



# HHS Public Access

Author manuscript

*J Chem Inf Model.* Author manuscript; available in PMC 2022 November 26.

Published in final edited form as:

*J Chem Inf Model.* 2022 March 28; 62(6): 1479–1488. doi:10.1021/acs.jcim.2c00047.

## Optimizing Multisite $\lambda$ -Dynamics Throughput with Charge Renormalization

Jonah Z. Vilseck<sup>1,3,||</sup>, Luis F. Cervantes<sup>4,||</sup>, Ryan L. Hayes<sup>1</sup>, Charles L. Brooks III<sup>\*1,2</sup>

<sup>1</sup>Department of Chemistry, University of Michigan, Ann Arbor, MI 48109.

<sup>2</sup>Biophysics Program, University of Michigan, Ann Arbor, MI 48109.

<sup>3</sup>Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana, 46202, United States.

<sup>4</sup>Department of Medicinal Chemistry College of Pharmacy, University of Michigan, Ann Arbor, MI 48109, United States.

### Abstract

With the ability to sample combinations of alchemical perturbations at multiple sites off a small molecule core, multisite  $\lambda$ -dynamics (MS $\lambda$ D) has become an attractive alternative to conventional alchemical free energy methods for exploring large combinatorial chemical spaces. However, current software implementations dictate that combinatorial sampling with MS $\lambda$ D must be performed with a multiple topology model (MTM), which is non-trivial to create by hand, especially for a series of ligand analogs which may have diverse functional groups attached. This work introduces an automated workflow, referred to as *msld\_py\_prep*, to assist in the creation of a MTM for use with MS $\lambda$ D. One approach for partitioning partial atomic charges between ligands to create a MTM, called charge renormalization, is also presented and rigorously evaluated. We find that *msld\_py\_prep* greatly accelerates the preparation of MS $\lambda$ D ready-to-use files and that charge renormalization can provide a successful approach for MTM generation, as long as bookending calculations are applied to correct small differences introduced by charge renormalization. Charge renormalization also facilitates the use of many different force field parameters with MS $\lambda$ D, broadening the applicability of MS $\lambda$ D for computer-aided drug design.

### Graphical Abstract

<sup>\*</sup>**Corresponding Author** Charles L. Brooks III, 930 N. University Ave., Department of Chemistry and Biophysics, Ann Arbor, MI 48109, brookscl@umich.edu.

<sup>||</sup>These authors contributed equally.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

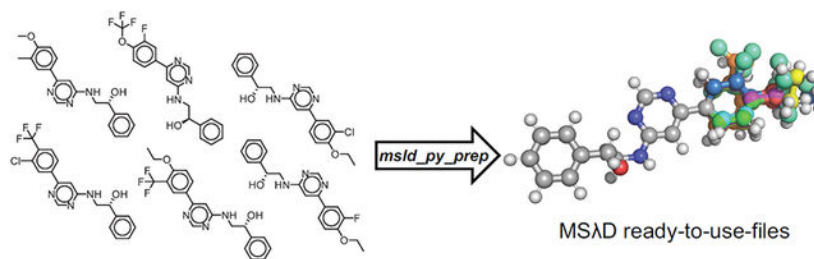
Supporting Information.

The following files are available free of charge.

Supporting figures and tables: chemical structures of all ligands investigated, thermodynamic cycles, overlapping energy distributions, computed hydration free energies, computed FAAH-ligand binding free energies, and FAAH ligand dihedral comparison. (PDF)

The authors declare no competing financial interests.

Data and Software Availability: *msld\_py\_prep* scripts and an accompanying tutorial can be found at [github.com/Vilseck-Lab/msld-py-prep](https://github.com/Vilseck-Lab/msld-py-prep) for download.



## INTRODUCTION

Alchemical free energy calculations have become an integral part of hit-to-lead optimization in both academic and industrial structure-based drug design.<sup>1-7</sup> The past two decades have seen a flurry of developments to improve both accuracy and efficiency of these methods, enabling more reliable prospective binding affinity predictions for small molecule derivatization and optimization.<sup>1-10</sup> Of note, large efficiency gains have been observed from the collective sampling of multiple ligands in a single simulation with  $\lambda$ -dynamics,<sup>11-13</sup> enveloping distribution sampling,<sup>14-17</sup> or  $\lambda$ -local elevation umbrella sampling ( $\lambda$ -LEUS) methods.<sup>18-20</sup>  $\lambda$ -Dynamics, the subject of this manuscript, has similarly undergone a variety of advancements since its introduction in 1996 to position it as an attractive alternative to traditional free energy perturbation (FEP)<sup>21</sup> or thermodynamic integration (TI)<sup>22</sup> methods. Several recent reports have shown the utility of  $\lambda$ -dynamics for drug discovery,<sup>23-25</sup> and a commercial workflow for widespread application of this approach has been developed.<sup>6</sup>

One advancement of  $\lambda$ -dynamics, termed multisite  $\lambda$ -dynamics (MS $\lambda$ D), enables multiple substituents to be alchemically perturbed at one or more sites of attachment off a common molecular core.<sup>13</sup> This approach enables MS $\lambda$ D to break through an inherent scalability limitation associated with FEP or TI to sample large combinatorial alchemical spaces very efficiently. Baseline estimates suggest up to an order of magnitude speed up is obtainable with MS $\lambda$ D over TI for investigating a congeneric series of 10-30 compounds, and even larger efficiency gains of 20-75 times have been reported for sampling many hundreds of alchemical end-states.<sup>6,23-25</sup>

In the current implementation of MS $\lambda$ D in the CHARMM molecular simulation package,<sup>26,27</sup> the alchemical sampling of multiple substituents is accomplished by using a multiple topology model (MTM).<sup>24</sup> Similar to the double topology model used with FEP/TI methods,<sup>28</sup> a MTM represents all alchemical substituent atoms explicitly and scales their atomic interactions with neighboring atoms and molecules via the alchemical coupling parameter,  $\lambda$ . In MS $\lambda$ D, each  $\lambda_{s,i}$ , for the  $i^{\text{th}}$  substituent at site  $s$  off a ligand core, is holonomically constrained to be  $0 < \lambda_{s,i} < 1$ , and all site specific  $\lambda_{s,i}$  must sum to 1.0 to ensure that only one physical end state is sampled at any given point in time. With these constraints in place, the  $\lambda_{s,i}$  parameters can then dynamically sample between interacting ( $\lambda_{s,i} \cong 1.0$ ) and non-interacting ( $\lambda_{s,i} \cong 0.0$ ) states in a continuous manner in conjunction with the forces in a molecular dynamics (MD) simulation using an extended Lagrangian approach.<sup>13,29</sup> The dynamic nature of the  $\lambda_{s,i}$  parameters removes the need to define discrete

intermediate  $\lambda_{s,i}$  states and allows relative free energy differences between many ligands to be readily computed from a single MS $\lambda$ D simulation.

Though use of a MTM facilitates efficient sampling of multiple substituents within a single simulation, adaptation of force field parameters across a congeneric ligand series to fit this topological representation can be difficult. This is especially true when considering the varying electrostatic characteristics different functional groups may present. For example, it is well known that electron-withdrawing or electron-donating groups can shift the distribution of electrostatic charge around an aromatic ring.<sup>30</sup> Thus, to create a MTM from a pre-selected group of ligand analogs, a single set of unperturbed core atoms and alchemical substituents that should be scaled by  $\lambda$  must be defined in such a way that the combination of any one substituent at each perturbation site combined with the core will yield a complete molecule with integer net charge. Correctly partitioning diverse sets of ligand analogs, which may feature substituents that vary in size, flexibility, or polarity, and their accompanying force field parameters to fit into a MTM is a non-trivial problem to solve manually. In this work, we first derive a workflow, expressed as a series of Python scripts collectively referred to as *msld\_py\_prep*, to automate the generation of MTM files for use with MS $\lambda$ D in the CHARMM software package. We then discuss our approach for partitioning force field partial atomic charges for an MTM, which we call charge renormalization, and show that it enables electrostatically different functional groups to be collectively sampled with MS $\lambda$ D.

## ALGORITHMIC WORKFLOW

In this section we discuss our workflow for the creation of a multiple topology model to represent multisite substituent perturbations with MS $\lambda$ D in the CHARMM molecular simulation package. Collectively we refer to this algorithm as *msld\_py\_prep*, and a series of Python scripts have been written to perform each operation in the workflow (Figure 1). The details of each step of *msld\_py\_prep* are described below. As an overview, a maximum common substructure search (MCSS) is first performed to identify a set of core atoms with matching chemical environments in all ligands of interest. These core atoms are characterized as sharing identical atom types and bonded neighbors; atoms with different chemical environments are categorized into alchemical functional groups. Second, appropriate partial atomic charges for core atoms and alchemical atoms are determined to yield correct net charges for all combinatorial molecule end-states. All other intramolecular parameters for the combinatorial system are then identified. Finally, all necessary files for running a MS $\lambda$ D simulation in CHARMM are written to disk. Prior to performing *msld\_py\_prep*, all ligand analogs should be docked and aligned within a protein's binding site so that the identified ligand cores are spatially near one other. This is helpful for successful completion of the MCSS step of *msld\_py\_prep*, and it ensures that the ligands maintain the same binding orientation and pose within the binding site, thus enabling the sampling of a diverse set of ligands. Furthermore, molecule coordinate files should employ the Tripos Mol2 format with one molecule per file, all ligand atoms should have unique atom names per file, and appropriate topology and parameter files formatted for CHARMM should be generated for each molecule prior to running *msld\_py\_prep*. Incidentally, the Mol2 coordinate file requirements match what is needed to obtain CGenFF parameters from the ParamChem/CGenFF or MATCH atom typing programs, and *msld\_py\_prep* works with

outputs obtained from both utilities.<sup>31-34</sup> The following subsections describe each step of *msld\_py\_prep* in more detail.

### Maximum Common Substructure Search (*msld\_mcs.py*).

The first step in preparing a MTM for MS $\lambda$ D is a MCSS to identify which atoms will be in the ligand core and which will be in an alchemical substituent. With *msld\_mcs.py*, a MCSS is performed by identifying atoms in separate molecules with identical force field atom types and lists of bonded neighbor atom types. This creates an effective fingerprint of each atom's chemical environment, as defined by the force field. Similarities between atoms in different molecules are identified by matching atoms with the same atom type and then matching atoms with identical bonded-neighbor lists. This analysis generally resolves most atom-to-atom matches between molecules; however, any remaining discrepancies are resolved by calculating a coordinate root mean square deviation (RMSD) between potential atom matches, emphasizing the need for all ligand cores to be aligned prior to MTM construction. Atoms within a predefined RMSD threshold value of 0.8 Å are considered a match. Several redundancy checks are then performed to ensure that the same number of core atoms are matched from each molecule, resulting in a single uniform set of core atoms for use with MS $\lambda$ D. In a second stage of the MCSS routine, all non-core atoms in each molecule are categorized as alchemical atoms. Bonds between alchemical atoms are used to form alchemical fragments, and each substituent is grouped by its site of attachment to the ligand core. Bonds between alchemical substituents and different core atoms define unique perturbation sites around the core, and fragments at each site are filtered to remove possible duplicates. Finally, a formatted text file of the MCSS output is printed. This file identifies all core atom names from each molecule, the number of alchemical perturbation sites, and a list of all fragments and fragment atom names at each perturbation site. Within the *msld\_py\_prep* workflow, MCSS by *msld\_mcs.py* is typically performed as an independent step with manual visualization of the MCSS results prior to charge renormalization (Figure 1). This ensures that the user can visually examine the MCSS output and confirm that the ligand system has been partitioned into core and alchemical components correctly before proceeding with *msld\_py\_prep*. To assist with visual inspection of the MCSS output, the Python script *vis\_check.py* can be used to quickly visualize core and fragment atoms in PyMOL (Figure S1).<sup>35</sup> We also note that many other MCSS routines exist in a variety of open-source programs, such as RDKit, that could also be used for this step of the workflow.<sup>36</sup> The advantages of the *msld\_mcs.py* MCSS procedure used here is the ability to separate atoms based on their force field atom types and bonded-environments, rather than elemental identities alone.

### Charge Renormalization (*msld\_crn.py*).

Following the correct identification of core atoms and alchemical functional groups for a series of congeneric ligands, a process called charge renormalization (CRN) begins. In this routine, partial atomic charges for each MTM group are identified such that any one alchemical fragment, at each perturbation site, combined with the ligand core will result in a physical ligand end-state with integer net charge (Figure 2). To begin, a preliminary set of partial atomic charges are identified for all core atoms by averaging together the atomic charges of equivalent atoms from each ligand. These initial charges are later adjusted

to neutralize a molecule's net charge after determining the atomic charges for atoms in each alchemical fragment. Charges for atoms in alchemical fragments are obtained next. To facilitate correct combinatorial sampling with MS $\lambda$ D, each alchemical substituent at a single perturbation site must maintain the same net group charge, or in cases where one substituent may feature a charge change, have the same offset group charge with respect to the core's net charge to yield an integer charge for the full molecule. Thus, to generate appropriate partial atomic charges to meet this criterion, a site-specific target group charge is first determined by averaging the net charges of all substituents at a single perturbation site. For substituents with charge changes, group charges are first offset by their charge change prior to averaging to avoid large deviations in computed averages. The difference between the site-specific target charge and a substituent's actual group charge is calculated, and atomic charges within that group are iteratively adjusted by  $\pm 0.000001 e^-$  until the substituent's net charge matches the target group charge. This is performed for all alchemical fragments at all perturbation sites connected to the common core. Finally, the partial atomic charges of the core atoms are revisited to ensure an integer net charge is obtained for a complete ligand end-state. A final target charge for the core is determined by subtracting each site-specific alchemical group charge from the expected final net charge of the ligand series. The initial core atomic charges are then iteratively adjusted by  $\pm 0.000001 e^-$  until the core's net charge matches the target charge. At the conclusion of charge renormalization, all functional group combinations should be feasibly sampled without introducing artifacts into the simulation caused by non-integer net charges for different physical end-states. Though charge renormalization formally changes the force field parameters of a ligand, we've found in practice that the changes are small. Functional groups that differ drastically in size or electrostatic characteristics can also be grouped by the user in such a way to minimize differences introduced by charge renormalization. As shown in the Results section below, errors introduced into computed free energy differences with this approach are mostly within statistical noise. However, because some outliers are observed, which may be difficult to predict in advance, quick bookending calculations (described below) can be performed *post hoc* to compute free energy differences between charge renormalized and original force field charge states to rigorously correct for differences introduced by charge renormalization.

### Ligand Parameter Consolidation (*msld\_prm.py*).

Following the conclusion of charge renormalization, the remaining intramolecular and Lennard-Jones force field parameters for the MTM are created. Because no atom type modifications occur with charge renormalization, or in *msld\_py\_prep* in general, all original intramolecular and Lennard Jones parameters can be consolidated and printed to a single CHARMM formatted parameter file. Thus, in this step, all force field parameters are extracted from each ligand analog, checked for duplication or potential differences in bond, angle, or dihedral parameter definitions, and written to disk.

### File Generation (*msld\_wrt.py*).

Finally, all files that are needed to run MS $\lambda$ D in CHARMM are written to disk. Core atom coordinates and topology definitions are generated first. All alchemical functional group coordinates and topology definitions are then written in a site-specific manner. Finally, a generic CHARMM script input file to setup and run ligand perturbations with

MSAD is created. This input script is system specific to the ligand series processed with *mstd\_py\_prep*, and requires minimal modification prior to initiating a MSAD simulation, such as inserting the appropriate protein, solvent, and ion file designations.

## COMPUTATIONAL DETAILS

With the assistance of *mstd\_py\_prep*, the generation of a MTM for examining a wide range of small molecule perturbations with MSAD in CHARMM is readily accomplished. MSAD can then be used to compute changes in solvation and binding free energies between multiple ligand analogs. Because charge renormalization formally changes a force field's partial atomic charge model for a given molecule, it is important to evaluate how impactful these changes are to the final solvation or binding free energy results computed with MSAD. Small molecule hydration free energy calculations have long served as a test for both new alchemical free energy methods and force field parameterization schemes.<sup>37-42</sup> Thus, to assess the impact of charge renormalization, relative hydration free energies between original force field (FF) and renormalized charges ( $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$ ) were computed for a variety of small molecule analogs grouped into congeneric-like series (Figure S2 - S7). Specifically, four sets of small molecules originating from CGenFF and two sets of drug-like molecules were investigated, including one set of compounds obtained from the ChEMBL database and another of known fatty acid amide hydrolase (FAAH) receptor inhibitors.<sup>31-33,43-45</sup> The impact of charge renormalization on computed binding affinities was additionally investigated using the same group of FAAH inhibitors bound to the FAAH receptor.

### CGenFF Small Molecule Setup.

A common functionalization approach in computer-aided lead optimization is to explore a variety of substituents at ortho, meta, and/or para positions off aromatic rings on a lead compound. Thus, four groups of aromatic compounds explicitly parameterized in the CGenFF force field were examined.<sup>31-33</sup> These are referred to as “123benz”, “12benz”, “12fused”, and “1benz” to reflect their patterns of substitution (Figures S2 - S5). A pdb file for each molecule was built with CHARMM,<sup>26,27</sup> using the internal coordinate table definitions in the CGenFF topology file (version 4.1).<sup>31-33</sup> A mol2 file was obtained with OpenBabel,<sup>46</sup> and a CGenFF stream file was generated with the ParamChem/CGenFF program.<sup>32,33</sup> Spatial alignment was performed in Chimera,<sup>47</sup> and a MTM was created with *mstd\_py\_prep*.

### ChEMBL Small Molecule Setup.

A series of drug-like molecules obtained from the ChEMBL database was examined next (Figure S6; referred to as “ChEMBL”).<sup>43,44</sup> By scanning the ChEMBL database with RDKit, a total of 45 molecules containing a common ‘c1ccc2c(N)ncnc2c1’ substructure was selected for study. This scaffold is a common bicycle frequently found in drug candidates.<sup>48,49</sup> Filtering of compounds with this scaffold core was performed to allow for two sites of substitution off the ligand. Compound SMILES strings were used to generate mol2 files with RDKit,<sup>36</sup> which were then used to obtain CGenFF parameters with the

ParamChem/CGenFF program.<sup>31-33</sup> Ligands were aligned with RDKit and a MTM was created with *msld\_py\_prep*.

### FAAH Inhibitor Setup.

Finally, a series of 32 FAAH inhibitors,<sup>45</sup> which have been studied previously with MS $\lambda$ D,<sup>6</sup> was examined (Figure S7). Initial 3D ligand conformations were generated with RDKit from SMILES strings using the ETKDG method.<sup>50</sup> The molecules were then aligned and minimized in RDKit using the MMFF94 force field.<sup>51-52</sup> For hydration calculations, all ligands were aligned with respect to ligand 31. For FAAH binding calculations, the ligands were aligned to the prepared FAAH receptor's co-crystallized ligand structure. All alignments were performed with RDKit.<sup>36,45</sup> MCSS and charge renormalization were then performed with *msld\_py\_prep*.

### FAAH Receptor Preparation.

Initial coordinates for the FAAH receptor-ligand complex were obtained from PDB ID: 6MRG.<sup>45</sup> Protonation states for titratable residues consistent with a pH of 7.4 were assigned with ProPKA.<sup>53</sup> With the CHARMM-GUI, the protein-ligand complex was then solvated in a cubic box of TIP3P water with an NaCl concentration of 0.145 M, allowing for a minimum of 7.0 Å between the complex and the edge of the box.<sup>54,55</sup> The protein was parameterized with the CHARMM36 force field.<sup>56,57</sup> This setup matches what was used previously by Raman, *et al.*, facilitating direct comparison between previously published MS $\lambda$ D results for FAAH-ligand binding with MATCH-parameterized ligands versus the use of ParamChem/CGenFF-parameterized ligands reported herein.<sup>6</sup>

### Hydration Free Energy Calculations.

Following the creation of a MTM for each ligand series, including CGenFF, ChEMBL, and FAAH compounds, hydration free energies between original and renormalized partial atomic charges were computed for each small molecule. Each molecule was first solvated in a cubic box of TIP3P water with a 12.0 Å buffer between solute and box edge with the MMTSB utility *convpdb.pl*.<sup>58</sup> Molecular dynamics simulations were then performed in vacuum and water-solvated states with both original and renormalized force field charges. MD simulations were run for 5 ns each, in the NPT ensemble at 25 °C and 1 atm, with CHARMM and the domain decomposition (*domdec*) module for running on graphic processing units (GPUs).<sup>26,27,59</sup> Periodic boundary conditions were employed, and long-range interactions were truncated with force-switching between 10 – 12 Å.<sup>60</sup> MD frames were saved every 500 steps, and a timestep of 2 fs was used. Following the conclusion of MD sampling, the trajectories were postprocessed and hydration free energy differences ( $\Delta G_{hyd}^{FF \rightarrow CRN}$ ) were calculated with the Bennett acceptance ratio (BAR) free energy estimator.<sup>61</sup> Free energies and percent overlap between end-point energy distributions were computed with *pymbar.py*.<sup>62,63</sup> Final relative free energy differences with respect to the original force field ( $\Delta \Delta G_{hyd}^{FF \rightarrow CRN}$ ) were obtained with a thermodynamic cycle for modeling solvation processes (Figure S8A).

### MS $\lambda$ D FAAH Binding Calculations.

Retrospectively, FAAH – ligand binding affinities were investigated with MS $\lambda$ D for 32 FAAH inhibitors,<sup>45</sup> modeled with the CGenFF force field.<sup>31-33</sup> The simulation setup and details matched what was performed previously by Raman, *et al.*, except that the simulations were performed directly in CHARMM.<sup>6</sup> Free energy differences were calculated by subdividing the full set of 32 ligands into 5 separate MS $\lambda$ D calculations. Perturbations were performed with ligands solvated in solution (“water-unbound”) and bound to FAAH (“protein-bound”), in accordance with a protein-ligand binding thermodynamic cycle (Figure S8B). Bias optimization with the Adaptive Landscape Flattening (ALF)<sup>23</sup> algorithm was performed in three stages, with 100 iterations of 100 ps long MS $\lambda$ D simulations, followed by 13 iterations of 1 ns long MS $\lambda$ D simulations, and concluding with 5 independent trials of 5 ns long simulations. Production MS $\lambda$ D simulations were then run with 5 independent trials for 20 ns and 50 ns for water-unbound and protein-bound states, respectively. Binding free energy differences ( $G_{\text{bind}}$ ) for each state were calculated as a ratio of the amount of sampling compared to a single reference compound, using a  $\lambda$ -cutoff value of 0.99 to approximate the  $\lambda = 1.00$  end-states.<sup>13,23-25</sup> Final relative differences in free energies of binding ( $G_{\text{bind}}$ ) between ligands were computed by taking the difference between water-unbound and protein-bound ligand states. As mentioned above, a MTM is necessary to sample multiple perturbations simultaneously, thus these MS $\lambda$ D binding calculations were performed with renormalized partial atomic charges ( $\Delta\Delta G_{\text{bind}}^{\text{CRN}}$ ). Absolute, computed binding free energies ( $G_{\text{comp}}$ ) were determined with the following equation, which utilizes known experimental binding free energies ( $G_{\text{expt}}$ ) and the computed relative binding free energies ( $G_{\text{comp}}$ )<sup>4</sup>:

$$\Delta G_{\text{comp}} = \Delta\Delta G_{\text{comp}} - \left( \frac{\sum \Delta\Delta G_{\text{comp}}}{n} - \frac{\sum \Delta G_{\text{expt}}}{n} \right) \quad (1)$$

### MS $\lambda$ D Bookending Corrections.

Similar to the hydration free energy calculations described above, single step charge perturbations were performed with BAR between original force field and renormalized atomic charges to assess the impact of charge renormalization on MS $\lambda$ D binding predictions. As shown in Figure 3, “bookending” correction terms ( $\Delta G_{\text{bind}}^{\text{FF} \rightarrow \text{CRN}}$ ) were computed *post hoc* to correct the MS $\lambda$ D  $\Delta G_{\text{bind}}^{\text{CRN}}$  results obtained with renormalized charges to the true force field free energy result ( $\Delta G_{\text{bind}}^{\text{FF}}$ ). These correction terms were calculated by running MD simulations in water-unbound and protein-bound ligand states, with original and renormalized atomic charges separately. Simulation parameters matched what was used for the MS $\lambda$ D binding calculations to ensure sufficient overlap between simulations would be obtained, and simulations were run for 5 ns each. Although one bookending correction value requires four additional simulations to be run, these simulations are independent from MS $\lambda$ D and can be run in parallel with MS $\lambda$ D production calculations.



## RESULTS

To test the effectiveness of charge renormalization and measure the extent of error introduced by adjusting a small molecule's partial atomic charges to fit within a MTM, hydration and binding free energy calculations were performed on a total of 191 compounds spanning 6 ligand series. The degree of change introduced by charge renormalization was analyzed as a charge root mean square deviation (Q-RMSD), and by measuring changes in dipole moment magnitudes ( $\mu$ ) and directions ( $\mu(\theta)$ ). Complete data for each ligand and ligand series can be found in Tables S1 - S6 of the Supporting Information. As shown in Table 1, the average  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  for all molecules is ca. 0.1 kcal/mol, which is generally within the limits of statistical noise for solvation transfer processes investigated with alchemical free energy methods.<sup>37-42</sup> Overall, the Q-RMSD and the  $\mu$  metrics, which represent the degree of change between renormalized and original force field charges, tended to be small. This suggests that, on average, differences introduced by charge renormalization should also be small. However, when larger deviations in Q-RMSD or  $\mu$  occur, for instance for the 12fused series of molecules, larger differences in  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  are subsequently observed. This trend is clear in Figure 4, which plots  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  in relation to Q-RMSD and  $\mu$ . As Q-RMSD grows larger than  $0.001 e^-$ ,  $\log_{10}(\text{Q-RMSD}) > -3.0$ , large deviations in both  $\mu$  and  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  can occur. Above this Q-RMSD threshold, while most  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  remains small, e.g.,  $< 0.5$  kcal/mol, some outliers with  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN} > 1.0$  kcal/mol are also observed. The largest outliers unanimously come from the 12fused series of molecules. In this series, the core is defined as two carbon atoms and two hydrogen atoms that are located opposite the bond connecting the fused rings. A deeper analysis of the electrostatic descriptions of these 41 compounds indicated that many of these molecules have significant charge discrepancies for these four equivalent core atoms. Thus, while there is sufficient atom type overlap to identify these atoms as core atoms by the MCSS routine in *msld\_py\_prep*, there is poorer electrostatic overlap which leads to large Q-RMSD and  $\mu$  by charge renormalization and, subsequently, large  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  differences. In ligands that have a larger number of core atoms, this situation may be remedied by manually assigning core atoms with large electrostatic discrepancies into adjacent alchemical substituents, thereby reducing the magnitude of charge modifications by charge renormalization. This can be easily performed with the "InFrag" option in *msld\_py\_prep*. Checking Q-RMSD values after running *msld\_py\_prep* can help identify if this is necessary. Furthermore, to eliminate potential  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  errors, we recommend calculating  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  bookending corrections whenever charge renormalization is used to create an MTM. Encouragingly, the drug-like ChEMBL and FAAH ligand series showed among the lowest  $\Delta\Delta G_{hyd}^{FF \rightarrow CRN}$  changes. Future applications of MSAD to ligand design and lead optimization will show how impactful charge renormalization is for analyzing many diverse sets of drug-like compounds.

To begin to explore this idea, binding free energy calculations were computed with MSAD for the 32 FAAH ligands bound to the FAAH receptor. The analogs contained a wide variety of phenyl ring substitutions with differing chemical properties. Charge renormalization

was performed on the entire set of ligands, and relative binding free energies were calculated from 5 separate MS $\lambda$ D calculations. Initial MS $\lambda$ D  $\Delta G_{bind}^{CRN}$  results obtained with renormalized charges were then corrected with bookending calculations to obtain  $\Delta G_{bind}^{FF}$ , the CGenFF force field result. Absolute free energies of binding are shown in Figure 5 for both data sets (see also Table S7). Overall, the uncorrected MS $\lambda$ D results yielded a Pearson correlation of 0.54 ( $\pm$  0.001), a Spearman correlation of 0.56 ( $\pm$  0.001), and a Kendall tau of 0.38 ( $\pm$  0.001); yet the MUE was low at 1.07 ( $\pm$  0.001) kcal/mol, a common target for alchemical binding free energy studies.<sup>4-10</sup> Symmetric confidence intervals were computed at the 95% confidence level with the leave-one-out method using 1000 iterations.<sup>64</sup> The corrected MS $\lambda$ D results, however, showed improved Pearson, Spearman, and Kendall tau coefficients of 0.60 ( $\pm$  0.001), 0.64 ( $\pm$  0.001), and 0.45 ( $\pm$  0.001), respectively, and the MUE slightly improved to 1.00 ( $\pm$  0.001) kcal/mol. Though Table 1 suggests that the Q-RMSD and  $\mu$  are small for the FAAH ligands, these binding results indicate that charge renormalization does affect MS $\lambda$ D free energy calculations in a non-trivial manner and that improved accuracy can be obtained by correcting MS $\lambda$ D results obtained with renormalized charges to the true force field result. Fortunately, a single step charge perturbation from original force field charges to renormalized charges allows charge renormalization differences to be corrected quickly without significantly impacting MS $\lambda$ D scalability. For example, the MS $\lambda$ D FAAH binding calculations performed herein utilized a total of 2.23  $\mu$ s of sampling, with an extra 0.64  $\mu$ s of MD for bookending corrections. Raman, *et al.*'s estimate for running FEP or TI calculations for all 32 FAAH ligands, with cycle closure loops and 5 duplicate production calculations to match our MS $\lambda$ D procedure, suggests 14.4  $\mu$ s of sampling would be required with FEP/TI methodologies to obtain equivalent  $G_{bind}$  results. Based solely on amounts of sampling, MS $\lambda$ D alone is ca. 6.5 times more efficient than FEP/TI. This comparison negates the contribution of uncertainty in computed estimates of computational efficiency. MS $\lambda$ D efficiency dips only slightly to ca. 5.0 times more efficient when bookending correction simulations are included. Thus, MS $\lambda$ D remains highly efficient, in terms of sampling along, even when  $G^{FF \rightarrow CRN}$  corrections are included. It is important to note that accuracy and correlation to experiment are force field dependent and, thus, an ancillary focus of this paper. Fortunately, charge renormalization provides a means to employ new or non-CHARMM-based force field parameters with MS $\lambda$ D, allowing accuracy to be tuned if necessary for specific systems of interest, while still using a highly efficient and scalable free energy method.

Finally, the results from a previous MS $\lambda$ D investigation<sup>6</sup> of the same set of 32 FAAH inhibitors bound to the FAAH receptor can be discussed.<sup>45</sup> That work employed commercial MATCH/CGenFF ligand parameters<sup>34</sup> and reported a MUE of 0.70 kcal/mol for computed free energies of binding compared to experiment.<sup>6</sup> An analysis of their force field parameters was performed to understand potential sources of differences compared to the ParamChem/CGenFF parameters we used. This revealed large deviations in the partial atomic charges, with an average Q-RMSD of 0.069  $e^-$ . The trends in Figure 4 suggest large relative free energy differences would be highly likely, a pertinent reminder that electrostatic definitions can have strong impacts on  $G_{hyd}$  or  $G_{bind}$  calculations. Differences in dihedral parameters were also observed. Table S8 shows 6 dihedral angles found in the FAAH ligands where energetic differences between MATCH and ParamChem/CGenFF dihedral

definitions were notable (RMSD > 1.0 kcal/mol). Most of these definitions can be mapped to the more flexible regions of the FAAH molecule core, such as torsions in the linker between the 4-aminopyrimidine and hydroxyl groups. Differences in torsional preferences could lead to sampling alternative ligand conformations or lead to increased ligand strain upon binding, which could also help explain MUE differences between this work and Raman, *et al.* Thus, this comparison emphasizes the importance of obtaining accurate ligand dihedral parameters and partial atomic charges when performing free energy calculations.

## DISCUSSION AND CONCLUSIONS

In this work, we've derived a new framework, instantiated as a set of Python scripts collectively called *msld\_py\_prep*, to automate the creation of multiple topology models to investigate single- and multi-site perturbations of small molecule functional groups with MSAD in the CHARMM molecular simulation package. As a key step in this workflow, a process called charge renormalization has been introduced to facilitate the collective modeling of a series of structurally diverse ligands with substituent changes at potentially many different sites of attachment. Within *msld\_py\_prep*, a MCSS partitions ligands into a set of core atoms and groups of alchemical fragments for sampling with MSAD; charge renormalization then processes and adjusts atomic charges to facilitate sampling of many alchemical substituents with a single ligand core, while maintaining correct integer net charges for each ligand end-state. Our evaluation of the effects of using charge renormalization for computing free energy differences has shown that while most differences introduced by charge renormalization are small ( $G^{FF \rightarrow CRN} < 0.5$  kcal/mol), significant outliers ( $G^{FF \rightarrow CRN} > 1.0$  kcal/mol) are sometimes observed. Since the  $G^{FF \rightarrow CRN}$  magnitude cannot be estimated *a priori* for an arbitrary ligand series, it is recommended that *post hoc* bookending corrections be calculated and applied whenever charge renormalization is used. Fortunately, these corrections are easy to setup and can be run in parallel with an MSAD calculation. Because charge renormalization makes only minor adjustments to a molecule's original set of atomic charges, bookending calculations can consist of a single step charge perturbation with BAR.<sup>61</sup> Convergence in these calculations is established by measuring the degree of overlap between energy distributions of ligand states featuring original or renormalized charges with *pymbar*.<sup>62,63</sup> In this work, high levels of overlap were observed for all charge perturbations (Figures S9 - S15). Furthermore, we note that charge normalization is only required when alchemical functional groups have different net group charges. For example, with the MATCH atom typing program,<sup>34</sup> atoms are charged in such a way that groups of nearby atoms have a neutral or integer net charge; this is not true of ParamChem/CGenFF parameters or most other force fields in popular use. This means that with MATCH parameters it is often feasible to define alchemical fragments in the MTM such that no charge renormalization is needed at all, and in such instances bookending calculations are unnecessary because *msld\_py\_prep* has no effect on the original MATCH parameters. In addition, charge renormalization facilitates the use of many additional force fields with MSAD, which may be useful for modeling many diverse series of congeneric ligands. Finally, files generated with *msld\_py\_prep* are fully compatible with other MSAD utilities, specifically the ALF algorithm,<sup>23</sup> thus adding to the continued and accelerated use of MSAD in a variety of research endeavors, including structure-based drug discovery.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENT

We gratefully acknowledge the many Brooks lab members who helped beta test the *mstd\_py\_prep* workflow.

### Funding Sources

The National Institutes of Health, through grants GM037554, GM130587, and GM107233, are gratefully acknowledged for financial support.

## ABBREVIATIONS

<b>MS<math>\lambda</math>D</b>	multisite $\lambda$ -dynamics
<b>MTM</b>	multiple topology model
<b>FEP</b>	free energy perturbation
<b>TI</b>	thermodynamic integration
<b>MD</b>	molecular dynamics
<b>MCSS</b>	maximum common substructure search
<b>RMSD</b>	root mean square deviation
<b>CRN</b>	charge renormalization
<b>GPU</b>	graphic processing unit
<b>BAR</b>	Bennett acceptance ratio
<b>Q-RMSD</b>	charge root mean square deviation

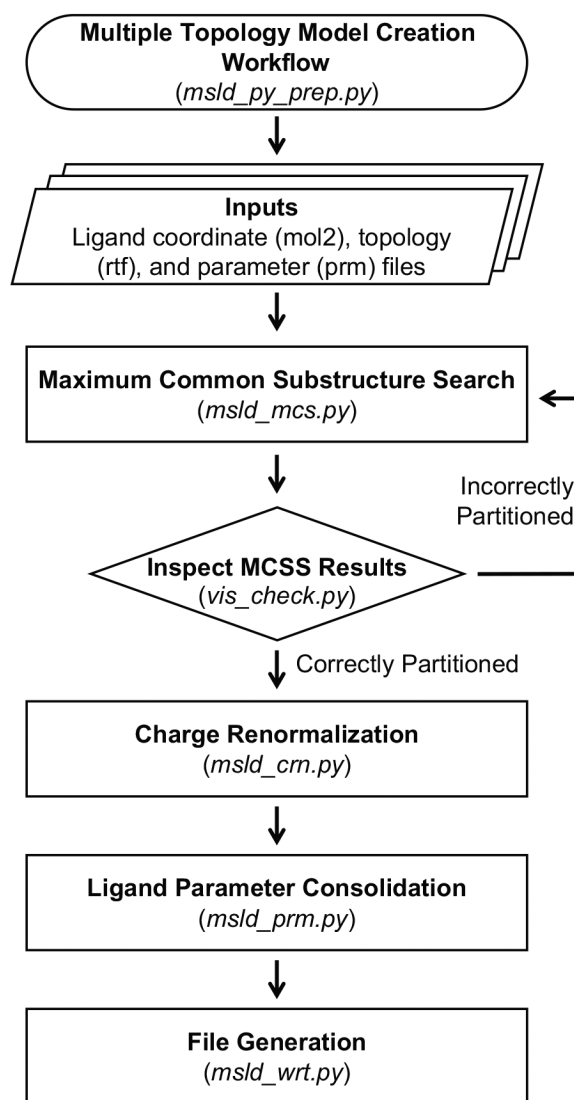
## REFERENCES

- Jorgensen WL The Many Roles of Computation in Drug Discovery. *Science* 2004, 303, 1813–1818. [PubMed: 15031495]
- Song LF; Merz KM Jr. Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model* 2020, 60, 5308–5318. [PubMed: 32818371]
- 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. [PubMed: 29243483]
- Wang L; Wu Y; Deng Y; Kim B; Pierce L; Krilov G; Lypyan 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; Westler 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. *J. Am. Chem. Soc* 2015, 137, 2695–2703. [PubMed: 25625324]
- Schindler CEM; Baumann H; Blum A; Böse D; Buchstaller H-P; Burgdorf L; Cappel D; Chekler E; Czodrowski P; Dorsch D; Eguida MKI; Follows B; Fuchß T; Grädler U; Gunera J; Johnson T; Jorand Lebrun C; Karra S; Klein M; Knehans T; Koetzner L; Krier M; Leiendecker M; Leuthner B; Li L; Mochalkin I; Musil D; Neagu C; Rippmann F; Schiemann K; Schulz R; Steinbrecher T;

- Tanzer E-M; Unzue Lopez A; Viacava Follis A; Wegener A; Kuhn D Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model* 2020, 60, 5457–5474. [PubMed: 32813975]
6. Raman EP; Paul TJ; Hayes RL; Brooks CL III Automated, Accurate, and Scalable Relative Protein-Ligand Binding Free-Energy Calculations Using Lambda Dynamics. *J. Chem. Theory Comput* 2020, 16, 7895–7914. [PubMed: 33201701]
  7. Abel R; Wang L; Mobley DL; Friesner RA A Critical Review of Validation, Blind Testing, and Real-World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. *Curr. Top. Med. Chem* 2017, 17, 2577–2585. [PubMed: 28413950]
  8. Jespers W; Esguerra M; Åqvist J; Gutiérrez-de-Terán H QligFEP: an automated workflow for small molecule free energy calculations in Q. *J. Cheminform* 2019, 11(26), 1–16. [PubMed: 30604073]
  9. Gapsys V; Pérez-Benito L; Aldeghi M; Seeliger S; van Vlijmen H; Tresadern G; de Groot BL Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci* 2020, 11, 1140–1152.
  10. Kuhn M; Firth-Clark S; Tosco P; Mey ASJS; Mackey M; Michel J Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *J. Chem. Inf. Model* 2020, 60, 3120–3130. [PubMed: 32437145]
  11. Kong X; Brooks CL III  $\lambda$  Dynamics: A New Approach to Free Energy Calculations. *J. Chem. Phys* 1996, 105, 2414–2423.
  12. Knight JL; Brooks CL III  $\lambda$  Dynamics Free Energy Simulation Methods. *J. Comput. Chem* 2009, 30, 1692–1700. [PubMed: 19421993]
  13. Knight JL; Brooks CL III Multisite  $\lambda$  Dynamics for Simulated Structure-Activity Relationship Studies. *J. Chem. Theory Comput* 2011, 7, 2728–2739. [PubMed: 22125476]
  14. Christ CD; van Gunsteren WF Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. *J. Chem. Phys* 2007, 126, 184110. [PubMed: 17508795]
  15. Christ CD; van Gunsteren WF Simple, Efficient, and Reliable Computation of Multiple Free Energy Differences from a Single Simulation: A Reference Hamiltonian Parameter Update Scheme for Enveloping Distribution Sampling (EDS). *J. Chem. Theory Comput* 2009, 5, 276–286. [PubMed: 26610104]
  16. Perthold JW; Oostenbrink C Accelerated Enveloping Distribution Sampling: Enabling Sampling of Multiple End States while Preserving Local Energy Minima. *J. Phys. Chem. B* 2018, 122, 5030–5037. [PubMed: 29669415]
  17. Sidler D; Schwaninger A; Riniker S Replica exchange enveloping distribution sampling (RE-EDS): A robust method to estimate multiple free-energy differences from a single simulation. *J. Chem. Phys* 2016, 145, 154114. [PubMed: 27782485]
  18. Bieler NS; Häuselmann R; Hünenberger PH Local Elevation Umbrella Sampling Applied to the Calculation of Alchemical Free-Energy Changes via  $\lambda$ -Dynamics: the  $\lambda$ -LEUS Scheme. *J. Chem. Theory Comput* 2014, 10, 3006–3022. [PubMed: 26588272]
  19. Bieler NS; Hünenberger PH Communication: Estimating the initial biasing potential for  $\lambda$ -local-elevation umbrella-sampling ( $\lambda$ -LEUS) simulations via slow growth. *J. Chem. Phys* 2014, 141, 201101. [PubMed: 25429926]
  20. Bieler NS; Tschopp JP; Hünenberger PH Multistate  $\lambda$ -Local-Elevation Umbrella-Sampling (MS $\lambda$ -LEUS): Method and Application to the Complexation of Cations by Crown Ethers. *J. Chem. Theory Comput* 2015, 11, 2575–2588. [PubMed: 26575556]
  21. Zwanzig RW High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys* 1954, 22, 1420–1426.
  22. Kirkwood JG Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys* 1935, 3, 300–313.
  23. Hayes RL; Armacost KA; Vilseck JZ; Brooks CL III Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite  $\lambda$  Dynamics. *J. Phys. Chem. B* 2017, 121, 3626–3635. [PubMed: 28112940]
  24. Vilseck JZ; Armacost KA; Hayes RL Goh GB; Brooks CL III Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite  $\lambda$  Dynamics. *J. Phys. Chem. Lett* 2018, 9, 3328–3332. [PubMed: 29847134]

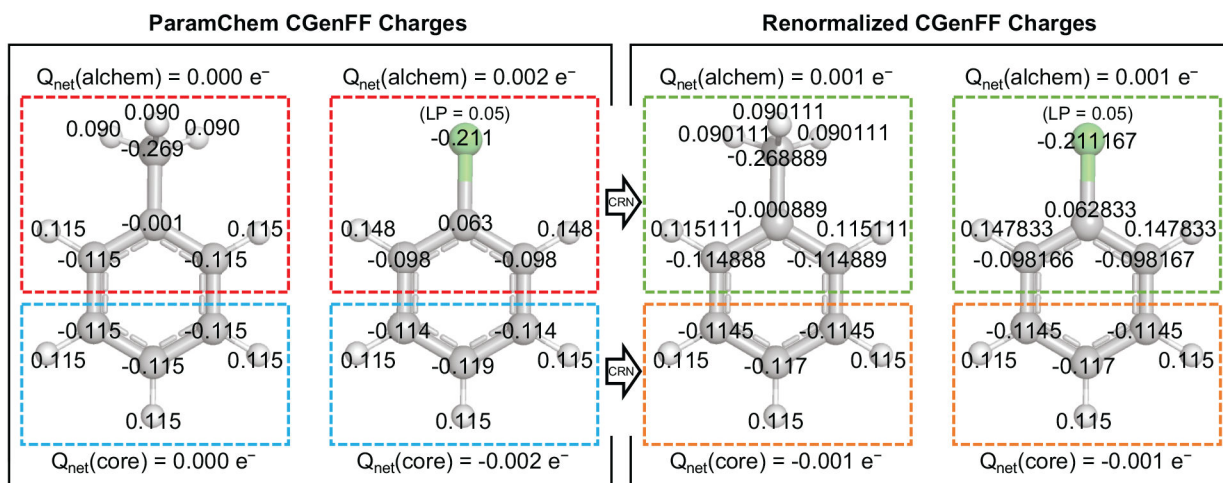
25. Vilseck JZ; Sohail N; Hayes RL; Brooks CL III Overcoming Challenging Substituent Perturbations with Multisite  $\lambda$ -Dynamics: A Case Study Targeting  $\beta$ -Secretase 1. *J. Phys. Chem. Lett* 2019, 10, 4875–4880. [PubMed: 31386370]
26. Brooks BR; Brooks CL III; MacKerell AD Jr; Nilsson L; 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; Hodosek M; Im W; Kuczera K; Lazaridis T; Ma J; Ovchinnikov V; Paci E; Pastor RW; Post CB; Pu JZ; Schaefer M; Tidor B; Venable RM; Woodcock HL; Wu X; Yang W; York DM; Karplus M CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem* 2009, 30, 1545–1614. [PubMed: 19444816]
27. Brooks BR; Brucoleri RE; Olafson BD; States BD; Swaminathan S; Karplus M CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem* 1983, 4, 187–217.
28. Pohorille A; Jarzynski C; Chipot C Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* 2010, 114, 10235–10253. [PubMed: 20701361]
29. Knight JL; Brooks CL III Applying Efficient Implicit Constraints in Alchemical Free Energy Simulations. *J. Comput. Chem* 2011, 32, 3423–3432. [PubMed: 21919014]
30. Smith BS; March J March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure, 6<sup>th</sup> ed.; John Wiley & Sons, Inc., 2007.
31. 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. *J. Comput. Chem* 2010, 31, 671–690. [PubMed: 19575467]
32. Vanommeslaeghe K; MacKerell AD Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model* 2012, 52, 3144–3154. [PubMed: 23146088]
33. Vanommeslaeghe K; Raman EP; MacKerell AD Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model* 2012, 52, 3155–3168. [PubMed: 23145473]
34. Yesselman JD; Price DJ; Knight JL; Brooks CL III MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. *J. Comput. Chem* 2012, 33, 189–202. [PubMed: 22042689]
35. The PyMOL Molecular Graphics System, Version 1.8, Schrodinger LLC, New York, New York, United States.
36. RDKit: Open-source cheminformatics. Landrum, G. 2010. <https://www.rdkit.org/>.
37. Vilseck JV; Tirado-Rives J; Jorgensen WL Evaluation of CM5 Charges for Condensed-Phase Modeling. *J. Chem. Theory Comput* 2014, 10, 2802–2812. [PubMed: 25061445]
38. Mobley DL; Guthrie JP FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput.-Aided Mol. Des* 2014, 28, 711–720. [PubMed: 24928188]
39. Matos GDR; Kyu DY; Loeffler HH; Chodera JD; Shirts MR; Mobley DL Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* 2017, 62, 1559–1569. [PubMed: 29056756]
40. Mobley DL; Bayly CI; Cooper MD; Shirts MR; Kill KA Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J. Chem. Theory Comput* 2009, 10, 350–358.
41. Shivakumar D; Deng Y; Roux B Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput* 2009, 5, 919–930. [PubMed: 26609601]
42. Shivakumar D; Williams J; Wu Y; Damm W; Shelley J; Sherman W Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput* 2010, 6, 1509–1519. [PubMed: 26615687]
43. Mendez D; Gaulton A; Bento AP; Chambers J; De Veij M; Félix E; Magariños MP; Mosquera JF; Mutowka P; Nowotka M; Gordillo-Marañón M; Hunter F; Junco L; Mugumbate G; Rodriguez-Lopez M; Atkinson F; Bosc N; Radoux CJ; Segura-Cabrera A; Hersey A; Leach AR ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 2019, 47, D930–D940. [PubMed: 30398643]

44. Davies M; Nowotka M; Papadatos G; Dedman N; Gaulton A; Atkinson F; Bellis L; Overington JP ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* 2015, 43, W612–W620. [PubMed: 25883136]
45. Saha A; Shih AY; Mirzadegan T; Seierstad M Predicting the Binding of Fatty Acid Amide Hydrolase Inhibitors by Free Energy Perturbation. *J. Chem. Theory Comput* 2018, 14, 5815–5822. [PubMed: 30289722]
46. O'Boyle NM; Banck M; James CA; Morley C; Vandermeersch T; Hutchison GR Open Babel: An open chemical toolbox. *J. Cheminform.* 2011, 3, 33. [PubMed: 21982300]
47. Pettersen EF; Goddard TD; Huang CC; Couch GS; Greenblatt DM; Meng EC; Ferrin TE UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem* 2004, 25, 1605–1612. [PubMed: 15264254]
48. Zdrzil B; Guha R The Rise and Fall of a Scaffold: A Trend Analysis of Scaffolds in the Medicinal Chemistry Literature. *J. Med. Chem* 2018, 61, 4688–4703. [PubMed: 29235859]
49. Taylor RB; MacCoss R; Lawson ADG Combining Molecular Scaffolds from FDA Approved Drugs: Application to Drug Discovery. *J. Med. Chem* 2017, 60, 1638–1647. [PubMed: 27935308]
50. Riniker S; Landrum GA Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J. Chem. Inf. Model* 2015, 55, 2562–2574. [PubMed: 26575315]
51. Tosco P; Stiefel N; Landrum G Bringing the MMFF Force Field to RDKit: Implementation and Validation. *Journal of Cheminformatics.* 2014, 6, 37.
52. Halgren TA Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 1996, 17, 490–519.
53. Olsson MH; Sondergaard CR; Rostkowski M; Jensen JH PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput* 2011, 7, 525–537. [PubMed: 26596171]
54. Jo S; Kim T; Iyer VG; Im W CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem* 2008, 29, 1859–1865. [PubMed: 18351591]
55. Lee J; Cheng X; Swails JM; Yeom MS; Eastman PK; Lemkul JA; Wei S; Buckner J; Jeong JC; Qi Y; Jo S; Pande VS; Case DA; Brooks CL III, MacKerell AD Jr, Klauda JB; Im W CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput* 2016, 12, 405–413. [PubMed: 26631602]
56. 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. *Biophys. J* 2012, 103, 1045–1051. [PubMed: 23009854]
57. 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  $\phi$ ,  $\psi$ , and Side-Chain  $\chi_1$  and  $\chi_2$  Dihedral Angles. *J. Chem. Theory Comput* 2012, 8, 3257–3273. [PubMed: 23341755]
58. Feig M; Karanicolas J; Brooks CL III MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graph. Model* 2004, 22, 377–395. [PubMed: 15099834]
59. Hynninen AP; Crowley MF New Faster CHARMM Molecular Dynamics Engine. *J. Comput. Chem* 2014, 35, 406–413. [PubMed: 24302199]
60. Steinbach PJ; Brooks BR New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J. Comp. Chem* 1994, 15, 667–683.
61. Shirts MR; Bair E; Hooker G; Pande VS Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett* 2003, 91, 140601. [PubMed: 14611511]
62. Shirts MR; Chodera JD Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys* 2008, 129, 124105. [PubMed: 19045004]
63. Klimovich PV; Shirts MR; Mobley DL Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des* 2015, 29, 397–411. [PubMed: 25808134]
64. Leave-One-Out Cross-Validation. In: Sammut C, Webb GI (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. 2011. 10.1007/978-0-387-30164-8\_469

**Figure 1.**

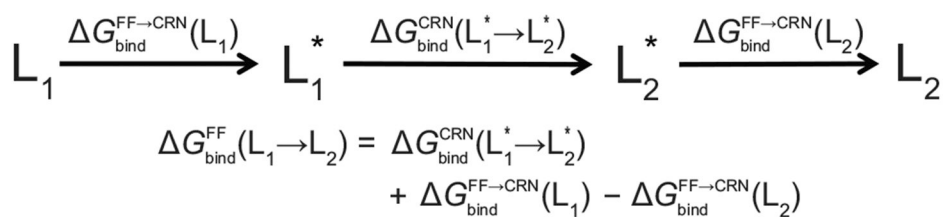
The *msld\_py\_prep* workflow for generating a multiple topology model for use with MSAD in the CHARMM molecular simulation package. Input files should be CHARMM formatted and generated prior to running *msld\_py\_prep*. While inspection of the MCSS results can be performed in any molecular visualization software package, *vis\_check.py* works specifically with PyMOL.



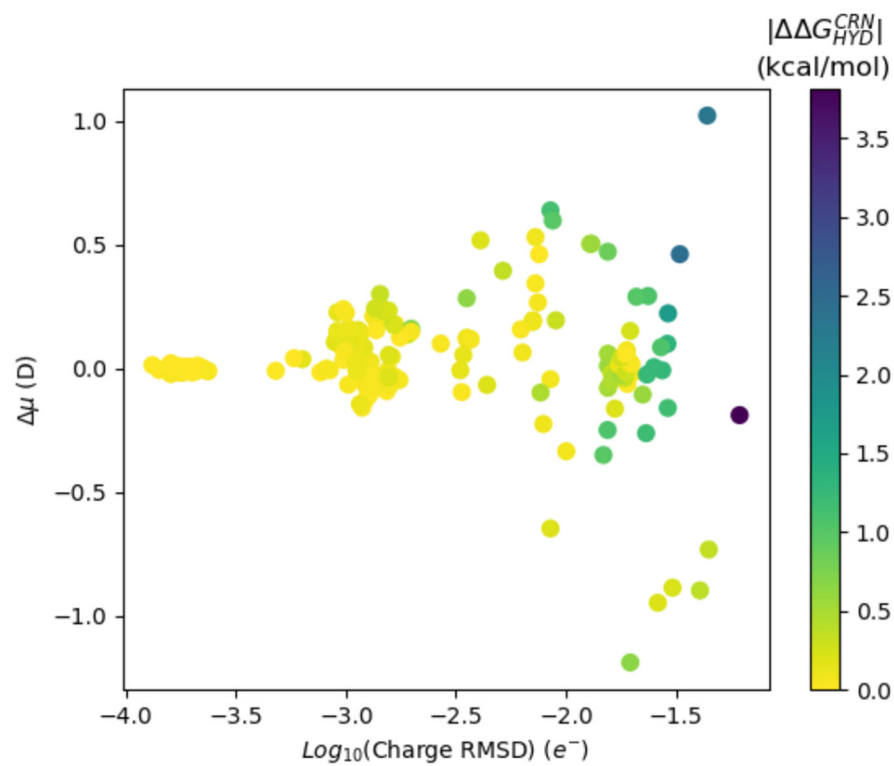


**Figure 2.**

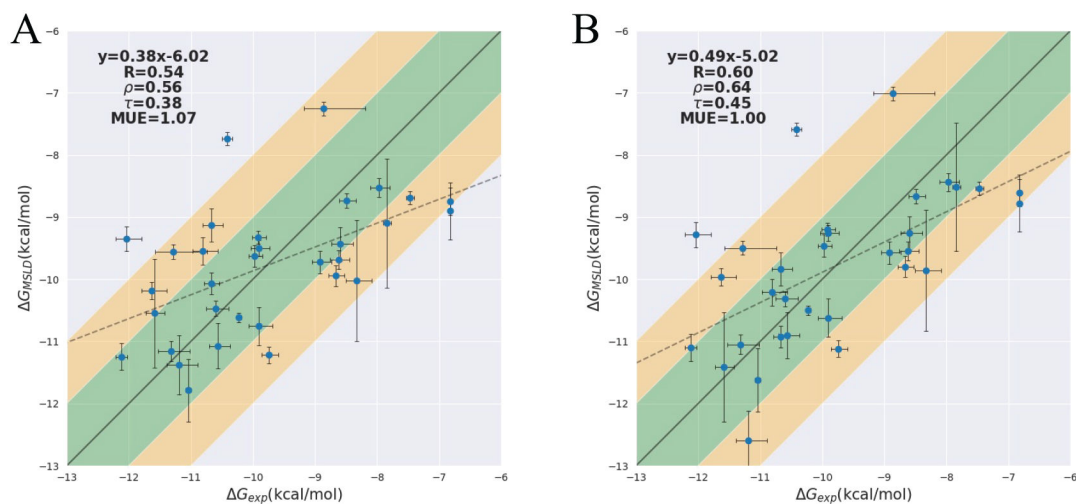
Example charge renormalization (CRN) of toluene and chlorobenzene into a dual topology model. On the left, CGenFF charges obtained from the ParamChem/CGenFF program have been grouped into core (blue box) and alchemical (red box) atom groups. Net group charges differ between toluene and chlorobenzene. Following small charge adjustments via CRN, net core (orange box) and alchemical (green box) group charges now match between these molecules, facilitating their use with MSAD.

**Figure 3.**

Thermodynamic bookending corrections for investigating alchemical transformations with MS $\lambda$ D and renormalized ligand charges within a specific state (gas, water-unbound, or protein-bound). The true force field  $L_1 \rightarrow L_2$  free energy difference ( $\Delta G_{\text{bind}}^{\text{FF}}(L_1 \rightarrow L_2)$ ) is composed of the MS $\lambda$ D free energy difference between two ligands with renormalized charges ( $\Delta G_{\text{bind}}^{\text{CRN}}(L_1^* \rightarrow L_2^*)$ ), plus bookending free energy terms ( $\Delta G_{\text{bind}}^{\text{FF} \rightarrow \text{CRN}}$ ) to correct for charge renormalization differences.



**Figure 4.** Charge RMSD ( $e^-$ ) and changes in dipole moments (D) for 191 small molecules due to charge renormalization. Computed relative free energies of hydration (kcal/mol) between original and renormalized partial atomic charges are color coded as a heatmap.

**Figure 5.**

(A) Correlation between experimental and MS $\lambda$ D computed FAAH – ligand binding affinities (kcal/mol) obtained with renormalized partial atomic charges. (B) Correlation between experimental and corrected MS $\lambda$ D computed binding affinities (kcal/mol), representing the original force field's free energy results.

**Table 1.**

Signed Average Deviations in Computed Relative Free Energies of Hydration (kcal/mol), Charge RMSD ( $e^-$ ), Dipole Moment Magnitudes (D) and Directions (radians) Induced by Charge Renormalization for Six Series of Small Molecules.

Series Name	$\langle \Delta \Delta G_{hyd}^{FF \rightarrow CRN} \rangle \pm \sigma$ (kcal/mol)	$\langle Q - RMSD \rangle (e^-)$	$\langle \mu \rangle$ (D) <sup>a</sup>	$\langle \mu(\theta) \rangle$ (radians) <sup>b</sup>
123benz	0.072 ± 0.002	0.006	-0.018	0.348
12benz	0.200 ± 0.002	0.008	0.112	0.254
12fused	0.451 ± 0.004	0.022	0.034	0.146
1benz	0.028 ± 0.000	0.001	-0.001	0.072
ChEMBL	-0.098 ± 0.002	0.003	0.090	0.077
FAAH	-0.044 ± 0.001	0.002	0.015	0.042
<b>all series</b>	<b>0.101 ± 0.002</b>	<b>0.007</b>	<b>0.039</b>	<b>0.157</b>

<sup>a</sup>Difference in dipole moment magnitudes between original and renormalized charges

<sup>b</sup>Angle between dipole vectors