

# Are Alchemical Free Energy Calculations Reproducible? (draft)

Hannes H. Loeffler,<sup>\*,†</sup> Stefano Bosisio,<sup>‡</sup> Julien Michel,<sup>‡</sup> Guilherme Duarte Ramos  
Matos,<sup>¶</sup> David L. Mobley,<sup>§</sup> Donghyuk Suh,<sup>||</sup> and Benoit Roux<sup>||</sup>

<sup>†</sup>*Science & Technology Facilities Council, Daresbury, Warrington, WA4 4AD, United  
Kingdom*

<sup>‡</sup>*EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road,  
Edinburgh EH9 3FJ, UK*

<sup>¶</sup>*Department of Chemistry, University of California, Irvine*

<sup>§</sup>*Departments of Pharmaceutical Sciences and Chemistry, University of California, Irvine*  
<sup>||</sup>*University of Chicago*

E-mail: Hannes.Loeffler@stfc.ac.uk

Phone: +44 1925 603367

## Abstract

Alchemical free energy calculations are an increasingly important modern simulation technique. Contemporary Molecular Dynamics and Monte Carlo software such as AMBER, CHARMM, GROMACS and SOMD include support for the method. Implementation details vary among those codes but users expect reliability and reproducibility, i.e. a simulation must yield a comparable free energy within statistical bounds regardless of code used. *Relative* alchemical free energy simulation has been less well tested than its absolute counterpart. However, relative transformations are thought to be computationally and statistically more efficient. Thus, reproducibility of computed

relative free energies across simulation packages is crucial. Here we present the results for relative alchemical free energy (RAFE) calculations for hydration free energies of a set of small organic molecules and show that free energies can be satisfactorily reproduced with aforementioned codes, but this requires care and attention to detail. The benchmarks and protocols reported here may be used to validate new and future versions of free energy calculations software.

1.Unify language!

2.Objectives: Do the relative free energy results produced across different softwares agree with each other?

What if they don't?

3.Objectives: Can we outline what should be a standard procedure to calculate relative free energies of hydration alchemically?

4.Objectives: We want to emphasize that we aim at improving codes/protocols/practices and not highlight "bad" codes

5.Be clear on what reproducibility means here e.g. no exact numerical

## 1 Introduction

The free energy is a fundamental function of thermodynamics and kinetics as it explains how processes in nature evolve. The equilibrium balance of products and reactants in a hypothetical chemical reaction can be immediately determined from the knowledge of the free energy difference of reactants and products and their concentrations. The free energy landscape of a given system, however, can be very complicated and rugged with barriers which impose limits on how fast the process can take place. It is therefore of little surprise that the determination of free energy changes is of utmost importance to all natural sciences e.g. for binding and molecular association, solvation and solubility, protein folding and stability, partition and transfer, and design and improvement of force fields.

The calculation of free energies via molecular simulations<sup>1-5</sup> has been particularly attrac-

tive as it promises to circumvent certain limitations of experimental approaches. Specifically, processes can be understood at the molecular and atomic level and there is the potential that computational techniques can be more cost and time effective. Thus, a multitude of methods have been devised to make reversible work estimates accessible through computation.<sup>1-5</sup> However, the reliability of estimates is still very much a matter of concern.<sup>2,6</sup> Roughly speaking, fast methods tend to be less accurate while more accurate methods tend to be slow. Here we are interested in alchemical free energy methods because they are firmly rooted in statistical thermodynamics and thus give asymptotically correct free energy estimates i.e. they are correct in the limit of sufficient simulation time.

One such method is the so-called *alchemical* free energy approach which is argued to be the most accurate method in quantitative prediction of free energies.<sup>1,7-9</sup> The method has been applied in various forms for many decades now since the early days of computer simulation.<sup>10-15</sup> The method has gained renewed attention in recent years — concomitant with improvements in computer hardware design — both within the traditional equilibrium framework<sup>16-18</sup> but also increasingly in combination with non-equilibrium techniques.<sup>19-21</sup> The name comes from the nonphysical intermediates that often need to be created to smoothly interpolate between end states and because parts or all of a molecule may “appear” or “disappear” in a transformation. In the context of force field methods the transformation takes place in “parameter space”, i.e. the various force field parameters are varied by scaling. This can be a particular efficient approach as it does not require translocation in configuration space. For instance, the dissociation of a ligand from a large biomolecule may involve many degrees of freedom while, at the same time, it is generally unclear along which coordinates a translocation simulation should take place.

Alchemical free energy simulations rely on the concept of thermodynamic cycles.<sup>14</sup> As the free energy is a state function, the sum of free energy changes computed around any closed cycle must be zero. This also implies that the reversible work can be computed arbitrarily along conveniently chosen legs of the cycle. E.g. in Fig. 1 the relative free energy

of hydration can be computed along the vertical legs, that is, following the physical process of moving a molecule from the gas phase to the liquid phase, or along the horizontal legs in an nonphysical but computationally more efficient alchemical calculation.

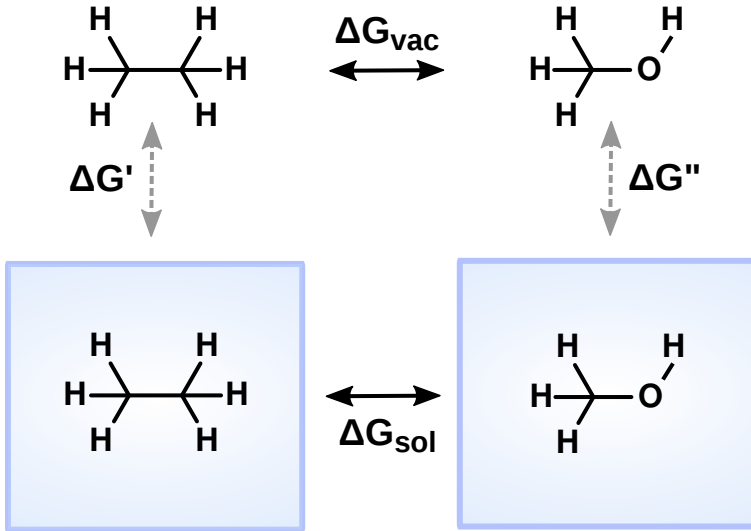


Figure 1: The thermodynamic cycle to compute the relative free energy of hydration  $\Delta\Delta G_{\text{hydr}} = \Delta G_{\text{sol}} - \Delta G_{\text{vac}} = \Delta G'' - \Delta G'$ . The example is for the ethanol  $\leftrightarrow$  methanol transformation. Alchemical simulations are performed along the non-physical horizontal legs while vertical legs illustrate the physical process of moving a molecule from the vacuum to the solution. The latter is also accessible through absolute alchemical free energy simulation, see e.g. Ref. 22.

Absolute (standard) alchemical free energy calculation has been a particular focus of recent years.<sup>16–19,21</sup> *Absolute* here really means that the equilibrium constant of a physical reaction, e.g. binding and dissociation, can be calculated directly by completely decoupling or annihilating a whole molecule from its environment and the term is mostly being used to discriminate against techniques usually referred to as *relative* (see below). We emphasize that the “absolute” approach will still result in a *relative* free energy between the solute fully interacting with its environment and one where it does not. Decoupling means the scaling of the non-bonded *inter*-molecular interactions between the perturbed group (all atoms that differ in at least one force field parameter between the end states) and its environment. We distinguish this from “annihilation” which also scales the *intra*-molecular interactions in addition to the inter-molecular interactions. These schemes may require two simulations

along the opposite edges of a quadrilateral thermodynamic cycle but approaches that produce the reversible work directly in one simulation have been proposed too.<sup>23,24</sup>

Relative alchemical free energy (RAFE) calculations “mutate” one molecule into another one. RAFEs are useful for instance in ranking which one of a set of molecules binds strongest to a chosen target. This approach has recently gained increased traction in the context of relative free binding energies between small molecules e.g. drug or lead like molecules and biomolecules.<sup>25,26</sup>

RAFEs are readily implemented<sup>27,28</sup> by making use of the single topology method.<sup>15,27,29</sup> Single topology means that there is only one representation of the molecule to be mutated, also implying a single set of coordinates. Thus, atom types are directly transformed into the new type, typically by linearly scaling the force field parameters. In typical implementations, disappearing/appearing atoms need to be balanced with “dummy” atoms to ensure constant number of atoms in both end states. Dummy atoms have no non-bonded interactions in the end state but retain the bonded terms of the original atom to avoid complications with unbound atoms<sup>29</sup> (see also “wandering” ligand problem below). However this need not be. For instance the AMBER implementation is a special case as it does not require the user to define dummy atoms explicitly, i.e. as atoms with coordinates and zero end state parameters in the topology, but only to mark the disappearing and appearing atoms in the control file of the MD engine. The contributions of bonded terms that involve at least one dummy atom will not be factored into the free energy as it is assumed that those contributions will perfectly cancel in the thermodynamic cycle.<sup>30,31</sup>

The single topology approach<sup>27</sup> requires topological and structural similarity. But also chemical similarity is of importance e.g. chirality and binding modes where the relative three dimensional arrangement in space must be taken into account. Furthermore, ring breaking is technically challenging<sup>26</sup> but it has also been shown that this can be done in certain circumstances.<sup>30,31</sup>

When the two molecules are structurally dissimilar or do not spatially overlap, the dual

topology method<sup>27,29</sup> can be applied to compute relative free energies. In this approach *all* atoms of the end states are duplicated and thus both sets are present at all times but don’t interact with each other. Only non-bonded interactions need to be scaled such that the disappearing end state corresponds to an ideal gas molecule.<sup>29</sup> This, however, comes with additional complications as two independent molecules can drift apart and so suffer from the “wandering” ligand problem as in absolute transformations.<sup>16–18</sup> Technically, a dual topology calculation is the same as two absolute calculations run simultaneously in opposite directions. It has been shown though that with the introduction of special restraints or constraints this can be a viable option.<sup>32–34</sup> Restraints between corresponding atoms, thus emulating the single topology approach, can also be used without affecting the free energy.<sup>34</sup> A recent alternative<sup>6</sup> considered molecules with a common core where all atom types are the same and the corresponding atom charges were made equal. This means that the core does not need to be duplicated. But all the chemistry is now exclusively handled by, possibly, just a few non-core atoms such that this approach may only be of limited use. A covalent link, e.g. as in side-chain mutation simulations, provides a natural restraint such that dual topology simulations can be applied without further problems. Modern MD software e.g. AMBER,<sup>35</sup> CHARMM,<sup>36</sup> GROMACS,<sup>37</sup> GROMOS<sup>38</sup> and SOMD<sup>39,40</sup> offer a hybrid single/dual topology approach i.e. the user can specify which part of a perturbed group should be handled by which method.<sup>31</sup>

As alluded to above, reliability is a principal matter of concern. In particular, we need to ensure reproducibility of free energy results among computer codes. To the best of our knowledge this has not been systematically tested yet for a set of different MD packages. However, there have been some recent efforts to test *energy* reproducibility across packages<sup>41</sup> — a necessary but not sufficient prerequisite. Given a predefined force field and run-time parameters we should be able to obtain comparable free energy results within statistical convergence limits. In practice, we have the problem, however, that the methods and algorithms used in one MD program are not always present in another package or are the same,

such as algorithms for pressure and temperature scaling, integrators, etc. Nevertheless, it is critical that free energy changes computed with different simulation software should be reproducible within statistical error, as this otherwise limits the transferability of potential energy functions, and the relevance of properties computed from a molecular simulation. This is especially important as the community increasingly combines or swaps different simulation packages within workflows aimed at addressing challenging scientific problems.<sup>42–46</sup>

In this work we present the results of relative hydration free energies of a set of small organic molecules (see Fig. 2). Solvation free energies have a wide range of uses and various methods exist to compute them.<sup>47</sup> They are also needed to calculate binding free energies where the simulation in solution (see Fig. 1) is combined with a mutation of the molecule bound to a partner, and other important physical properties.<sup>47</sup> A large database of hydration free energies computed from alchemical free energy (AFE) simulation, FreeSolv, has been presented recently<sup>48</sup> and was just updated.<sup>22</sup> Here, we focus on the reproducibility of RAFE with the simulation programs AMBER, CHARMM, GROMACS and SOMD. We will discuss the reversible work results obtained with these packages and make recommendations regarding simulation protocols, setup procedures and analysis techniques. We will also deliberate on what needs to be done to progress the field, both from a usability perspective as well as from the view point of code development.

## 2 Methods

We will compare relative free energies calculated via simultaneous parameter change simulations — our “unified protocol” — and separated parameter change simulations — our “split protocol”— to relative free energy calculations obtained from two absolute hydration free energies – our “absolute protocol”. The unified protocol changes partial charges, van der Waals parameters, and bond parameters simultaneously along the alchemical path, while the split protocol separates the transformation of the van der Waals parameters from the

charge transformation. The order in which this has to be done is detailed in section 2 of the SI. The scaling of the bonded terms can be combined with either transformation.

## 2.1 Alchemical Free Energy Implementations

We begin by working out the differences in the alchemical free energy implementations of the four MD codes AMBER, CHARMM, GROMACS and SOMD. One key difference is in the softcore functions<sup>49,50</sup> implemented in each code as summarised in section 1 of the SI. Softcore functions are used to avoid numerical and thus stability problems of the conventional van der Waals and Coulombic potentials<sup>51</sup> as they have singularities at zero distance (vertical asymptotes). Direct scaling of these potentials causes the functions to increasingly behave like hard-sphere potentials as  $\lambda \rightarrow 0$ . This implies a higher probability of other atoms to penetrate into the highly repulsive short-range portion of the potential which can lead to strongly fluctuating forces/energies and to severe instabilities in the integrator, as well as problems with analysis even when simulations do terminate normally.<sup>49–51</sup>

Another difference is how the codes scale force field parameters (“parameter scaling”) and/or the energy (“energy scaling”).<sup>29</sup> In the former case each parameter is scaled individually e.g. in the case of a harmonic bond or angle term this is the force constant and the equilibrium distance/angle while in the latter case the total energy term resulting from the potential is scaled at once. The two approaches are not mathematically equivalent and so the pathway is different. It should be noted, however, that the non-bonded interactions are handled through softcore functions in modern codes and thus parameter scaling<sup>27,52</sup> is now only used for bonded terms in some codes.

One more important question is if the code allows constraints for bonds with changing bond lengths. These and other details will be outlined below. The perturbed group consists of all atoms that need to be transformed, i.e. any atom that differs in at least one force field parameter in the other end state. In our case this means the whole solute.



**AMBER** This code is strictly dual topology and all terms are energy-scaled. The code allows, however, mapping of atoms in a single topology fashion and computes these non-softcore atoms by linearly scaling the energy and forces for each atom in the pair. The perturbed group must be entirely duplicated i.e. for **sander** this means two topology files with one end state each, and for **pmemd** both end states in one topology file. The softcore potential applies to any atom chosen by the user which may include atoms that actually have an equivalent in the other state. Non-softcore atoms must still match 1:1 in both states. Explicit dummy atoms are not needed as the code will only compute bonded contributions for “real” atoms and thus ignores bonded energies involving dummy atoms. We will call this the “implicit dummy protocol”. The code cannot handle bond length changes involving a constraint. There is only one global  $\lambda$  for parameter transformation. Protocols that couple only some parameters (separated protocols, see below) must be emulated through careful construction of topologies by keeping force field parameters that are not being modified constant along  $\lambda$ .

**CHARMM** The PERT module duplicates the topology similar to **sander** but mapped atoms are given in the topology only once. The module requires balancing with explicit dummy atoms. All terms are energy-scaled. The PSSP softcore potential is applied to *all* atoms in the perturbed group (see section 1 in the SI). The code can handle constraints of changing bond lengths in the perturbed group but this may cause incorrect results with PSSP softcores (Stefan Boresch, private communication). There is only one global  $\lambda$  for parameter transformation, however, the scripting facilities in CHARMM allow run time modification of topologies e.g. by setting charges or vdW parameters to arbitrary values.

**GROMACS** This code uses a single topology description. Bonded terms are strictly parameter-scaled, which requires proper balancing of multi-term dihedrals i.e. each individual term in the Fourier series must have an equivalent in both end states. If the term does not exist it must be created with parameters zeroing its energy. The softcore potential applies to

dummy atoms only determined from atoms having zero vdW parameters in the end states. The code allows changing bond lengths involving constraints within the perturbed group but this can lead to instabilities and wrong results (Michael Shirts, private communication). There are separate  $\lambda$ s for vdW, Coulomb and bonded parameters (and some other) which allows easy implementation of separated protocols.

**Sire/SOMD** This code uses a single topology description. The final state is constructed at run time from the initial state with a “patch” (list of force field parameters to be modified). All dummy atoms needed to describe the transformation must be present in the initial state. Bond and angle terms are parameter-scaled while the dihedral term is energy-scaled. The softcore potential applies to atoms that become dummy atoms in one end-state. Dummy atoms are specified by a keyword in the patch file. The code cannot handle constraints of changing bond lengths in the perturbed group. There is only one global  $\lambda$  for parameter scaling. Separated protocols (see below) must be emulated through careful construction of the patch file. SOMD is software built by linking of the Sire<sup>39</sup> and OpenMM<sup>40</sup> molecular simulation libraries.

## 2.2 RAFF Setup

The setup for all relative free energy simulations has been carried out with the tool FESetup<sup>45</sup> in version 1.2. FESetup is a perturbed topology writer for AMBER, CHARMM, GROMACS, SOMD and also NAMD<sup>53</sup> (within the limits of the dual topology approach). The tool makes use of a maximum common substructure search algorithm to automatically compute atoms that can be mapped i.e. atoms that have a direct relationship to an equivalent atom in the other state which means atom type to atom type conversion. The only current limit is that rings are required to be preserved.<sup>30</sup> With this strategy a maximal single topology description is achieved: any atom that does not match will be made a dummy atom. FESetup allows equilibration of the solvated simulation systems and ensures that “forward” and “backward”

simulations have the same amount of total atoms. The masses of perturbed hydrogen atoms are set to the mass of the heavy atom for SOMD. The other codes use the atom masses of the initial state (AMBER, CHARMM?) or allow the user to define how masses vary as a function of lambda (GROMACS). The tool creates all input files with control parameters, topologies and coordinates as required for RAFe simulations. Full details on FESetup can be found in Ref. 45.

Figure 2 shows all 18 transformation considered in this study including “forward” and “backward” mutations. RAFe simulations do not have a directionality with respect to the coupling (order) parameter  $\lambda$ . But to test for possible discrepancies we have run simulations in both directions. As we shall discuss in the Results section we do see differences in some cases.

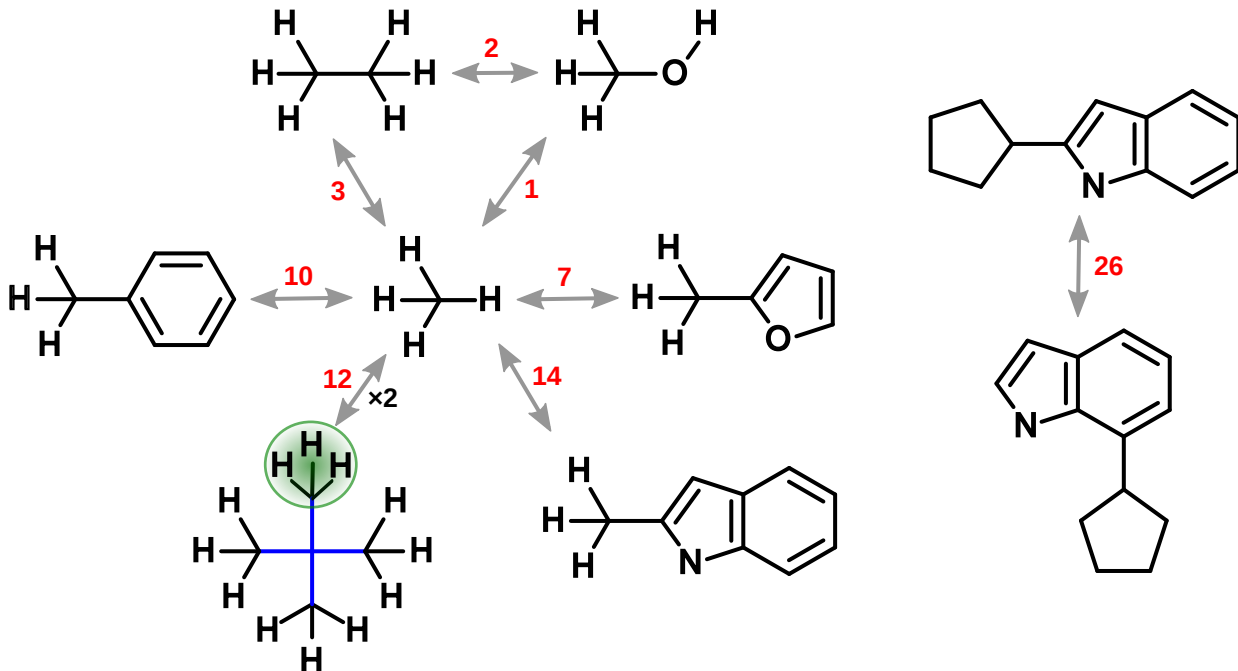


Figure 2: The thermodynamic cycles considered in this study. To compute the free energy of hydration all pair-wise transformations have to be carried out once in solution and once in vacuum. Green and blue colours in neopentane show two alternative mappings for methane. The numbers in red denote the number of dummy atoms.

The ethane  $\rightarrow$  methanol transformation is traditionally regarded as a standard test for RAFe simulations.<sup>15,54</sup> The other transformations are centered around mutations from and

to methane, and are meant to mimic typical transformations that could be attempted in the context of e.g. protein-ligand binding calculations. The 2-cyclopentanylindole to 7-cyclopentanylindole transformation has been added to include both deletion as well as insertion of sub-parts of the perturbed group in one simulation (in the following CPI will stand for cyclopentanylindole). For neopentane  $\rightarrow$  methane we point out that there are two alternative mappings possible, see Figure 2. One in which methane is mapped with a terminal methyl (green) and the other one where the methane carbon is mapped with the central carbon in neopentane (blue). The first approach will be called “terminally mapped” and the second one “centrally mapped”.

### 2.3 RAFF Simulation Protocols

One of the major concerns in a reproducibility study is to ensure consistency in the applied protocols. This is complicated by the fact that a given MD software may employ a wide range of methods and algorithms that may not be available in other MD software. For example, pressure and temperature scaling, integrators and other algorithms can be very different. It is also unclear if and how implementation details can affect results, in particular see subsection 2.1 discussing the implementation details of alchemical free energy simulation in code.

In this study we look at a set of simple organic molecules (see Figure 2). As the focus here is on probing for reproducibility among various MD packages, we chose fairly small, rigid and neutral molecules to keep problems with sampling low, and avoid difficulties with charged particles.<sup>55,56</sup> The force field was chosen to be GAFF<sup>57</sup> (version 1.8), utilising AM1/BCC charges<sup>58,59</sup> for the solute and TIP3P<sup>60</sup> for the solvent. Charges were computed with the `antechamber` program and missing bonded and vdW terms were generated with the `parmchk2` program, both from the AmberTools16 distribution. The quality of free energies of various small molecule force fields has been shown elsewhere, see e.g. Refs. 61,62.

While the MD packages principally allow a “one-step” transformation<sup>63</sup> that is both van der Waals and Coulombic softcore potentials vary simultaneously, it can be more efficient to carry out a separated protocol.<sup>64,65</sup> In such a protocol the charges are transformed linearly between the end states followed by a mutation of the van der Waals parameters using a softcore potential<sup>49,50</sup> (see section 1 in the SI for details) on the vdW term only. It is important to note that in the separated protocol charges have to be switched off before vdW parameters (and vice versa for the transformation in opposite direction) to avoid collapse of other atoms, e.g. solvents, onto a “naked” charge, see section 2 in the SI.

All simulations were started from the simulation box which has been created with FES-*etup*.<sup>45</sup> It should be noted, however, that in constructing the system steric overlaps between the solute and the solvent may happen. This is because each unperturbed solute is independently equilibrated and the perturbed system combined from those two. The number of atoms are always chosen to be the same for forward and backward setups by using the larger box of the two unperturbed systems. Thus, in transformation from a smaller to a larger solute water molecules may be in close distance to the solute. The production simulations were run at 298.15 K and 1.0 bar.

**AMBER** The starting coordinates were usually taken directly from the pre-equilibrated setup step but no further  $\lambda$  specific equilibration was carried out, i.e. RAFE MD simulations were started with new velocities appropriate for the final simulation temperature. In a very few cases it was necessary to use coordinates from a nearby  $\lambda$  state because of simulation instabilities. This happened in transformations with a larger number of dummy atoms. Water hydrogens (TIP3P) were constrained with SHAKE but none of the atoms in the perturbed group and so the time step was set to 1 fs (see SOMD protocol below for an alternative approach). The temperature was controlled through a Langevin thermostat with a friction constant of  $2.0 \text{ ps}^{-1}$  and pressure rescaling through a Monte Carlo barostat with 100 steps between isotropic volume change attempts. Long-range electrostatics in solution

was handled with PME and a cutoff of 8.0 Å for both Coulomb and vdW interactions. In vacuum no cutoff was used.

**CHARMM** The CHARMM c40b1 molecular dynamics package was used to carry out all relative hydration free energy simulations. 21 evenly spaced windows were used and all windows were run for 1.5 ns with a timestep of 1 fs. Most windows used the same pre-equilibrated configuration, but a few windows were started with their own equilibration. Constant temperature and pressure control was done with the Berendsen weak coupling method, with a compressibility of  $4.63 \times 10^{-5} \text{ atm}^{-1}$  and temperature and pressure coupling constants of  $5.0 \text{ ps}^{-1}$ . A 12.0 Å atom-based cutoff was used for non-bonded interactions and Particle Mesh Eward (PME) with order of 6 was used to calculate electrostatic forces along with periodic boundary condition. SHAKE was used to constrain water hydrogens only. The PERT module in CHARMM was used to handle alchemical transformation in conjunction and PSSP softcore potential functions were used to avoid numerical issue at the end point lambdas.

**GROMACS** We used GROMACS 4.6.7 to carry out simulations for the relative hydration free energies ( $\Delta\Delta G_{\text{hydr}}$ ). Each transformation had its Gibbs free energy calculated: (i) in a single topology approach in which van der Waals energy terms were changed separately from the electrostatic and bonded components; (ii) in a single topology approach in which bonded, van der Waals, and electrostatic terms are changed together; and (iii) via the difference between two absolute calculations. In the first two cases, each alchemical transformation was described by 31 and 16 states, respectively, and simulated for 4.2 ns with time steps of 1.0 fs in water and vacuum. Atomic masses were not changed along the alchemical path; their change affect kinetic energy only and do not contribute to the free energy change on this case. We used the Langevin integrator implemented in GROMACS with a default friction coefficient of  $1.0 \text{ ps}/m_{\text{atom}}$ . Absolute hydration free energies were calculated from 5 ns Langevin dynamics at 298.15 K. We used the 20-step alchemical protocol where charge

coupling and van der Waals coupling were dealt with separately along the path.<sup>22,48</sup> A Parrinello–Rahman barostat with  $\tau_p = 10$  ps and compressibility equal to  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. We used two methods to calculate electrostatic interactions: Particle Mesh Ewald (PME) and charge group-based Reaction Field with a dielectric of 78.3, as implemented in the software. We set the electrostatic and non-bonded cutoffs to 10.0 Å; a switch was applied to the latter starting at 9.0 Å. PME calculations were of order 6 and had a tolerance of  $1.0 \times 10^{-6}$ , with a grid spacing of 1.0 Å. All transformations required the use of softcore potentials to avoid numerical problems in the free energy calculation. We chose the 1–1–6 softcore potential for Lennard-Jones terms and used the default softcore Coulomb implementation in paths where charges, van der Waals, and bonded terms were modified together, but no soft core potentials were applied to Coulomb interactions when Coulomb interactions were modified separately.

**SOMD** All simulations were carried out with SOMD from the Sire 2016.1 release<sup>39,40</sup> Each alchemical transformation was divided into 17 evenly spaced windows and simulated for 2 ns both in water and in vacuum. A velocity-Verlet integrator was employed with a 2 fs time step, constraining only those hydrogen bonds which are not alchemically transformed. This approach will be called the “unperturbed protocol”. Referring to eq. S1 the softcore parameters were set to default values for all the transformations, specifically  $n = 0$  for Coulombic interactions and  $\alpha = 2.0$  for the Lennard-Jones potential.<sup>32</sup> A shifted atom-based Barker–Watts reaction field<sup>66</sup> with a dielectric constant of 78.3 was adopted for the solution phase simulations with a cutoff of 10 Å. A similar cutoff was used for Lennard-Jones interactions. Temperature control was achieved with the Andersen thermostat<sup>67</sup> with a coupling constant of 10 ps<sup>-1</sup>. A Monte Carlo barostat assured pressure control, attempting isotropic box edge scaling every 25 time steps. The reaction field was not employed in the vacuum legs, where a Coulombic potential without cutoff was used. This leads to an inconsistent description of the intramolecular electrostatic interactions of the solute in the solvated and vacuum phases.

To maintain a consistent description of intramolecular energetics across vacuum and water legs, a free energy correction term  $G_c$  was evaluated as detailed in Ref. 68). The  $G_c$  term was obtained via post-processing of the end state trajectories of each water phase simulation, using the Zwanzig relationship:<sup>69</sup>

$$G_c = -\beta^{-1} \ln \langle \exp [-\beta(U_{ic,nc}(\mathbf{r}) - U_{ic,sim}(\mathbf{r}))] \rangle_{sim} \quad (1)$$

where  $U_{ic,nc}(\mathbf{r})$  is the solute intramolecular electrostatic-no cutoff potential that depends on the coordinates  $\mathbf{r}$  of the solute and is given by Coulomb’s law computed without cutoffs.  $U_{ic,sim}(\mathbf{r})$  is the intramolecular electrostatic potential term as computed in the simulation with the shifted atom-based Barker-Watts Reaction Field cutoff.

To compute absolute free energy of hydration a double annihilation technique<sup>16,68,70</sup> was adopted. Initially the solute atoms’ partial charges are turned off both in water and in vacuum, discharging step, giving a discharging free energy change  $\Delta G_{solv}^{elec}$  and  $\Delta G_{vac}^{elec}$  respectively. Secondly, the molecule is fully decoupled from the surrounding environment by switching off the van der Waals terms, vanishing step, retrieving a  $\Delta G_{solv}^{vdW}$  in water phase and  $\Delta G_{vac}^{vdW}$  in vacuum. Eventually, the absolute free energy of hydration  $\Delta G$  can be computed as:

$$\Delta G = (\Delta G_{solv}^{elec} + \Delta G_{solv}^{vdW}) - (\Delta G_{vac}^{elec} + \Delta G_{vac}^{vdW}) \quad (2)$$

The same setup for relative free energy calculations was followed to compute absolute free energy of hydration. Thus, 17 evenly spaced  $\lambda$  windows were used and simulations were run for 2 ns employing a velocity-Verlet integrator with 2 fs time step, constraining only the hydrogen bonds which are not alchemically transformed.

## 2.4 Analysis

Various estimators have been proposed to obtain the free energy from AFE simulations. Early work by Kirkwood<sup>71</sup> expresses the free energy as an ensemble average of the derivative



of the Hamiltonian with respect to the coupling parameter  $\lambda$ . The method is now known as Thermodynamic Integration (TI). Zwanzig devised the exponential formula<sup>69</sup> (EXP), also known as Free Energy Perturbation (FEP) or thermodynamic perturbation (TP), which calculates the free energy from the exponential average of the energy difference of the end states. The energy difference is computed from the energy with the configuration of one end state and reusing it to compute the energy with the force field parameters of the other end state. This obviously assumes that the configuration is representative of both state. As the phase-space overlap needs to be sufficiently large<sup>72,73</sup> the EXP approach typically needs intermediates, controlled by  $\lambda$ . However, it has been shown that EXP has an asymmetric bias depending on the directionality of  $\lambda$ <sup>74</sup> and that the Bennett Acceptance Ratio (BAR) method<sup>11</sup> is considerably more effective in obtaining an accurate result.<sup>75</sup> BAR is a generalization of EXP by making explicit use of the “forward” ( $\lambda_i \rightarrow \lambda_{i+1}$ ) and “backward” ( $\lambda_{i+1} \rightarrow \lambda_i$ ) estimates. Later it was demonstrated that this can be effectively extended to incorporate all other  $\lambda$ s. This approach has been called multi-state BAR (MBAR).<sup>76</sup> MBAR has been shown to have the lowest variance of any known estimator.<sup>77</sup>

In this work we primarily focus on TI as this is supported by all MD packages “out-of-the-box”, whereas BAR and MBAR are not. Equation 3 computes the free energy as

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\mathcal{H}(\mathbf{q}, \mathbf{p}; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (3)$$

where  $\mathcal{H}(\mathbf{q}, \mathbf{p}; \lambda)$  is the Hamiltonian as a function of the coordinate vectors  $\mathbf{q}$  and the momentum vectors  $\mathbf{p}$ , and parametric dependence on the coupling parameter  $\lambda$ . The angle brackets denote the ensemble average of the gradient of the Hamiltonian with respect to  $\lambda$  at a given  $\lambda$  value. An AFE simulation is typically carried out in a series of equilibrium simulations at discrete values of  $\lambda$  but the gradient can also be evaluated with a continuously varying coupling parameter as a function of the simulation time. The free energy is finally computed through a suitable numerical integration method.

Results from additional estimators will be given where available. We have used the alchemical analysis tool<sup>78</sup> for analysis of all free energies. This tool provides various estimators such as TI, TI with cubic splines, BAR and MBAR. All data can be sub-sampled to eliminate correlated data.

All RAFF simulations were run in triplicate in forward as well as backward direction for a total of 6 simulations per mutation. The final hydration free energy  $\Delta\Delta G_{\text{hydr}}$  was computed as the average for each direction separately. For comparison we have also calculated the absolute (standard) hydration free energies for all molecules in Figure 2. These energies are less implementation-dependent and thus provide an additional check.

To estimate the reliability and convergence of the results, the standard error of the mean (SEM) has been calculated. The SEM is defined as

$$\text{err}(\Delta\Delta G_{\text{hydr}}) = \frac{\sigma}{\sqrt{n}} \quad (4)$$

where  $\sigma$  is the sample standard deviation and  $n$  is the size of the uncorrelated sample. The SEM for component free energies is combined as

$$\text{err}(\text{combined}) = \sqrt{\sum_i \sigma_i^2}. \quad (5)$$

which is appropriate if the property to be computed is a sum of contributions.

We also make use of the mean absolute error MAE (also called mean unsigned error, MUE) to compare data sets.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (6)$$

where  $n$  is the total number of samples,  $y_i$  and  $x_i$  are the  $i$ -th datum to be compared.

### 3 Results

We will use absolute hydration free energies here as our standard point of comparison because these are easily calculated with considerable precision<sup>22</sup> and are considerably simpler to set up and implement than relative calculations. Prior work actually compared calculated absolute hydration free energies across different codes with considerable success<sup>79</sup> (see Fig. 7 in the reference), further supporting this view.

Table 1 summarizes results for relative free energies of hydration obtained from absolute transformations which are summarized in Table S1. The table shows the data from simulations with the recommended protocol as discussed in detail in the following subsections for each MD code. The  $\Delta\Delta G_{\text{hydr}}$  obtained with the various MD packages in this way agree quite well, although some larger deviations are apparent as well. GROMACS predicts a smaller  $\Delta\Delta G_{\text{hydr}}$  for the methanol transformations by about 0.2 kcal mol<sup>-1</sup>. Similarly, there are some discrepancies in the toluene, 2-methylfuran and 2-methylindole cases where it is mostly CHARMM that produces somewhat smaller  $\Delta\Delta G_{\text{hydr}}$  with 0.4 kcal mol<sup>-1</sup> difference between the smallest and highest values. The largest deviation can be found in the largest transformation where the AMBER result is higher by 0.4–0.5 kcal mol<sup>-1</sup> than the other MD packages.

Table 1: Comparing relative free energies of hydration for various MD packages as obtained from absolute transformations.

transformation		AMBER <sup>a</sup>	CHARMM <sup>a</sup>	GROMACS <sup>b</sup>	SOMD <sup>a</sup>
ethane	methane	$-0.02 \pm 0.01$	$-0.08 \pm 0.01$	$-0.04 \pm 0.01$	$-0.05 \pm 0.02$
methanol	methane	$6.20 \pm 0.01$	$6.15 \pm 0.02$	$5.95 \pm 0.01$	$6.21 \pm 0.06$
ethane	methanol	$-6.22 \pm 0.01$	$-6.23 \pm 0.02$	$-5.98 \pm 0.01$	$-6.26 \pm 0.05$
toluene	methane	$3.19 \pm 0.01$	$3.00 \pm 0.01$	$3.16 \pm 0.01$	$3.07 \pm 0.03$
neopentane	methane	$-0.13 \pm 0.02$	$-0.18 \pm 0.02$	$-0.14 \pm 0.01$	$-0.19 \pm 0.06$
2-methylfuran	methane	$2.96 \pm 0.02$	$2.79 \pm 0.01$	$2.95 \pm 0.01$	$2.90 \pm 0.03$
2-methylindole	methane	$8.72 \pm 0.01$	$8.36 \pm 0.02$	$8.79 \pm 0.02$	$8.57 \pm 0.03$
2-CPI	7-CPI	$0.39 \pm 0.04$	$-0.10 \pm 0.04$	$0.02 \pm 0.05$	$0.08 \pm 0.14$

<sup>a</sup>unified protocol.

<sup>b</sup>split protocol with PME.

Tab 2 shows the MAE between SOMD, GROMACS, AMBER and CHARMM. The MAE appears to be fairly small generally and reaches just about  $0.2 \text{ kcal mol}^{-1}$  between CHARMM and GROMACS, and CHARMM and AMBER.

Table 2: MAE ( $\text{kcal mol}^{-1}$ ) comparing relative free energies from absolute simulations between SOMD, GROMACS, AMBER and CHARMM.

Package	GROMACS	AMBER	CHARMM
SOMD	0.12	0.09	0.08
GROMACS		0.13	0.17
AMBER			0.16

Having established the predictions from absolute transformations we now turn to computing  $\Delta\Delta G_{\text{hydr}}$  from relative mutations. Table 3 summarizes the results for the four MD packages. Again the data is from the recommended protocol (see detailed discussions in the following subsections).

Overall, the predicted  $\Delta\Delta G_{\text{hydr}}$  valued from the four MD packages are in good agreement with each other. CHARMM tends to show smaller relative free energies in a number of transformations: toluene, neopentane, 2-methylfuran and 2-methylindole. SOMD displays smaller  $\Delta\Delta G_{\text{hydr}}$  in the methanol and toluene transformations. The largest discrepancy, however, is in the neopentane transformation with central mapping where AMBER with implicit dummy atoms is about  $0.5 \text{ kcal mol}^{-1}$  higher and CHARMM about  $0.3 \text{ kcal mol}^{-1}$  lower than the other two codes. The terminal mapped neopentane reveals AMBER to be in line with GROMACS and SOMD while CHARMM’s results is essentially the same as in the central mapped transformation. AMBER deviates also quite strongly from the other codes in the cyclopentanyl indole cases.

The MAEs of the relative free energy simulations are presented in Table 4. These are continuously larger than the MAEs from the absolute simulations (Table 2) and reach  $0.3 \text{ kcal mol}^{-1}$  for CHARMM compared with AMBER and GROMACS.

Table 3: Comparing relative free energies of hydration for various MD packages as obtained from relative transformations. Signs of the backward transformation have been reverted to correspond to the forward transformation.

		AMBER <sup>a</sup>		CHARMM <sup>b</sup>	GROMACS <sup>a</sup>	SOMD <sup>b</sup>
		implicit	explicit			
ethane	methane	$0.02 \pm 0.01$	$-0.13 \pm 0.02$	$-0.11 \pm 0.02$	$-0.04 \pm 0.02$	$-0.01 \pm 0.05$
methane	ethane	$0.00 \pm 0.03$	$-0.19 \pm 0.03$	$-0.06 \pm 0.01$	$-0.02 \pm 0.01$	$-0.04 \pm 0.02$
methanol	methane	$6.19 \pm 0.01$	$6.20 \pm 0.02$	$6.17 \pm 0.01$	$6.20 \pm 0.01$	$5.99 \pm 0.05$
methane	methanol	$6.20 \pm 0.03$	$6.15 \pm 0.01$	$6.20 \pm 0.01$	$6.20 \pm 0.01$	$5.96 \pm 0.04$
ethane	methanol	$-6.20 \pm 0.01$	$-6.27 \pm 0.01$	$-6.26 \pm 0.01$	$-6.19 \pm 0.01$	$-6.09 \pm 0.03$
methanol	ethane	$-6.20 \pm 0.01$	$-6.25 \pm 0.01$	$-6.29 \pm 0.01$	$-6.19 \pm 0.01$	$-6.09 \pm 0.02$
toluene	methane	$3.24 \pm 0.02$	$3.39 \pm 0.02$	$2.93 \pm 0.02$	$3.21 \pm 0.01$	$2.89 \pm 0.09$
methane	toluene	$3.42 \pm 0.03$	$3.52 \pm 0.03$	$2.96 \pm 0.02$	$3.20 \pm 0.01$	$3.06 \pm 0.02$
neopentane <sup>c</sup>	methane	$0.32 \pm 0.04$	$-0.03 \pm 0.06$	$-0.44 \pm 0.01$	$-0.15 \pm 0.02$	$-0.20 \pm 0.05$
methane <sup>c</sup>	neopentane	$0.25 \pm 0.03$	$-0.07 \pm 0.03$	$-0.33 \pm 0.02$	$-0.16 \pm 0.05$	$-0.13 \pm 0.05$
neopentane <sup>d</sup>	methane	$-0.13 \pm 0.01$	$-0.12 \pm 0.02$	$-0.65 \pm 0.02$	$-0.14 \pm 0.01$	$-0.11 \pm 0.01$
methane <sup>d</sup>	neopentane	$-0.13 \pm 0.03$	$-0.12 \pm 0.03$	$-0.49 \pm 0.02$	$-0.18 \pm 0.03$	$-0.10 \pm 0.06$
2-methylfuran	methane	$3.09 \pm 0.01$	$3.10 \pm 0.01$	$2.76 \pm 0.03$	$2.93 \pm 0.05$	$2.92 \pm 0.05$
methane	2-methylfuran	$3.10 \pm 0.03$	$3.15 \pm 0.03$	$2.76 \pm 0.02$	$2.96 \pm 0.01$	$2.83 \pm 0.03$
2-methylindole	methane	$8.78 \pm 0.03$	$8.78 \pm 0.04$	$8.34 \pm 0.01$	$8.73 \pm 0.03$	$8.64 \pm 0.06$
methane	2-methylindole	$9.14 \pm 0.02$	$9.13 \pm 0.03$	$8.42 \pm 0.02$	$8.74 \pm 0.01$	$8.67 \pm 0.08$
2-CPI <sup>e</sup>	7-CPI	$0.36 \pm 0.03$	$0.63 \pm 0.06$	$-0.01 \pm 0.01$	$-0.03 \pm 0.03$	$-0.11 \pm 0.07$
7-CPI <sup>e</sup>	2-CPI	$0.34 \pm 0.05$	$0.50 \pm 0.03$	$0.03 \pm 0.01$	$-0.20 \pm 0.04$	$-0.01 \pm 0.08$

<sup>a</sup>split protocol.

<sup>b</sup>unified protocol.

<sup>c</sup>central mapping.

<sup>d</sup>terminal mapping.

<sup>e</sup>partial re/discharge i.e. only the charges of the appearing and the disappearing 5-rings are switched.

Table 4: MAE (in kcal mol<sup>-1</sup>) comparing relative free energies from relative simulations between SOMD, GROMACS, AMBER and CHARMM.

Package	GROMACS	AMBER	CHARMM
SOMD	0.18	0.23	0.19
GROMACS		0.14	0.31
AMBER			0.32

### 3.1 AMBER

Using AMBER for RAFF simulations has revealed several problems with the implementation. Some bugs were identified and have been fixed for AMBER16 release by the developers, e.g. energy minimization in **sander** led to diverged coordinates for mapped atoms. For a single topology description, however, it is necessary to have the same coordinates. Other issues are that vacuum simulations can only be carried out with the **sander** program because **pmemd** cannot handle AFE simulations in vacuum at the moment. This will, however, be rectified in future versions.<sup>80</sup> A disadvantage of **sander** is that it cannot be used to simulate the  $\lambda$  end points<sup>81</sup> such that the TI gradients need to be extrapolated (minimum and maximum allowed  $\lambda$ s are 0.005 and 0.995). Also, **sander** considers the whole system as the perturbed region while **pmemd** restricts this to a user chosen atom selection. This has obvious implications for performance.<sup>81</sup>

We also found that, in contrast to the other three codes, AMBER cannot correctly reproduce relative free energies in the unified protocol i.e. when all force field parameters are scaled simultaneously (see Table S2). This appears to be a problem when more than a few dummy atoms are involved while the unified protocol works for the smaller transformations (refer to Figure 2). The separated RAFF protocol and absolute free energies, however, are very close to the other MD packages as demonstrated in Table 5.

End point geometries appear to be another issue with AMBER simulations in both solution and vacuum. This is most obvious in the neopentane  $\rightarrow$  methane test case with central mapping (see RAFF Setup and Figure 1). As shown in Figure S3, the methane end state exhibits incorrect distances between the carbon and the four attached hydrogens of approximately 1.23 Å. This value is about 1.12 Å for the terminal dummy atoms in the other test cases but still higher than the expected 1.09 Å on average. Figure S3 demonstrates how this depends on the number of dummy atoms immediately surrounding the central atom.

We also compare free energies obtained from the implicit dummy approach in AMBER with results from explicit dummy atom simulations and results from absolute transforma-

tions, see Tables 1 and 3. The relative simulations have been carried out with the separated protocol while the absolute simulations used a unified protocol throughout. SHAKE was explicitly deactivated for all bonds in the perturbed region in these protocols.

Table 5 shows selected results for transformations with SHAKE enabled for all bonds to hydrogens except those bonds that change bond length during transformation. These

Table 5: Comparing AMBER results for simulations with various protocols. The emphasis is here on the data with SHAKE enabled and a time step of 2 fs (last column). Implicit, explicit and absolute protocols had SHAKE disabled and a time step of 1 fs. Signs of the backward transformation have been reverted to correspond to the forward transformation.

transformation		implicit $\Delta\Delta G$	explicit $\Delta\Delta G$	absolute $\Delta G$	SHAKE <sup>a</sup> $\Delta\Delta G$
ethane	methanol	$-6.20 \pm 0.01$	$-6.27 \pm 0.01$	$-6.22 \pm 0.01$	$-6.20$
methanol	ethane	$-6.20 \pm 0.01$	$-6.25 \pm 0.01$		
toluene	methane	$3.24 \pm 0.02$	$3.39 \pm 0.02$	$3.19 \pm 0.01$	$3.29$
methane	toluene	$3.42 \pm 0.03$	$3.52 \pm 0.03$		
neopentane <sup>b</sup>	methane	$0.32 \pm 0.04$	$-0.03 \pm 0.06$	$-0.13 \pm 0.02$	$0.37$
methane <sup>b</sup>	neopentane	$0.25 \pm 0.03$	$-0.07 \pm 0.03$		
neopentane <sup>c</sup>	methane	$-0.13 \pm 0.01$	$-0.12 \pm 0.02$		
methane <sup>c</sup>	neopentane	$-0.13 \pm 0.03$	$-0.12 \pm 0.03$		

<sup>a</sup>implicit dummy atom protocol with  $\delta t = 2$  fs and SHAKE on all H-bonds except perturbed bonds.

<sup>b</sup>central mapping.

<sup>c</sup>terminal mapping.

free energies are computed from a single run as we generally observe very small SEM and in practice one simulation is sufficient to obtain converged free energy averages for all the systems studied here. The time step has been increased from 1 fs as used in the other three protocols to 2 fs. As the results are essentially the same as the non-SHAKE simulations, this SHAKE protocol appears to be a viable solution to increase the performance of RAFF simulations. We have repeated this protocol with AMBER in response to the good results obtained with SOMD using the same. From a practical point of view, AMBER uses an *atom* based mask for bond SHAKEs such that the mask must be set for the hydrogens in question while the same is not possible for their non-H counter-part in the other state because *all* bonds emanating from this atom would be affected.

We can also compute the cycle closure error from Table 5 for the closed cycle ethane  $\rightarrow$  methanol  $\rightarrow$  methane  $\rightarrow$  ethane (see Figure 2). The free energy difference within a closed thermodynamic cycle must necessarily be zero. For the implicit dummy simulation we calculate a cycle error for  $\Delta\Delta G_{\text{hydr}}$  of  $(0.069 \pm 0.041) \text{ kcal mol}^{-1}$  and for the explicit dummy simulation the error is  $(-0.016 \pm 0.047) \text{ kcal mol}^{-1}$ .

In general, the free energies computed with each approach are in good agreement with each other and with the results of the other MD packages. There are, however, a few notable deviations. Neopentane  $\rightarrow$  methane with central mapping differs from the result with terminal mapping by about  $0.4 \text{ kcal mol}^{-1}$  (cf. Table 5). The terminal mapping and the free energies from the explicit dummy simulations are, however, consistent with the absolute transformations. We also observe a systematic deviation between forward and backward vacuum transformations in particular in the 2-methylindole simulation (see Table S3. A discrepancy of consistently  $0.2\text{--}0.4 \text{ kcal mol}^{-1}$  is evident from every  $\lambda$  step of the vdW plus bonded transformation with both implicit and explicit dummy atoms).

### 3.2 CHARMM

**6.bug in TI gradient accumulation in parallel runs (does not affect serial?, does not affect EXP)**

**7.cannot handle LRC: test with larger cutoffs and/or LRC correction with arbitrary, single structure; check**

**<http://pubs.acs.org/doi/abs/10.1021/jp0735987>**

### 3.3 GROMACS

GROMACS has some run input options which can simplify the procedure for setting up free energy calculations. Specifically, `couple-moltype` implicitly defines the initial and final states by giving a special tag to a molecule and controls whether intramolecular interactions of the tagged molecule are retained or not along the alchemical path. It should be used in absolute free energy calculations to tag the molecule which will be decoupled from the rest of the system. `couple-lambda0` and `couple-lambda1` control the interactions of the



molecule specified by `couple-moltype` with its surroundings. The entries `vdw-lambdas` and `fep-lambdas` define the lambda schedule. The former indicates the value of the  $\lambda$  vector component that modifies van der Waals interactions for each state, while the latter changes all  $\lambda$  vector components that are not specified in the `.mdp` file. For instance, in split protocol simulations, these entries are sets such that the components of the energy are modified in different stages. If the transformation involves particle deletion ("forward process"), `fep-lambdas` is set to change charges and bonds before `vdw-lambdas` changes van de Waals components. If the process involves particle insertion ("backward process") we reverse the roles. In this work, `mass-lambdas` were all set to zero to not included mass changes in the free energy. Unified protocols set all  $\lambda$  vectors the same.

Table 6 lists the relative free energies obtained from GROMACS simulations. Relative free energies are in good agreement with each other and with  $\Delta\Delta G_{\text{hydr}}$  obtained from the other software used in this study (cmp. Tables 1 and 3). A noteworthy exception is the difference between the unified and split results of methane  $\rightarrow$  methanol and its reverse process, which was investigated with additional split protocol simulations using Coulomb softcore potentials (Table 7).

We noticed a difference of approximately  $1.5 \text{ kcal mol}^{-1}$  between the split protocol without Coulomb softcore potentials and both protocols that use it. Data suggests that softcore use in electrostatic interactions requires adjustments in the  $\lambda$ -distance between states in the rapidly varying part of the  $\partial\mathcal{H}/\partial\lambda$ , see Figure S7. A test with combining the bonded terms with the vdW transformation did not change this result. The split protocol without Coulomb softcore potentials is the most effective way to calculate relative free energy calculations. We also find that there is no significant difference in calculated free energies depending on the choice of Particle Mesh Ewald or Reaction Field for calculation of long-range electrostatic interactions.

One particularity of the software worth mentioning is that relative free energy simulations will blow up if a hydrogen alchemically becomes a heavy atom if the simulation employs

Table 6: Relative hydration free energies obtained from GROMACS simulations in  $kcal \cdot mol^{-1}$ . Signs of the backward transformation have been reverted to correspond to the forward transformation.

transformation	split <sup>a</sup>		unified <sup>b</sup>		absolute <sup>c</sup>	
	RF $\Delta\Delta G$	PME $\Delta\Delta G$	RF $\Delta\Delta G$	PME $\Delta\Delta G$	RF $\Delta\Delta G$	PME $\Delta\Delta G$
ethane	-0.03 $\pm$ 0.01	-0.04 $\pm$ 0.02	-0.02 $\pm$ 0.00	-0.03 $\pm$ 0.00	-0.06 $\pm$ 0.01	-0.04 $\pm$ 0.01
methane	-0.01 $\pm$ 0.02	-0.02 $\pm$ 0.01	0.046 $\pm$ 0.020 <sup>d</sup>	0.01 $\pm$ 0.02		
methanol	6.16 $\pm$ 0.01	6.20 $\pm$ 0.00	7.30 $\pm$ 0.02	7.38 $\pm$ 0.01	5.77 $\pm$ 0.01	5.95 $\pm$ 0.01
methane	6.17 $\pm$ 0.01	6.20 $\pm$ 0.01	7.09 $\pm$ 0.02	7.17 $\pm$ 0.02		
ethane	-6.12 $\pm$ 0.01	-6.19 $\pm$ 0.01	-7.12 $\pm$ 0.01	-7.21 $\pm$ 0.02	-5.83 $\pm$ 0.01	-5.98 $\pm$ 0.01
methanol	-6.12 $\pm$ 0.01	-6.19 $\pm$ 0.00	-7.34 $\pm$ 0.00	-7.40 $\pm$ 0.00		
toluene	3.22 $\pm$ 0.01	3.21 $\pm$ 0.01	3.23 $\pm$ 0.01	3.22 $\pm$ 0.01	2.97 $\pm$ 0.01	3.16 $\pm$ 0.01
methane	3.25 $\pm$ 0.01	3.20 $\pm$ 0.01	3.22 $\pm$ 0.01	3.21 $\pm$ 0.00		
neopentane <sup>e</sup>	-0.10 $\pm$ 0.01	-0.15 $\pm$ 0.02	-0.08 $\pm$ 0.02	-0.18 $\pm$ 0.03	-0.18 $\pm$ 0.01	-0.14 $\pm$ 0.01
methane <sup>e</sup>	-0.11 $\pm$ 0.02	-0.16 $\pm$ 0.05	0.00 $\pm$ 0.03	-0.18 $\pm$ 0.03		
neopentane <sup>f</sup>	-0.12 $\pm$ 0.01	-0.13 $\pm$ 0.01	-0.14 $\pm$ 0.01	-0.14 $\pm$ 0.01		
methane <sup>f</sup>	-0.10 $\pm$ 0.03	-0.18 $\pm$ 0.03	-0.09 $\pm$ 0.01	-0.15 $\pm$ 0.02		
2-methylfuran	2.99 $\pm$ 0.01	2.93 $\pm$ 0.05	3.05 $\pm$ 0.01	3.00 $\pm$ 0.01	2.87 $\pm$ 0.01	2.95 $\pm$ 0.01
methane	3.01 $\pm$ 0.00	2.96 $\pm$ 0.01	3.06 $\pm$ 0.01	3.01 $\pm$ 0.01		
2-methylindole	8.71 $\pm$ 0.02	8.73 $\pm$ 0.03	8.73 $\pm$ 0.01	8.80 $\pm$ 0.03	8.44 $\pm$ 0.02	8.79 $\pm$ 0.02
methane	8.73 $\pm$ 0.03	8.74 $\pm$ 0.01	8.30 $\pm$ 0.02	8.77 $\pm$ 0.04		
2-CPI	-0.07 $\pm$ 0.02	-0.03 $\pm$ 0.03	-0.10 $\pm$ 0.05	-0.2 $\pm$ 0.1	-0.02 $\pm$ 0.05	0.02 $\pm$ 0.02
7-CPI	-0.12 $\pm$ 0.06	-0.20 $\pm$ 0.04	-0.04 $\pm$ 0.06	-0.14 $\pm$ 0.09		

<sup>a</sup>results obtained from alchemical transformations with electrostatic and bonded scaling separate from vdW parameter change.

<sup>b</sup>results obtained from alchemical transformation with all parameters scaling together.

<sup>c</sup>results obtained from absolute free energy calculations.

<sup>d</sup>inverted sign

<sup>e</sup>central mapping

<sup>f</sup>terminal mapping

Table 7: Relative hydration free energies of methanol  $\rightarrow$  methane and methane  $\rightarrow$  methanol transformations without and with the use of Coulomb softcore potentials. The complete version of this table is in the SI.

		split		split+sc		absolute	
		RF	PME	RF	PME	RF	PME
transformation		$\Delta\Delta G$	$\Delta\Delta G$	$\Delta\Delta G$	$\Delta\Delta G$	$\Delta\Delta G$	$\Delta\Delta G$
methanol	methane	$6.16 \pm 0.01$	$6.20 \pm 0.00$	$7.32 \pm 0.03$	$7.42 \pm 0.04$	$5.77 \pm 0.01$	$5.95 \pm 0.01$
methane	methanol	$6.17 \pm 0.01$	$6.20 \pm 0.01$	$7.14 \pm 0.03$	$7.21 \pm 0.03$		

hydrogen bond constraining algorithms such as SHAKE or LINCS. Successful simulations require turning off the constraint and decreasing the time step, but there are alternatives that require some scripting and changes in the topology file: the use of algorithms mentioned above and the definition of a new atom type for the hydrogen atoms that will be alchemically changed, or the use of `constraint = none` and of explicit constraints in the topology file.

### 3.4 SOMD

Fig. S4 compares relative free energy of hydration  $\Delta\Delta G$  according to the 'unperturbed H bond' protocol, with relative  $\Delta\Delta G$  obtained from two absolute free energy calculations. Table 3 summarizes all the computed relative free energy of hydration for the dataset in Fig. 2. A very good agreement is observed between both methodologies ( $R^2 = 0.99 \pm 0.01$  and  $\text{MAE} = (0.10 \pm 0.03) \text{ kcal mol}^{-1}$ ), highlighting internal consistency within SOMD.

To achieve this level of reproducibility within SOMD it was crucial to use the "unperturbed H bond" protocol. Specifically, bonds that involve unperturbed hydrogen atoms are constrained. Bonds involving hydrogen atoms that are perturbed to a heavy elements are unconstrained. Additionally the atomic mass of the perturbed hydrogen atom is set to the mass of the heavy atom it is perturbed to. Bonds involving hydrogen atoms that are perturbed to another hydrogen atom type are constrained. This protocol suppresses high frequency vibrations in flexible bonds involving hydrogen atoms, thus enabling a time step of 2 fs, whilst giving essentially negligible errors due to the use of constraints for perturbed bonds.

This is apparent from the comparison with the absolute hydration free energy calculations. Additionally, the protocol yields relative hydration free energy extremely similar (MAE =  $0.09 \text{ kcal mol}^{-1}$ ) to those computed from simulations where no constraints are applied on the solutes and a timestep of 1 fs is used (Figure S6).

By contrast, a protocol that constrains all bonds in a solute leads to significant differences with the absolute hydration free energies. For instance neopentane  $\rightarrow$  methane (centrally mapped) gives a RAFF  $\Delta\Delta G = (2.04 \pm 0.01) \text{ kcal mol}^{-1}$  whereas the absolute hydration free energy calculations give  $\Delta\Delta G = (-0.19 \pm 0.06) \text{ kcal mol}^{-1}$  as shown in tab. S6 and fig. S6.

This discrepancy occurs because in the SOMD implementation, the energies of constrained bonds are not evaluated, but the calculation of the energies of the solute at perturbed  $\lambda$  values is carried out using the coordinates of the reference  $\lambda$  trajectory. This leads to a neglect of contributions of the bonded term (and associated coupled terms) to the free energy change. The effect is more pronounced for perturbations that feature a large change in equilibrium bond lengths, such as those where a hydrogen atom is perturbed to/from a heavy atom. This problem was already discussed by Pearlman<sup>82</sup> and Boresch,<sup>29,54</sup> as a missing bond-stretching term, which introduce a free energy offset estimated as:

$$\Delta G = RT \ln \left( \frac{r_i}{r_f} \right) \quad (7)$$

where  $r_i$  and  $r_f$  are the initial and final bond lengths, respectively.

The reaction fields implemented in SOMD and GROMACS differ somewhat (atom-based shifted Barker Watts<sup>66</sup>) vs group based switched Barker Watts), but nevertheless SOMD and GROMACS RF produce comparable results with a MAE of  $0.18 \text{ kcal mol}^{-1}$ .

Overall, the SOMD free energy estimations are in good agreement with the other MD packages, as the MAE suggests (see Table XX). Reaction field and PME results are in good agreement. In particular, the unperturbed hydrogen bonds constraint protocol allows a timestep of 2 fs. All SOMD RAFF simulations were carried out with simultaneous transfor-

mation of Lennard-Jones, charges, and bonded terms. This suggests that the failure of the GROMACS "unified protocol" in some instances may be due to differences in the softcore implementations. As regards methane  $\rightarrow$  neopentane transformation SOMD is able to maintain consistency between central and terminal mapping, as shown in tab S5 while GROMACS RF displays a  $\Delta\Delta G$  of  $(-0.20 \pm 0.01)$  kcal mol $^{-1}$  for the centrally mapped transformation and  $\Delta\Delta G = (-0.05 \pm 0.02)$  kcal mol $^{-1}$  for the terminally mapped transformation. The SOMD RF calculations show a MAE = 0.18 kcal mol $^{-1}$  and MAE = 0.23 kcal mol $^{-1}$  with GROMACS PME and AMBER respectively, similar to the MAE obtained with GROMACS RF.

## 4 Discussion

8.recommended protocols

9.protocols to avoid

10.lessons learned

11.2-cyclopentanylindole to 7-cyclopentanylindole: better to go through intermediates?

12.what do we need to progress the field e.g. automation to make things easy but also consistent and thus more reproducible (FESetup also for reproducibility); GPU: GROMACS, SOMD but not AMBER (yet) and CHARMM; alternative softcore functions?; sampling?; force field improvements?; analysis?

13.developer notes: constraints, both appearing/disappearing; lambda paths for AMBER (relative), absolute: crgmask requires vacuum corr if separated protocol

14.further investigation required: binding RAFEs?

## ToDo

	P.
1. Unify language! . . . . .	2
2. Objectives: Do the relative free energy results produced across different softwares agree with each other? What if they don't? . . . . .	2

3. Objectives: Can we outline what should be a standard procedure to calculate relative free energies of hydration alchemically? . . . . .	2
4. Objectives: We want to emphasize that we aim at improving codes/protocols/practices and not highlight "bad" codes . . . . .	2
5. Be clear on what reproducibility means here e.g. no exact numerical . . . . .	2
6. bug in TI gradient accumulation in parallel runs (does not affect serial?, does not affect EXP) . . . . .	24
7. cannot handle LRC: test with larger cutoffs and/or LRC correction with arbitrary, single structure; check <a href="http://pubs.acs.org/doi/abs/10.1021/jp0735987">http://pubs.acs.org/doi/abs/10.1021/jp0735987</a> . . . . .	24
8. recommended protocols . . . . .	29
9. protocols to avoid . . . . .	29
10. lessons learned . . . . .	29
11. 2-cyclopentanylindole to 7-cyclopentanylindole: better to go through intermediates? 29	
12. what do we need to progress the field e.g. automation to make things easy but also consistent and thus more reproducible (FESetup also for reproducibility); GPU: GROMACS, SOMD but not AMBER (yet) and CHARMM; alternative softcore functions?; sampling?; force field improvements?; analysis? . . . . .	29
13. developer notes: constraints, both appearing/disappearing; lambda paths for AMBER (relative), absolute: crgmask requires vacuum corr if separated protocol . . . . .	29
14. further investigation required: binding RAFEs? . . . . .	29

## Acknowledgement

HHL is supported through an EPSRC provided SLA, funding the core support of CCP-BioSim. CCPBioSim is the Collaborative Computational Project for Biomolecular Simulation funded by EPSRC grants EP/J010588/1 and EP/M022609/1. JM is supported by a Royal Society University Research Fellowship. The research leading to these results has

received funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/2007–2013)/ERC Grant agreement No. 336289. GDRM appreciates the support from the Brazilian agency CAPES - Science without Borders program (BEX 3932-13-3). DLM appreciates support from the National Science Foundation (CHE 1352608), and computing support from the UCI GreenPlanet cluster, supported in part by NSF Grant CHE-0840513.

We thank Prof. Stefan Boresch for valuable discussions and making code modifications to CHARMM. We thank Dr. Ross Walker and Daniel Mermelstein for valuable discussions and making code modifications to AMBER. We thank Prof. Michael Shirts for valuable discussions about GROMACS.

We acknowledge use of Hartree Centre resources and the use of the SCARF HPC cluster in this work.

## References

- (1) Hansen, N.; Van Gunsteren, W. F. Practical aspects of free-energy calculations: A review. *Journal of Chemical Theory and Computation* **2014**, *10*, 2632–2647.
- (2) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *The Journal of Physical Chemistry B* **2010**, *114*, 10235–10253, PMID: 20701361.
- (3) Gallicchio, E.; Levy, R. M. In *Computational chemistry methods in structural biology*; Christov, C., Ed.; Advances in Protein Chemistry and Structural Biology; Academic Press, 2011; Vol. 85; pp 27 – 80.
- (4) Lu, X.; Fang, D.; Ito, S.; Okamoto, Y.; Ovchinnikov, V.; Cui, Q. QM/MM free energy simulations: recent progress and challenges. *Molecular Simulation* **2016**, *42*, 1056–1078, PMID: 27563170.

- (5) Rickman, J. M.; LeSar, R. Free-Energy Calculations in Materials Research. *Annual Review of Materials Research* **2002**, *32*, 195–217.
- (6) Wan, S.; Knapp, B.; Wright, D. W.; Deane, C. M.; Coveney, P. V. Rapid, Precise, and Reproducible Prediction of Peptide–MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. *Journal of Chemical Theory and Computation* **2015**, *11*, 3346–3356, PMID: 26575768.
- (7) Beveridge, D. L.; Dicapua, F. M. Free Energy Via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annual Review of Biophysics and Biophysical Chemistry* **1989**, *18*, 431–492.
- (8) Straatsma, T. P.; Mccammon, J. A. Computational Alchemy. *Annual Review of Physical Chemistry* **1992**, *43*, 407–435.
- (9) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews* **1993**, *93*, 2395–2417.
- (10) Squire, D. R.; Hoover, W. G. Monte Carlo Simulation of Vacancies in Rare–Gas Crystals. *The Journal of Chemical Physics* **1969**, *50*, 701–706.
- (11) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22*, 245–268.
- (12) Mruzik, M. R.; Abraham, F. F.; Schreiber, D. E.; Pound, G. M. A Monte Carlo study of ion–water clusters. *The Journal of Chemical Physics* **1976**, *64*, 481–491.
- (13) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. Thermodynamics of cavity formation in water. A molecular dynamics study. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55–67.
- (14) Tembe, B. L.; McCammon, J. Ligand-receptor interactions. *Computers & Chemistry* **1984**, *8*, 281 – 283.



- (15) Jorgensen, W. L.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *The Journal of Chemical Physics* **1985**, *83*, 3050–3054.
- (16) Gilson, M.; Given, J.; Bush, B.; McCammon, J. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical Journal* **1997**, *72*, 1047 – 1069.
- (17) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B* **2003**, *107*, 9535–9551.
- (18) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2009**, *113*, 2234–2246.
- (19) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. Comparison of free energy methods for molecular systems. *Journal of Chemical Physics* **2006**, *125*, 184114.
- (20) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of Computational Chemistry* **2015**, *36*, 348–354.
- (21) Sandberg, R. B.; Banchelli, M.; Guardiani, C.; Menichetti, S.; Caminati, G.; Procacci, P. Efficient Nonequilibrium Method for Binding Free Energy Calculations in Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2015**, *11*, 423–435, PMID: 26580905.
- (22) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *Journal of Chemical & Engineering Data* **0**, *0*, null.

- (23) Woods, C. J.; Malaisree, M.; Hannongbua, S.; Mulholland, A. J. A water–swap reaction coordinate for the calculation of absolute protein–ligand binding free energies. *The Journal of Chemical Physics* **2011**, *134*, 054114.
- (24) Woods, C. J.; Malaisree, M.; Michel, J.; Long, B.; McIntosh-Smith, S.; Mulholland, A. J. Rapid decomposition and visualisation of protein–ligand binding free energies by residue and by water. *Faraday Discuss.* **2014**, *169*, 477–499.
- (25) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society* **2015**, *137*, 2695–2703, PMID: 25625324.
- (26) Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *Journal of Chemical Theory and Computation* **2017**, *13*, 42–54, PMID: 27933808.
- (27) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *The Journal of Physical Chemistry* **1994**, *98*, 1487–1493.
- (28) Michel, J.; Essex, J. W. Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 639–658.
- (29) Boresch, S.; Karplus, M. The Role of Bonded Terms in Free Energy Simulations: 1. Theoretical Analysis. *The Journal of Physical Chemistry A* **1999**, *103*, 103–118.

- (30) Liu, S.; Wang, L.; Mobley, D. L. Is Ring Breaking Feasible in Relative Binding Free Energy Calculations? *Journal of Chemical Information and Modeling* **2015**, *55*, 727–735, PMID: 25835054.
- (31) Shobana, S.; Roux, B.; Andersen, O. S. Free Energy Simulations: Thermodynamic Reversibility and Variability. *The Journal of Physical Chemistry B* **2000**, *104*, 5179–5190.
- (32) Michel, J.; Verdonk, M. L.; Essex, J. W. Protein–Ligand Complexes: Computation of the Relative Free Energy of Different Scaffolds and Binding Modes. *Journal of Chemical Theory and Computation* **2007**, *3*, 1645–1655, PMID: 26627610.
- (33) Rocklin, G. J.; Mobley, D. L.; Dill, K. A. Separated topologies—A method for relative binding free energy calculations using orientational restraints. *The Journal of Chemical Physics* **2013**, *138*, 085104.
- (34) Axelsen, P. H.; Li, D. Improved convergence in dual-topology free energy calculations through use of harmonic restraints. *Journal of Computational Chemistry* **1998**, *19*, 1278–1283.
- (35) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26*, 1668–1688.
- (36) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.

- (37) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (38) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **1999**, *103*, 3596–3607.
- (39) Woods, C.; Mey, A. S.; Calabro, G.; Michel, J. Sire molecular simulations framework. 2016.
- (40) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation* **2013**, *9*, 461–469, PMID: 23316124.
- (41) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *Journal of Computer-Aided Molecular Design* **2017**, *31*, 147–161.
- (42) Pronk, S.; Larsson, P.; Pouya, I.; Bowman, G. R.; Haque, I. S.; Beauchamp, K.; Hess, B.; Pande, V. S.; Kasson, P. M.; Lindahl, E. Copernicus: A New Paradigm for Parallel Adaptive Molecular Dynamics. Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis. New York, NY, USA, 2011; pp 60:1–60:10.
- (43) Sadiq, S. K.; Wright, D.; Watson, S. J.; Zasada, S. J.; Stoica, I.; Coveney, P. V. Automated Molecular Simulation Based Binding Affinity Calculator for Ligand-Bound

- HIV-1 Proteases. *Journal of Chemical Information and Modeling* **2008**, *48*, 1909–1919, PMID: 18710212.
- (44) Lundborg, M.; Lindahl, E. Automatic GROMACS Topology Generation and Comparisons of Force Fields for Solvation Free Energy Calculations. *The Journal of Physical Chemistry B* **2015**, *119*, 810–823, PMID: 25343332.
- (45) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *Journal of Chemical Information and Modeling* **2015**, *55*, 2485–2490.
- (46) Balasubramanian, V.; Bethune, I.; Shkurti, A.; Breitmoser, E.; Hruska, E.; Clementi, C.; Laughton, C. A.; Jha, S. ExTASY: Scalable and Flexible Coupling of MD Simulations and Advanced Sampling Techniques. *CoRR* **2016**, *abs/1606.00093*.
- (47) Skyner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B. O. A Review of Methods for the Calculation of Solution Free Energies and the Modelling of Systems in Solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191.
- (48) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 711–720.
- (49) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters* **1994**, *222*, 529–539.
- (50) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard–Jones interactions in thermodynamic integration. *The Journal of Chemical Physics* **1994**, *100*, 9025–9031.

- (51) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *The Journal of Chemical Physics* **2007**, *127*, 214108.
- (52) Cross, A. J. Influence of hamiltonian parameterization on convergence of kirkwood free energy calculations. *Chemical Physics Letters* **1986**, *128*, 198 – 202.
- (53) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.
- (54) Boresch, S.; Karplus, M. The Role of Bonded Terms in Free Energy Simulations. 2. Calculation of Their Influence on Free Energy Differences of Solvation. *The Journal of Physical Chemistry A* **1999**, *103*, 119–136.
- (55) Rocklin, G. J.; Mobley, D. L.; Dill, K.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *The Journal of chemical physics* **2013**, *139*, 184103.
- (56) Loeffler, H. H.; Sotriffer, C. A.; Winger, R. H.; Liedl, K. R.; Rode, B. M. Calculation of sequence-dependent free energies of hydration of dipeptides formed by alanine and glycine. *Journal of Computational Chemistry* **2001**, *22*, 846–860.
- (57) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.
- (58) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **2000**, *21*, 132–146.

- (59) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641.
- (60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- (61) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *Journal of Chemical Theory and Computation* **2012**, *8*, 2553–2558, PMID: 26592101.
- (62) Hu, Y.; Sherborne, B.; Lee, T.-S.; Case, D. A.; York, D. M.; Guo, Z. The importance of protonation and tautomerization in relative binding affinity prediction: a comparison of AMBER TI and Schrödinger FEP. *Journal of Computer-Aided Molecular Design* **2016**, *30*, 533–539.
- (63) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-core potentials in thermodynamic integration: comparing one- and two-step transformations. *Journal of Computational Chemistry* **2011**, *32*, 3253–3263.
- (64) Naden, L. N.; Pham, T. T.; Shirts, M. R. Linear basis function approach to efficient alchemical free energy calculations. 1. Removal of uncharged atomic sites. *Journal of Chemical Theory and Computation* **2014**, *10*, 1128–1149.
- (65) Naden, L. N.; Shirts, M. R. Linear basis function approach to efficient alchemical free energy calculations. 2. Inserting and deleting particles with coulombic interactions. *Journal of Chemical Theory and Computation* **2015**, *11*, 2536–2549.
- (66) Barker, J.; Watts, R. Monte Carlo studies of the dielectric properties of water-like models. *Molecular Physics* **1973**, *26*, 789–792.

- (67) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* **1980**, *72*, 2384–2393.
- (68) Bosisio, S.; Mey, A. S. J. S.; Michel, J. Blinded predictions of distribution coefficients in the SAMPL5 challenge. *Journal of Computer-Aided Molecular Design* **2016**, *30*, 1101–1114.
- (69) Zwanzig, R. High-Temperature Equation of State by a Perturbation Method. I. Non-polar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.
- (70) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *The Journal of chemical physics* **1988**, *89*, 3742–3746.
- (71) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *Journal Of Chemical Physics* **1935**, *3*, 300–313.
- (72) Wu, D.; Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *Journal of Chemical Physics* **2005**, *123*.
- (73) Wu, D.; Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *Journal of Chemical Physics* **2005**, *123*.
- (74) Wu, D.; Kofke, D. A. Asymmetric bias in free-energy perturbation measurements using two Hamiltonian-based models. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* **2004**, *70*, 066702.
- (75) Lu, N.; Singh, J. K.; Kofke, D. A. Appropriate methods to combine forward and reverse free-energy perturbation averages. *The Journal of Chemical Physics* **2003**, *118*, 2977–2984.
- (76) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*.



- (77) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*, 124105.
- (78) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **2015**, *29*, 397–411.
- (79) Klimovich, P. V.; Mobley, D. L. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 307–316.
- (80) Lee, T.-S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *Journal of Chemical Theory and Computation* **0**, *0*, null, PMID: 28618232.
- (81) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *Journal of Chemical Theory and Computation* **2013**, *9*, 4131–4139.
- (82) Pearlman, D. A.; Kollman, P. A. The overlooked bond-stretching contribution in free energy perturbation calculations. *The Journal of chemical physics* **1991**, *94*, 4532–4545.

## Graphical TOC Entry

