefore the docking protocol was run again for all ligands with the FXR apo structure. The results are shown in Fig. 5 (B). Compared with the submitted docking results using the 3OKI and 3DCT structures shown in Fig. 4, the results using the apo structure became worse for benzimidazole ligands, as expected, but much better for non-benzimidazole ligands, especially the sulfonamide and spiro ligands. This makes sense because the benzimidazole holo structure is quite different from the apo structure, and the correct pose can not be found using a rigid receptor docking protocol. For non-benzimidazole ligands, using the top-1 ranked pose, 2 out of 3 sulfonamide ligands were docked correctly and all 3 spiro ligands were docked correctly.

### 3.2 FE1 Results

Free energy set 1 consisted of 15 ligands with multiple sites of chemical substitutions, including a ligand core change from a piperidinyl fused pyrazole bicyclic ring to a piperidinyl fused pyrrole. These perturbations were investigated by breaking up the set into 4 subgroups, as described earlier. Upon completion of the initial NVT calculations, two problems immediately stood out: first the equilibration procedure for the simulations did not equilibrate the pressure, and so the protein and solvent simulations ran at roughly −1150 and −600 atm, respectively. (Similar differences were observed in both FE1 and FE2.) Second, in FE1_1, ligand 101 was negatively charged, and this led to an obvious artifact, as 101 had a predicted relative binding free energy of 11.6 kcal/mol, relative to the reference compound 17, while other ligands were within a few kcal/mol. This is unsurprising, as it is well known that charge changes can lead to large errors in free energy calculations [40,24,21].

To address these issues, additional simulations with pressure coupling were run; unfortunately, these simulations did not finish in time for the D3R submission deadline. Furthermore, a modified FE1_1 system with a protonated (neutral) 101 ligand was run in the

NPT and NVT ensembles. In the neutralized FE1_1 system, only the NPT simulation finished in time, therefore NVT results were used for all ligands except 101. For 101, the neutral NPT result was used by finding the relative binding free energy of 101 relative to the mean of the other six ligands, and adding that to the mean of the other six ligands in the old charged NVT result. Uncertainties were obtained with bootstrap analysis by sampling from the 5 independent runs. Correlation with experiment is shown in Fig. 7a. The mean unsigned error (MUE) and root mean square error (RMSE) relative to ligand 17 was 1.77 and 2.08 kcal/mol, respectively (Table 2). The root mean square error with respect to the mean, or center, (RMSEc) was 2.00 kcal/mol. The Pearson correlation coefficient of 0.04 was poor. In most instances, large errors were obtained because ligands were predicted to bind more favorably than experimentally observed. For example, ligands 96 and 102, were predicted to be more favorable than reference compound 17 by ~0.5–1.5 kcal/mol. The three inactive compounds, 48, 49, and 99, were also predicted as false-positives. In contrast, ligands 91 and 101 were predicted to be less potent than compound 17 by ~1.0–2.0 kcal/mol more than experimentally observed (Table 2).

Biasing potential replica exchange (BP-REX) has been shown to substantially improve convergence of MS$\lambda$D results in previous studies [4], but was previously incompatible with the domdec module of CHARMM. As mentioned in the methods section, a domdec-compatible version of BP-REX was written, but was not ready in time for the submission deadline. Consequently, BP-REX was applied retrospectively to determine whether this enhancement would improve the computed free energies. BP-REX was applied with the NVT ensemble to neutral FE1_1 and FE1_2, and with the NPT ensemble to all four sets. Typically dihedrals are not scaled by $\lambda$ and NOE restraints are used to ensure overlap of analogous atoms in the ligands in order to improve sampling. These features were removed in the BP-REX simulations of FE1 out of concern they might introduce artifacts, though additional simulations suggested they had little effect.

BP-REX results are shown in Fig. 7b–c and Table 2. Results from the NVT ensemble simulations yielded an MUE, RMSE, and RMSEc of 1.64, 1.99, and 1.34 kcal/mol, respectively; the Pearson correlation coefficient was 0.01. Findings from the NPT ensemble simulations showed larger errors (MUE of 2.49, RMSE of 2.84, RMSEc of 1.37 kcal/mol, and a Pearson correlation coefficient of −0.16). In this case, the difference between RMSE depending on reference arises because of larger errors in the reference compound 17. This is especially acute in the NPT ensemble because the reference ligand and a few others with larger substituents became less favorable than in the NVT ensemble. This makes sense as the NVT results included both a component for binding, and an artifact due to transfer of the ligand from a higher pressure (solvent at −600 atm) to a lower pressure (protein at −1150 atm) system, which is favorable for larger ligands. Consequently NPT removed this artifact which had been making results appear artificially good.

### 3.3 FE2 Results

Free Energy Set 2 consisted of 18 ligands with two sites of chemical alterations. These ligand sets were divided into three subgroups (Fig. 3) to calculate relative free energies of binding with MS$\lambda$D. The first two groups, FE2_1 and FE2_2, maintained a constant R2

group of 4-COO-phenyl with an ionic charge of -1 and featured changes to the aromatic ring attached to the sulfonamide moiety of the ligand core (the R1 site). The third group, FE2_3, investigated changes at both R1 and R2 sites. Although many of the perturbations involved the addition or transformation of only a small handful of heavy atoms, usually considered ideal for pairwise FEP or TI calculations, the connectivity of the various groups made setting up MS$\lambda$D calculations more difficult. For example, 78, 77, and 83 all feature an R1 phenyl ring with two chlorine atoms attached (Fig. 3); however, the 2,6-, 2,3-, and 2,5-dichloro patterns of substitution meant that the entire aromatic ring for each ligand would have to be considered as a separate substituent with MS$\lambda$D. This was further reinforced by the need to sample between benzene and thiophene rings at R1. Similar to problems observed in the FE1 ligand set, ligand charge changes provided another complication. While the majority of the FE2 ligands contained a carboxylic acid group at R2, which was originally modeled in its anionic form, four ligands featured neutral R2 groups, thus requiring an ionic to neutral perturbation. Original attempts to model transitions between neutral and ionic substituents concurrently in FE2_3 resulted in extremely poor $\lambda$ sampling, significant end-point trapping, and a roughly 8 kcal/mol divide between neutral and charged ligands. Long range electrostatic corrections, such as the PME correction [14,16], were not incorporated into the present simulations, so it is possible that these difficulties were a result of how electrostatic interactions were represented. The difficulties encountered were alleviated by protonating ligands 10 and 12 in FE2_3 to make a separate, neutral subgrouping of ligands. No thermodynamic correction was performed to connect FE2_3 with the results from FE2_1 and FE2_2, and thus FE2_3 results are treated separately in the discussion below.

The FE2 results submitted to Grand Challenge 2 are graphically shown in Fig. 7d and reported in Table 3 (Submitted NVT). For all ligand perturbations in FE2, the MUE, RMSE, and RMSEc were 1.26, 1.67, and 1.62 kcal/mol, respectively, scoring 12th in RMSE accuracy among all participants. The Pearson correlation coefficient was 0.45. Standard deviations (SD) between independent runs were generally 0.8–0.9 kcal/mol, with one as high as 1.37 kcal/mol for ligand 78. The average SD is less than 0.20 kcal/mol for the solvated ligand, so most of the uncertainty originates from the protein simulations. Of the 18 relative free energies, only 6 predictions had errors greater than 2.0 kcal/mol, two of which were the inactive compounds 41 and 75. Unfortunately, 41 was predicted to be a false-positive with a computed free energy of binding of −2.16 kcal/mol, compared to the experimental lower limit of 1.70 kcal/mol. This may have occurred because the ester moiety in 41 and the carboxylic acid in 12 maintain a similar hydrogen bond to the neighboring Asn297 residue, yielding similar poses throughout most of the MD trajectories, and the carboxylic acid fails to form other strong interactions in its protonated state to differentiate its binding ability. A false positive was also observed with 38, although to a lesser extent. 75 shows the correct inactive result, though it is difficult to gauge its true accuracy when the reported $IC_{50}$ is a lower limit (>100 $\mu$M). Other molecules in the set seem to be well represented, including 84, 89, 82, 77, and 73, with unsigned errors near or below 0.5 kcal/mol. Most of these are found in FE2_1 and FE2_2, leading to the reduced MUE/RMSE errors for this subgrouping compared to FE2_3. Interestingly, ligand 12, present in both FE2_1 and FE2_3, is predicted to have $\Delta\Delta G_{bind}$ of −0.36 and −1.95 kcal/mol, respectively. Two possible sources of error for this discrepancy include incomplete sampling or

protonation state differences. Although both sources of error likely contribute, neither was able to be fully explored prior to the deadlines of the challenge. Retrospectively, the sampling issue was investigated with BP-REX enhanced sampling.

Upon completion of a domdec-compatible version of BP-REX, the above FE2 simulations were repeated in the NVT ensemble with replica exchange (BP-REX NVT in Table 3, Fig. 7e). Only three independent calculations for each subgroup were performed, and the standard deviations (SDs) averaged 0.4 kcal/mol, with a maximum SD of 0.81 kcal/mol. For FE2, slight improvements in accuracy were observed by employing the replica exchange algorithm, decreasing the MUE and RMSE to 1.11 and 1.54 kcal/mol. Some notable improvements are observed, for example 75 and 79 are closer to experiment with errors near 1.0 kcal/mol, compared to the 2.0–2.2 kcal/mol errors observed previously, but the majority of compounds maintained similar errors to those observed in the Submitted NVT set. Notably, however, the precision of the results is much improved. The reduction in standard deviation is accompanied by a slightly better correlation of 0.48 (Fig. 7e). Some outliers still occur, such as 78, 81, and 41, but most of the other data points nicely fall within ± 1.0 kcal/mol of the ideal = $x$ curve. Furthermore, single site transition rates increased from an average of 60 transitions per ns without replica exchange to 160 transitions per ns with BP-REX. Thus the replica exchange algorithm successfully improves $\lambda$ sampling and yields more precise free energy results. It is clear that for modeling large substituent changes with MS$\lambda$D, such as was performed for both FE1 and FE2, the BP-REX algorithm should be employed to obtain the most precise results [4].

Owing to the improvements in FE1 from running MS$\lambda$D simulations in the NPT ensemble, BP-REX simulations for FE2 were also repeated, but only a single calculation was performed for each subgroup (BP-REX NPT in Table 3, Fig. 7f). Because the substituted phenyl R1 and R2 rings all occupied about the same space at each respective site, the benefits of alleviating pressure artifacts observed with FE1 were largely absent here. And indeed, errors were observed to increase in moving from the NVT ensemble to the NPT ensemble. It is not immediately clear why poorer results are obtained and future investigations are pending to learn more from this system. A preliminary investigation of convergence for the BP-REX NPT simulations revealed poorer $\lambda$ sampling for substituents from 75 and 78, which also show the largest $\Delta G_{\text{bind}}$ errors of 4.6 and 5.5 kcal/mol. Therefore additional sampling or further optimization of biasing potentials could reduce noise in these ligands and improve correlation with experiment. Removing them from the analysis reduces the MUE to 1.49 kcal/mol.

## 4 Discussion

### 4.1 Docking

Rigid receptor docking with CDOCKER successfully identified the correct ligand bound poses in re-docking experiments where the correct holo structure of the protein was used. In contrast, it is also clear that large protein conformational changes induced by ligand binding significantly increased docking difficulty and limited predictive accuracies. In addition, side chain placement is also important when the receptor structure used for docking has multiple alternative conformations, which occurs when part of the receptor has partial occupancy. For

example, some spiros and sulfonamides, such as 10 or 15, show steric clashes with one of two alternative positions of His298 and Asn297 in the FXR apo structure, which prevents the rigid receptor docking from identifying the correct docking pose if the wrong alternative conformation of the receptor is used.

Retrospectively, we also tried adding side chain flexibility with flexible CDOCKER into the current docking protocol (the detailed results are not shown). We observed mixed results when the simulated annealing search remained the same as the rigid receptor docking procedure. The results from flexible side chain docking do look promising for several ligands, as it can identify the correct docking poses within the top 5 poses, which is impossible for rigid protein docking to identify, as there is collision between the correct poses and the apo protein structure. Adding protein side chain flexibility greatly increases the search space. More exhaustive conformational search will be needed in order to further improve the flexible protein side chain docking results. To accelerate the conformational search, in ongoing work we are parallelizing the search by moving the flexible protein side chain search on to GPUs.

### 4.2 Binding Free Energy Prediction

Ligand perturbations of the size present in the D3R challenge have been previously shown to require soft-core potentials to obtain consistent results and adaptive landscape flattening to sample effectively [22]. Consequently these corrections were used throughout this study. In addition, porting MS$\lambda$D calculations to GPUs provided speed-ups of a factor of 5–10 compared to prior CPU protocols, allowing more simulations to be run on a local GPU cluster within the timeframe of Grand Challenge 2 and during retrospective analysis. After submitting predictions for FE1 and FE2, two additional corrections were applied to improve accuracy. First, BP-REX was implemented with domdec to enable GPU acceleration of BP-REX simulations. BP-REX resulted in improved statistical precision and lower RMSE, suggesting simulations converged to force field values more efficiently. Second, simulations were run in the NPT ensemble rather than the NVT ensemble. Unfortunately, although this removed a source of error due to differing pressures in the solvent and protein-bound simulations, the results appear to be slightly poorer when the NPT ensemble is used. In FE1, these larger errors can be attributed primarily to increased errors in the reference ligand. In FE2, the increased errors can be partly attributed the fact that only a single 20 ns run was performed in the NPT ensemble, giving less converged estimates. In addition, sampling and convergence could be slower in the NPT ensemble, similar to what was observed with ligands 75 or 78 in FE2.

Ligand charge changes remain an unresolved difficulty encountered in Grand Challenge 2. It is undoubtedly necessary to account for long range electrostatic effects to obtain accurate results in ligand perturbations between ligands with different net charges [24,46]. Charge changes especially plagued FE2, where ligands were treated as negative in FE2_1 and FE2_2, but had to be neutralized in FE2_3 to connect with other neutral ligands. There is no reason to expect that free energy estimates will be the same between protonated and deprotonated ligands. Methods to account for long range electrostatic effects, such as particle mesh Ewald [14, 16], have been recently developed for Constant pH MD [23], a

special class of $\lambda$ dynamics, and needs to be fully incorporated into future MS$\lambda$D capabilities. Other potential sources of error beyond the reach of MS$\lambda$D are force field errors and discrepancies between the measured $IC_{50}$s and the true dissociation constants.

In conclusion, in the current work we show that the CHARMM-based docking and free energy approaches, CDOCKER and MS$\lambda$D, perform quite reasonably in pose prediction and free energy estimation. Furthermore, this D3R exercise allowed us to push ongoing developments to refine the algorithmic approaches we are using in this domain. However, we also note a significant weakness in the current structure of this and related exercises. While blind challenges are useful to refine and test various computational approaches for protein-ligand docking and free energy of protein-ligand binding methodologies, they do not enable participants to access the underlying models and force fields used by others. In the context of our CHARMM-based approaches, we are, for example, agnostic to the force fields utilized in the simulations. Thus, comparison of the core sampling methodologies could be directly assessed relative to the force fields used by other participants, permitting a clear assessment of whether observed differences arise from algorithms and sampling or force fields differences. We, therefore, strongly urge the organizers of D3R, and related blinded assessments, to move to a model whereby participants are able to share their force field models, thereby enabling truly “apples-to-apples” comparison between protocols and methods.

All topology and system set up files used for this challenge are available upon request.

## Acknowledgments

We thank Michael Crowley for sharing a new implementation of replica-exchange in CHARMM upon which we based BP-REX with domdec.

This work is supported by grants from the NIH (GM037554 and GM107233).

## References

1. Abel R, Mondal S, Masse C, Greenwood J, Harriman G, Ashwell MA, Bhat S, Wester R, Frye L, Kapeller R, Friesner RA. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. Current Opinion in Structural Biology. 2017; 43:38–44. [PubMed: 27816785]
2. Adcock SA, McCammon JA. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chemical Reviews. 2006; 106:1589–1615. [PubMed: 16683746]
3. Akwabi-Ameyaw A, Bass JY, Caldwell RD, Caravella JA, Chen L, Creech KL, Deaton DN, Jones SA, Kaldor I, Liu Y, Madauss KP, Marr HB, McFadyen RB, Miller AB, Navas F III, Parks DJ, Spearing PK, Todd D, Williams SP, Wisely GB. Conformationally constrained farnesoid X receptor (FXR) agonists: Naphthoic acid-based analogs of GW 4064. Bioorganic and Medicinal Chemistry Letters. 2008; 18:4339–4343. [PubMed: 18621523]
4. Armacost KA, Goh GB, Brooks CL III. Biasing potential replica exchange multisite $\lambda$-dynamics for efficient free energy calculations. Journal of Chemical Theory and Computation. 2015; 11:1267–1277. [PubMed: 26579773]
5. Baig MH, Ahmad K, Roy S, Ashraf JM, Adil M, Siddiqui MH, Khan S, Kamal MA, Provazník I, Choi I. Computer aided drug design: Success and limitations. Current Pharmaceutical Design. 2016; 22:572–581. [PubMed: 26601966]
6. Ban F, Dalal K, Li H, LeBlanc E, Rennie PS, Cherkasov A. Best practices of computer-aided drug discovery: Lessons learned from the development of a preclinical candidate for prostate cancer with

a new mechanism of action. Journal of Chemical Information and Modeling. 2017; 57:1018–1028. [PubMed: 28441481]

7. Best RB, Mittal J, Feig M, MacKerell AD Jr. Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of $\alpha$-helix and $\beta$-hairpin formation. Biophysical Journal. 2012; 103:1045–1051. [PubMed: 23009854]
8. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone $\varphi$, $\psi$ and side-chain $\chi_1$ and $\chi_2$ dihedral angles. Journal of Chemical Theory and Computation. 2012; 8:3257–3273. [PubMed: 23341755]
9. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. Journal of Computational Chemistry. 2009; 30:1545–1614. [PubMed: 19444816]
10. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry. 1983; 4:187–217.
11. Carlson HA, Smith RD, L K, Stuckey JA, Ahmed A, Convery MA, Somers DO, Kranz M, Elkins PA, Cui G, Peishoff CE, Lambert MH, Dunbar JB Jr. CSAR 2014: A benchmark exercise using unpublished data from pharma. Journal of Chemical Information and Modeling. 2016; 56:1063–1077. [PubMed: 27149958]
12. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. Molprobity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D: Biological Crystallography. 2010; 66(1):12–21. [PubMed: 20057044]
13. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: progress and challenges. Current Opinion in Structural Biology. 2011; 21:150–160. [PubMed: 21349700]
14. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. Journal of Chemical Physics. 1993; 98:10,089–10,092.
15. Ding X, Vilseck JZ, Hayes RL, Brooks CL. Gibbs sampler based $\lambda$-dynamics and rao-blackwell estimator for alchemical free energy calculation. Journal of Chemical Theory and Computation. 2017; 13:2501–2510. [PubMed: 28510433]
16. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh ewald method. Journal of Chemical Physics. 1995; 103:8577–8593.
17. Feng S, Yang M, Zhang Z, Wang Z, Hong D, Richter H, Benson GM, Bleicher K, Grether U, Martin RE, Plancher JM, Kuhn B, Rudolph MG, Chen L. Identification of an N-oxide pyridine GW4064 analog as a potent FXR agonist. Bioorganic and Medicinal Chemistry Letters. 2009; 19:2595–2598. [PubMed: 19328688]
18. Gagnon JK, Law SM, Brooks CL. Flexible cdocker: Development and application of a pseudo-explicit structure-based docking method within charmm. Journal of Computational Chemistry. 2016; 37:753–762. [PubMed: 26691274]
19. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB Jr, Carlson HA, Burley SK, Walters WP, Amaro RE, Feher VA, Gilson MK. D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity pre dictions. Journal of Computer-Aided Molecular Design. 2016; 30:651–668. [PubMed: 27696240]
20. Frisch, MJ., Trucks, GW., Schlegel, HB., Scuseria, GE., Robb, MA., Cheeseman, JR., Scalmani, G., Barone, V., Mennucci, B., Petersson, GA., Nakatsuji, H., Caricato, M., Li, X., Hratchian, HP., Izmaylov, AF., Bloino, J., Zheng, G., Sonnenberg, JL., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, JA., Jr, Peralta, JE., Ogliaro, F., Bearpark, M., Heyd, JJ., Brothers, E., Kudin, KN., Staroverov, VN., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, JC., Iyengar, SS., Tomasi, J., Cossi, M., Rega, N., Millam, JM., Klene, M., Knox, JE., Cross, JB., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, RE., Yazyev, O., Austin, AJ., Cammi, R.,

Pomelli, C., Ochterski, JW., Martin, RL., Morokuma, K., Zakrzewski, VG., Voth, GA., Salvador, P., Dannenberg, JJ., Dapprich, S., Daniels, AD., Farkas, O., Foresman, JB., Ortiz, JV., Cioslowski, J., Fox, DJ. Gaussian 09, Revision A.02. Gaussian, Inc; Wallingford CT: 2009.

21. Hansen N, van Gunsteren WF. Practical aspects of free-energy calculations: A review. Journal of Chemical Theory and Computation. 2014; 10:2632–2647. [PubMed: 26586503]
22. Hayes RL, Armacost KA, Vilseck JZ, Brooks CL III. Adaptive landscape flattening accelerates sampling of alchemical space in multisite $\lambda$ dynamics. Journal of Physical Chemistry B. 2017; 121:3626–3635.
23. Huang Y, Chen W, Wallace JA, Shen J. All-atom continuous constant pH molecular dynamics with particle mesh Ewald and titratable water. Journal of Chemical Theory and Computation. 2016; 12:5411–5421. [PubMed: 27709966]
24. Hummer G, Pratt LR, García AE. Free energy of ionic hydration. Journal of Physical Chemistry. 1996; 100:1206–1215.
25. Hynninen AP, Crowley MF. New faster CHARMM molecular dynamics engine. Journal of Computational Chemistry. 2014; 35:406–413. [PubMed: 24302199]
26. Jorgensen WL. Efficient drug lead discovery and optimization. Accounts of Chemical Research. 2009; 42:724–733. [PubMed: 19317443]
27. Jorgensen WL. Computer-aided discovery of anti-HIV agents. Bioorganic and Medicinal Chemistry. 2016; 24:4768–4778. [PubMed: 27485603]
28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics. 1983; 79:926–935.
29. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. Nature Reviews Drug Discovery. 2004; 3:935–949. [PubMed: 15520816]
30. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. Drug Discovery Today. 2006; 11:580–594. [PubMed: 16793526]
31. Knight JL, Brooks CL III. Applying efficient implicit nongeometric constraints in alchemical free energy simulations. Journal of Computational Chemistry. 2011; 32:3423–3432. [PubMed: 21919014]
32. Knight JL, Brooks CL III. Multisite $\lambda$ dynamics for simulated structure-activity relationship studies. Journal of Chemical Theory and Computation. 2011; 7:2728–2739. [PubMed: 22125476]
33. Kong X, Brooks CL III. $\lambda$-dynamics: A new approach to free energy calculations. Journal of Chemical Physics. 1996; 105:2414–2423.
34. Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. Current Medicinal Chemistry. 2013; 20:2839–2860. [PubMed: 23651302]
35. MarvinSketch 16.6.6.0. ChemAxon; 2016. URL www.chemaxon.com
36. Richter HG, Benson G, Bleicher K, Blum D, Chaput E, Clemann N, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin R, Plancher JM, Rudolph M, Schuler F, Taylor S. Optimization of a novel class of benzimidazole-based farnesoid X receptor (FXR) agonists to improve physicochemical and ADME properties. Bioorganic and Medicinal Chemistry Letters. 2011; 21:1134–1140. [PubMed: 21269824]
37. Richter HG, Benson GM, Blum D, Chaput E, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin RE, Plancher JM, Rudolph MG, Schuler F, Taylor S, Bleicher KH. Discovery of novel and orally active FXR agonists for the potential treatment of dyslipidemia and diabetes. Bioorganic and Medicinal Chemistry Letters. 2011; 21:191–194. [PubMed: 21134747]
38. Skjærven L, Jariwala S, Yao XQ, Grant BJ. Online interactive analysis of protein structure ensembles with bio3d-web. Bioinformatics. 2016; 32:3510–3512. [PubMed: 27423893]
39. Stewart JJP. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. Journal of Molecular Modeling. 2007; 13:1173–1213. [PubMed: 17828561]
40. Straatsma TP, Berendsen HJC. Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. Journal of Chemical Physics. 1988; 89:5876–5886.

41. Trott O, Olson AJ. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry. 2010; 31:455–461. [PubMed: 19499576]
42. van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. Molecular Physics. 1977; 34:1311–1327.
43. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry. 2010; 31:671–690. [PubMed: 19575467]
44. Vanommeslaeghe K, MacKerell AD Jr. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. Journal of Chemical Information and Modeling. 2012; 52:3144–3154. [PubMed: 23146088]
45. Vanommeslaeghe K, Raman EP, MacKerell AD Jr. Automation of the CHARMM general force field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. Journal of Chemical Information and Modeling. 2012; 52:3155–3168. [PubMed: 23145473]
46. Wallace JA, Shen JK. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. Journal of Chemical Physics. 2012; 137:184, 105.
47. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Frye L, Farid R, Lin T, Mobley DL, Jorgensen WL, Berne BJ, Friesner RA, 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. Journal of the American Chemical Society. 2015; 137:2695–2703. [PubMed: 25625324]
48. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. Journal of molecular biology. 1999; 285(4):1735–1747. [PubMed: 9917408]
49. Wu G, Robertson DH, Brooks CL, Vieth M. Detailed analysis of grid-based molecular docking: A case study of cdockera charmm-based md docking algorithm. Journal of Computational Chemistry. 2003; 24:1549–1562. [PubMed: 12925999]
50. Yesselman JD, Price DJ, Knight JL, Brooks CL III. Match: An atom-typing toolset for molecular mechanics force fields. Journal of Computational Chemistry. 2011; 33:189–202. [PubMed: 22042689]
51. Yin J, Henriksen NM, Slochower DR, Shirts MR, Chiu MW, Mobley DL, Gilson MK. Overview of the SAMPL5 host-guest challenge: Are we doing better. Journal of Computer-Aided Molecular Design. 2017; 31:1–19. [PubMed: 27658802]
52. Zwanzig RW. High-temperature equation of state by a perturbation method. i. nonpolar gases. Journal of Chemical Physics. 1954; 22:1420–1426.

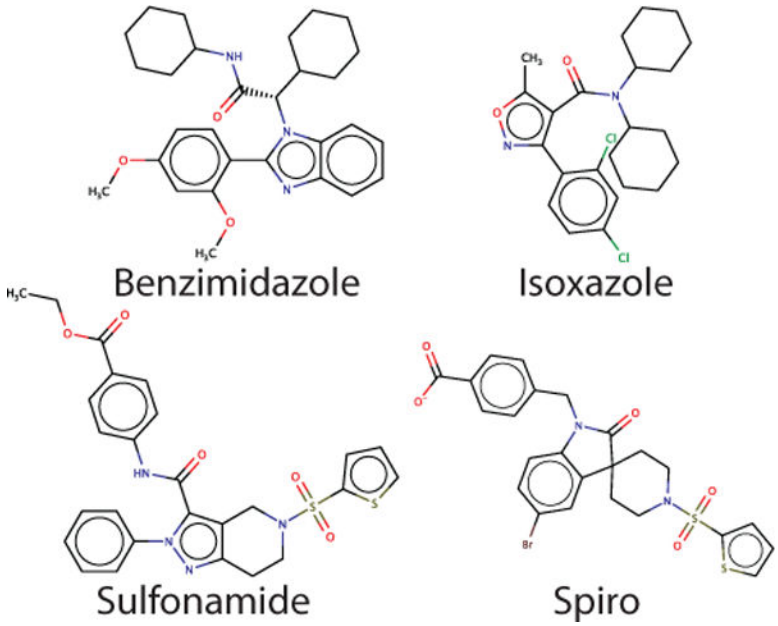
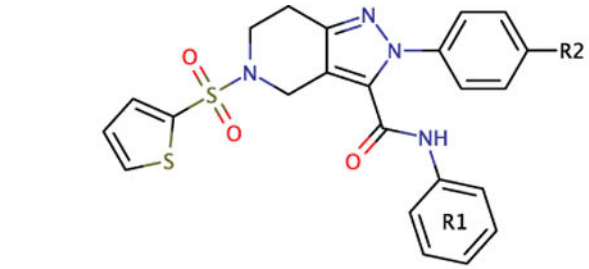



**Fig. 1.**

Representative ligands from each class: benzimidazole (6), isoxazole (4), sulfonamide (17), and spiro (10).

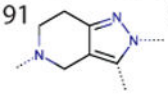



| FE1_1 | $R_1$ |
|---|---|
| 17 | 4-$COOCH_2CH_3$ |
| 49 | 4-$COCH_3$ |
| 46 | 4-$CONH_2$ |
| 101 | 4-COOH |
| 98 | 4-$CONHCH_3$ |
| 96 | 4-$CON(CH_3)_2$ |
| 102 | 4-(CO-morpholino) |
| ($R_2$ = H) | |

| FE1_2 | $R_1$ |
|---|---|
| 17 | 4-$COOCH_2CH_3$ |
| 91 | H |
| 47 | 3-$COOCH_2CH_3$ |
| 99 | 4-$OCH_3$ |
| 95 | 4-$NHCOCH_3$ |
| 48 | 4-$CH_2COOCH_2CH_3$ |
| 100 | 4-$SO_2NH_2$ |
| ($R_2$ = H) | |

| FE1_3 | $R_2$ |
|---|---|
| 17 | H |
| 45 | $OCF_3$ |
| ($R_1$ = 4-$COOCH_2CH_3$) | |

FE1_4 core swap

91  ($R_1$ = H)

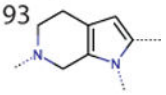
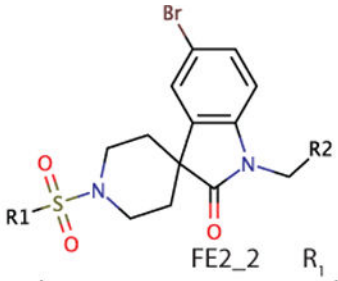
93  ($R_2$ = H)

**Fig. 2.**

The partitioning of ligands in free energy set 1 into subgroups for MS$\lambda$D calculations.




| FE2_1 | $R_1$ | FE2_2 | $R_1$ |
|---|---|---|---|
| 10 | 2-thienyl | 10 | 2-thienyl |
| 12 | 2-Cl-phenyl | 89 | 4-Cl-phenyl |
| 76 | phenyl | 82 | 2-F-3-Cl-phenyl |
| 84 | 2-F-phenyl | 78 | 2-Cl-6-Cl-phenyl |
| 74 | 2-Br-phenyl | 77 | 2-Cl-3-Cl-phenyl |
| 85 | 2-$CH_3$-phenyl | 83 | 2-Cl-5-Cl-phenyl |
| 88 | 2-$CF_3$-phenyl | 81 | 2-$CH_3$-3-Cl-phenyl |
| ($R_2$ = 4-COOH-phenyl) | | ($R_2$ = 4-COOH-phenyl) | |

| FE2_3 | $R_1$ = 2-thienyl | $R_1$ = 2-Cl-phenyl |
|---|---|---|
| $R_2$ = 4-COOH-phenyl | 10 | 12 |
| $R_2$ = 4-$COOCH_3$-phenyl | 38 | 41 |
| $R_2$ = 4-OH-phenyl | 73 | -- |
| $R_2$ = 4-pyridyl | 75 | -- |
| $R_2$ = 3-COOH-phenyl | 79 | -- |

**Fig. 3.**

The partitioning of ligands in free energy set 2 into subgroups for MS$\lambda$D calculations.

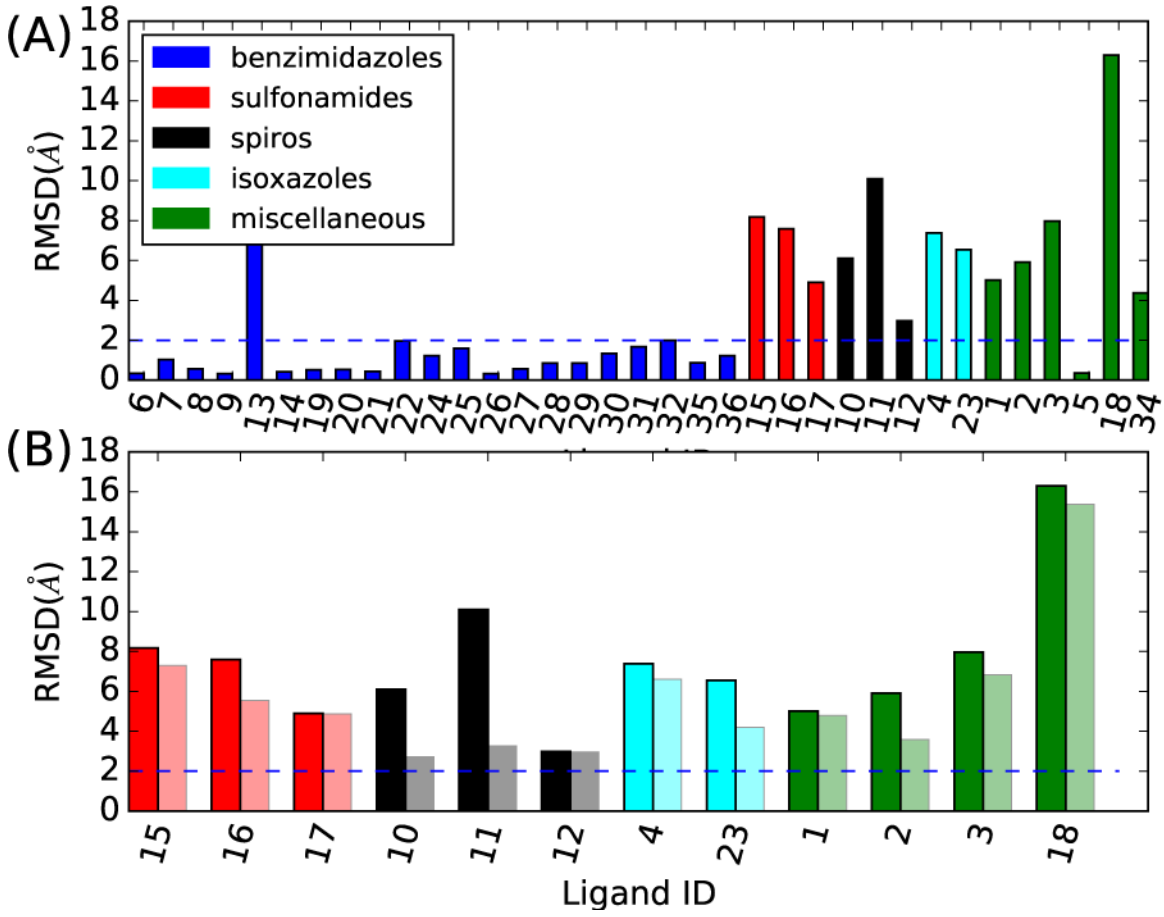



**Fig. 4.**

RMSD of docked poses with respect to native poses. (A) RMSD of the lowest energy poses for all 35 ligands sorted based on ligand types; (B) (opaque) RMSD of the lowest energy poses for non-benzimidazole ligands; (transparent) Lowest RMSD among 5 lowest energy poses for the corresponding ligands. The blue dash line in both (A) and (B) corresponds to an RMSD of 2 Å

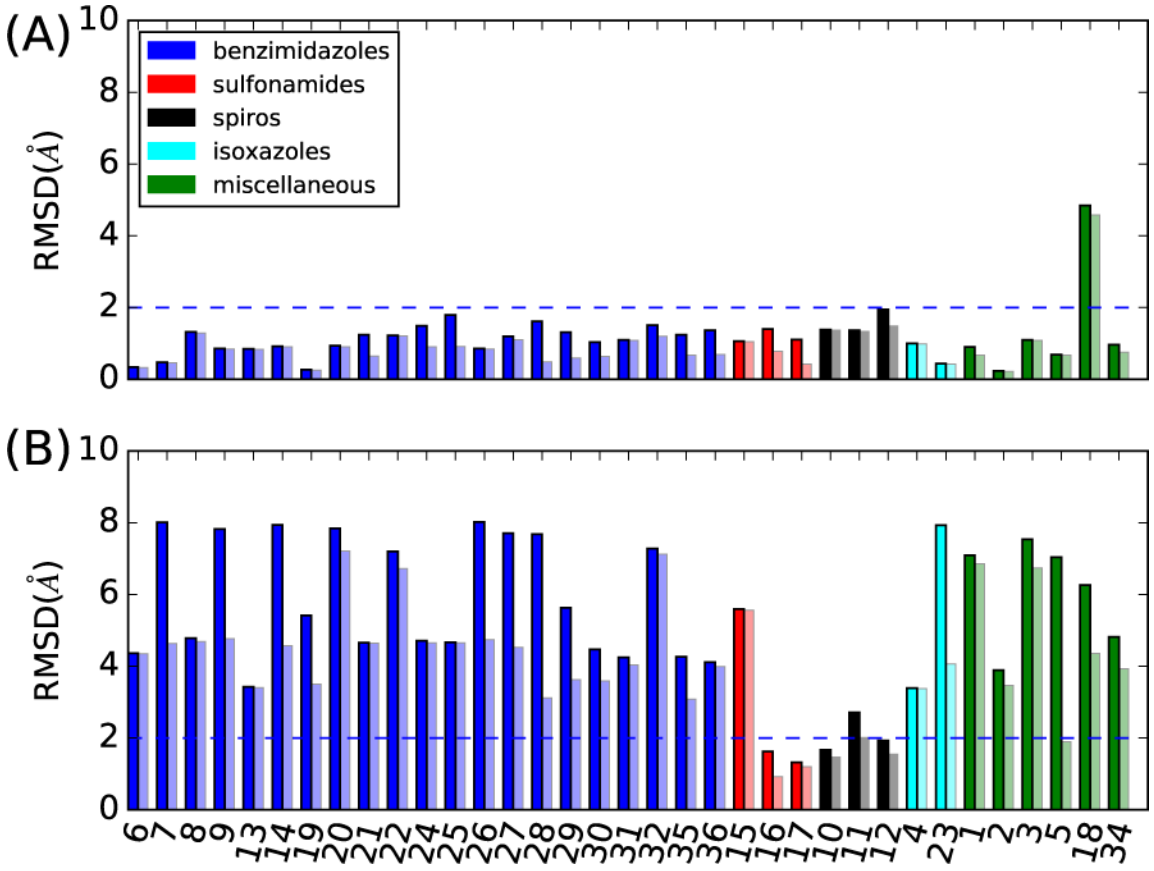



**Fig. 5.**

Root-mean-square deviation (RMSD) of docked poses with respect to native poses. (A) Results of re-docking, in which the corresponding holo structure of FXR was used. (B) Results of docking using the apo structure of FXR. (opaque) RMSD of the lowest energy poses; (transparent) Lowest RMSD among 5 lowest energy. The blue dash line corresponds to the RMSD of 2 Å.

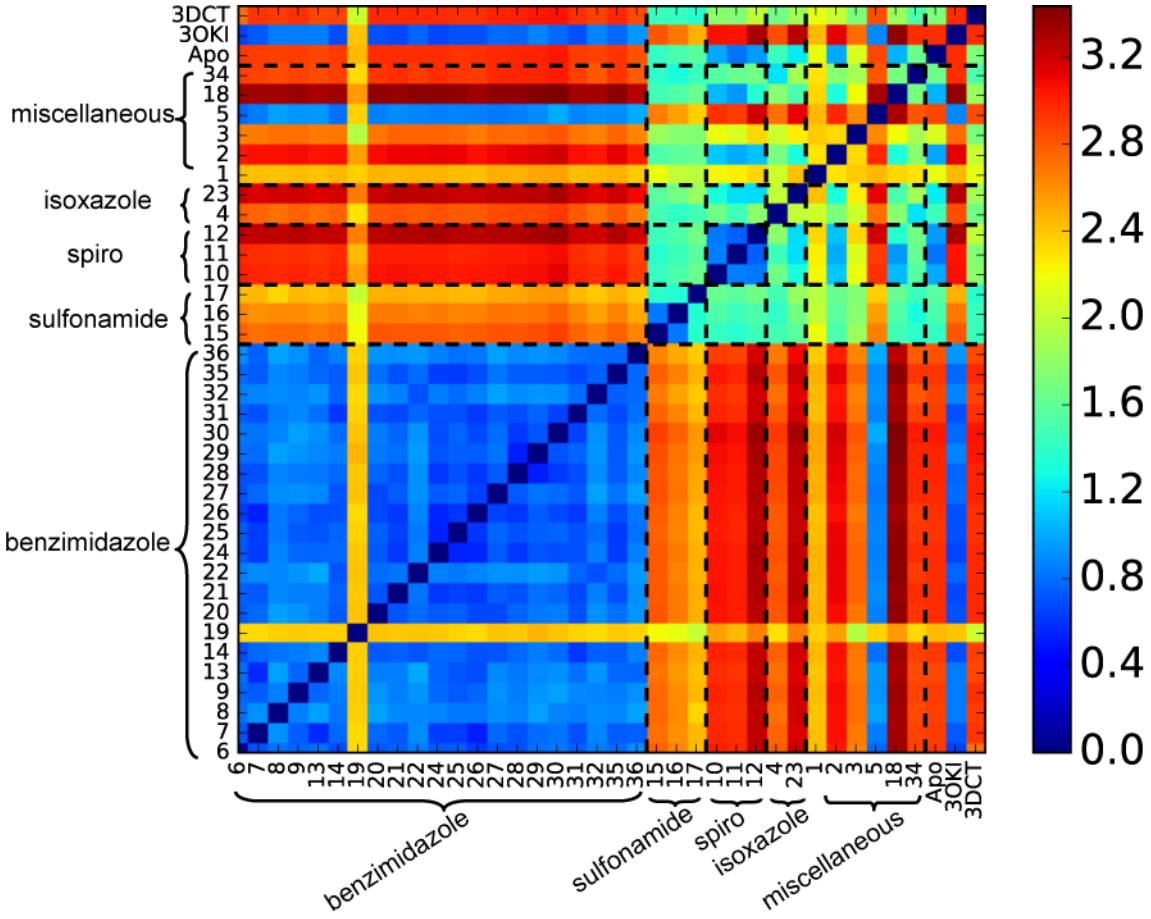



**Fig. 6.**

Root-mean-square deviation (RMSD) between protein structures of FXR. The black dashed lines are used to separate different ligand types.

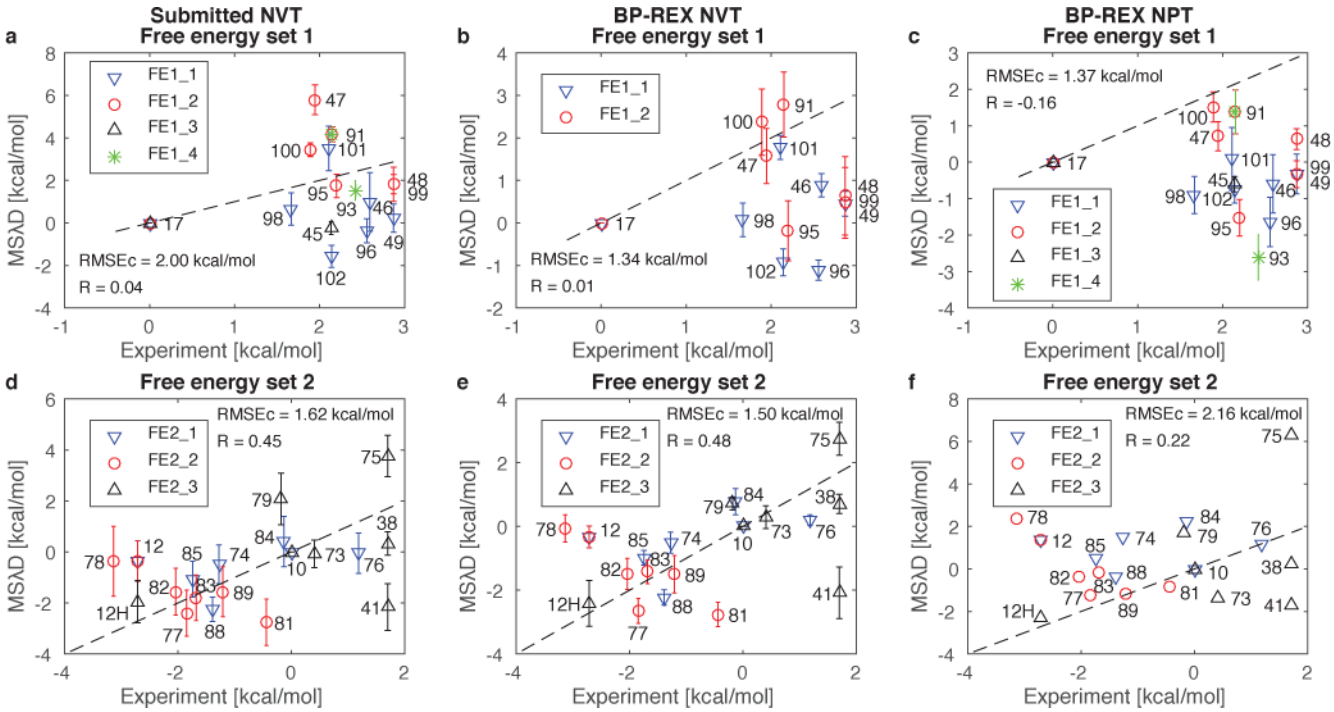



**Fig. 7.**

Computed free energies of binding (kcal/mol) versus experimental results for FE1 (a–c) and FE2 (d–f). Submitted results are shown (a and d) along with results of BP-REX simulations in the NVT (b and e) and NPT (c and f) ensembles. The black dashed lines represents the ideal $y = x$ correlation. Centered RMSE (RMSEc) and Pearson correlation coefficient (R) are given for each data set.

**Table 1**

The numerical values of $E_{max}$ (kcal/mol) used for van der Waals, electrostatic attractive, and repulsive energies in soft-core potentials.

| grid potential | $E_{max}$(vdW) | $E_{max}$(att) | $E_{max}$(rep) |
|---|---|---|---|
| soft-grid-1 | 0.6 | −0.4 | 8.0 |
| soft-grid-2 | 3.0 | −20.0 | 40.0 |

## Table 2

Relative Free Energies of Binding (kcal/mol) for Free Energy Set 1

| FXR Ligand | Expt. | Submitted NVT[a] | BP-REX NVT | BP-REX NPT |
|---|---|---|---|---|
| FE1_1 | | | | |
| 17 | 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| 46 | 2.59 | 0.98 ± 1.38 | 0.89 ± 0.28 | −0.59 ± 0.80 |
| 49 | 2.87[b] | 0.23 ± 0.66 | 0.45 ± 0.29 | −0.32 ± 0.54 |
| 96 | 2.56 | −0.37 ± 0.57 | −1.11 ± 0.24 | −1.64 ± 0.68 |
| 98 | 1.67 | 0.66 ± 0.76 | 0.07 ± 0.39 | −0.90 ± 0.51 |
| 101 | 2.11 | 3.50 ± 1.05 | 1.77 ± 0.28 | 0.09 ± 0.86 |
| 102 | 2.14 | −1.58 ± 0.52 | −0.92 ± 0.32 | −0.76 ± 0.36 |
| **MUE** | | **1.90** | **1.83** | **2.58** |
| **RMSE** | | **2.23** | **2.21** | **2.85** |
| FE1_2 | | | | |
| 17 | 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| 47 | 1.95 | 5.80 ± 0.70 | 1.58 ± 0.65 | 0.71 ± 0.40 |
| 48 | 2.87[b] | 1.82 ± 0.80 | 0.64 ± 0.93 | 0.63 ± 0.29 |
| 91 | 2.15 | 4.17 ± 0.35 | 2.78 ± 0.76 | 1.38 ± 0.60 |
| 95 | 2.20 | 1.74 ± 0.55 | −0.18 ± 0.70 | −1.53 ± 0.50 |
| 99 | 2.87[b] | 1.82 ± 0.46 | 0.47 ± 0.83 | −0.33 ± 0.37 |
| 100 | 1.89 | 3.45 ± 0.33 | 2.37 ± 0.78 | 1.52 ± 0.41 |
| **MUE** | | **1.43** | **1.21** | **1.65** |
| **RMSE** | | **1.84** | **1.57** | **2.12** |
| FE1_3 | | | | |
| 17 | 0.00 | 0.00 ± 0.00 | N/C[c] | 0.00 ± 0.00 |
| 45 | 2.14 | −0.24 ± 0.30 | N/C[c] | −0.60 ± 0.21 |
| FE1_4 | | | | |
| 91 | 2.15 | 4.17 ± 0.35 | N/C[c] | 1.38 ± 0.60 |
| 93 | 2.42 | 1.53 ± 0.46 | N/C[c] | −2.61 ± 0.62 |
| All FE1 Groups | | | | |
| **MUE** | | **1.77** | **1.64** | **2.49** |
| **RMSE** | | **2.08** | **1.99** | **2.84** |
| **RMSEc** | | **2.00** | **1.34** | **1.37** |

[a] Submitted results did not use BP-REX enhanced sampling.

[b] Inactive ligand ($IC_{50}$ >100 $\mu$M).

[c] Not calculated (N/C) for BP-REX NVT.

**Table 3**

Relative Free Energies of Binding (kcal/mol) for Free Energy Set 2

| FXR Ligand | Expt. | Submitted NVT[a] | BP-REX NVT | BP-REX NPT[c] |
|---|---|---|---|---|
| FE2_1 and FE2_2 | | | | |
| 12 | −2.71 | −0.36 ± 0.80 | −0.33 ± 0.34 | 1.37 |
| 10 | 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.00 |
| 76 | 1.19 | −0.05 ± 0.79 | 0.20 ± 0.17 | 1.19 |
| 84 | −0.13 | 0.41 ± 0.99 | 0.78 ± 0.41 | 2.25 |
| 74 | −1.28 | −0.49 ± 0.76 | −0.51 ± 0.33 | 1.49 |
| 85 | −1.74 | −1.07 ± 0.70 | −0.98 ± 0.23 | 0.52 |
| 88 | −1.39 | −2.25 ± 0.48 | −2.23 ± 0.24 | −0.36 |
| 89 | −1.21 | −1.56 ± 0.98 | −1.50 ± 0.59 | −1.15 |
| 82 | −2.04 | −1.56 ± 0.92 | −1.49 ± 0.49 | −0.37 |
| 78 | −3.14 | −0.36 ± 1.37 | −0.06 ± 0.43 | 2.35 |
| 77 | −1.85 | −2.40 ± 0.90 | −2.65 ± 0.39 | −1.21 |
| 83 | −1.68 | −1.80 ± 0.89 | −1.42 ± 0.37 | −0.14 |
| 81 | −0.44 | −2.76 ± 0.91 | −2.76 ± 0.38 | −0.83 |
| **MUE** | | **1.09** | **1.16** | **1.86** |
| **RMSE** | | **1.38** | **1.45** | **2.45** |
| FE2_3 | | | | |
| 12 | −2.71 | −1.95 ± 0.83 | −2.41 ± 0.72 | −2.32 |
| 41 | 1.70[b] | −2.16 ± 0.92 | −2.08 ± 0.81 | −1.67 |
| 10 | 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.00 |
| 38 | 1.70[b] | 0.33 ± 0.46 | 0.71 ± 0.30 | 0.26 |
| 73 | 0.41 | −0.08 ± 0.54 | 0.29 ± 0.36 | −1.35 |
| 75 | 1.70[b] | 3.76 ± 0.81 | 2.75 ± 0.52 | 6.31 |
| 79 | −0.18 | 2.08 ± 1.01 | 0.72 ± 0.20 | 1.71 |
| **MUE** | | **1.80** | **1.19** | **2.31** |
| **RMSE** | | **2.12** | **1.70** | **2.75** |
| All FE2 Groups | | | | |
| **MUE** | | **1.26** | **1.11** | **1.88** |
| **RMSE** | | **1.67** | **1.54** | **2.54** |
| **RMSEc** | | **1.62** | **1.50** | **2.16** |

[a] Submitted results did not use BP-REX enhanced sampling.

[b] Inactive ligand ($IC_{50}$ >100 $\mu$M).

[c] Only one production calculation was performed and hence standard deviations were not calculated.