Reproducing Relative Alchemical Free Energies of Hydration (draft)

Hannes H. Loeffler,*,† Stefano Bosisio,‡ Julien Michel,‡ Guilherme Duarte Ramos Matos,¶ David L. Mobley,§ Donghyuk Suh,‖ and Benoit Roux‖

†*Science & Technology Facilities Council, Daresbury, Warrington, WA4 4AD, United Kingdom*

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

¶*Department of Chemistry, University of California, Irvine*

§*Departments of Pharmaceutical Sciences and Chemistry, University of California, Irvine*

‖*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 Sire/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 also 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 given some level of care and attention to detail. We will also recommend simulation protocols, setup procedures and analysis techniques, and we discuss what needs to be done to ensure continued progress.

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 product 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 this reversible work 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 through computers[1–5] has been particularly attractive

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.[1–5] However, the reliability of estimates is still very much a matter of concern.[2,6] Roughly speaking, fast methods tend to be less accurate while more accurate methods tend to be slow. Nevertheless, rigorous methods are obligatory in obtaining accurate, precise and reliable results, while less accurate methods can be used as an input filter for computationally more expensive approaches. By rigorous we mean methods that 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 whose theory is firmly rooted in statistical thermodynamics and is argued to be the most accurate method in quantitative prediction of free energies.[1,7–9] The method has been applied in various forms for many decades now since the early days of computer simulation.[10–15] The method has gained renewed attention in recent years — concomitant with improvements in computer hardware design — both within the traditional equilibrium framework[16–18] but also increasingly in combination with non-equilibrium techniques.[19–21] 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 force field's parameters determining strength and equilibrium of interactions 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 are constructed around the concept of thermodynamic cycles.[14] As the free energy is a state function, the sum of free energy changes computed

around any closed cycle must sum to 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 alchemical calculation. As mentioned above, the latter may be computationally and statistically more efficient.

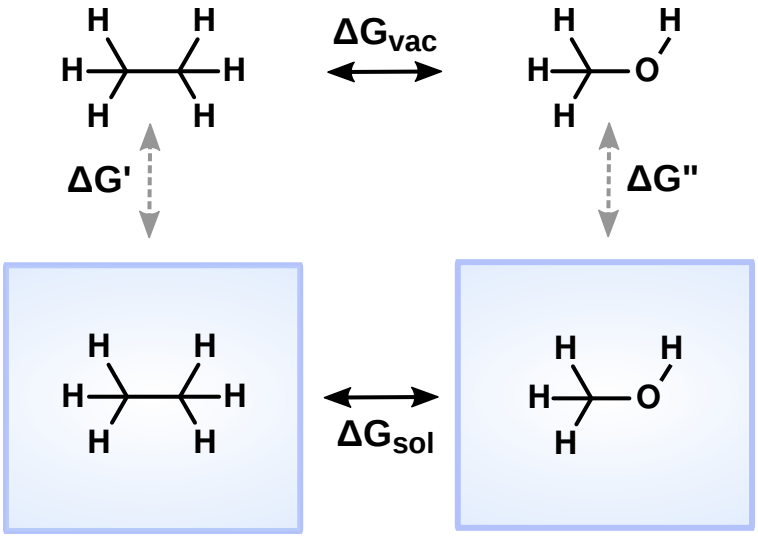

Figure 1: The thermodynamic cycle to compute the relative free energy of hydration $\Delta\Delta G_{\mathrm{hydr}} = \Delta G_{\mathrm{sol}} - \Delta G_{\mathrm{vac}} = \Delta G'' - \Delta G'$. The example is for the ethanol $\leftrightarrow$ methanol transformation. Alchemical simulations can be performed along the non–physical horizontal legs while vertical legs illustrate the physical process (which is directly 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.[16–19,21] *Absolute* here really means that the equilibrium constant of a physical reaction, e.g. binding and dissociation, can be calculated directly by completely decoupling a whole molecule from its environment and the term is mostly being used to discriminate against *relative* techniques (see below). 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 would also scale 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.[23,24]

Relative alchemical free energy (RAFE) calculations "mutate" one molecule into another one. This is most efficiently achieved[25,26] by making use of the single topology method.[15,25,27] 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. 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[27] (see also "wandering" ligand problem below). AMBER 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 for 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.[28,29]

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 more traction in the context of relative free binding energies between small molecules e.g. drug or lead like molecules and biomolecules.[30,31]

The single topology approach[25] requires a certain similarity between the two mutated states. This means primarily 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[31] but it has also been shown that this should be done only in certain circumstances.[28,29]

When the two molecules are very dissimilar, the dual topology method[25,27] can be applied

to compute relative free energies. In this approach all atoms in 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.[27] 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.[16–18] Topological similarity can only be exploited when the charges of the common core are explicitly made equivalent.[6] This approach, however, shifts all the chemical variability exclusively to the dummy atoms and is thus of only limited use.

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.[32,33] 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,[34] CHARMM,[35] GROMACS,[36] GROMOS[37] and Sire/SOMD[38,39] 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.[29]

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[40] — 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, the reversible work computed with any simulation software should be expected to be reproducible within statistical error. Modern MD packages support a wide range of force fields and meth-

ods such that these packages are replaceable with each other to an ever increasing extent and the choice of the right package for the user becomes less and less a matter of technical restrictions.

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.[41] 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.[41] A large database of hydration free energies computed from alchemical free energy (AFE) simulation, FreeSolv, has been presented recently[42] and was just updated.[22] Here, we are interested in the reproducibility of RAFE with the simulation programs AMBER, CHARMM, GROMACS and Sire/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

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 Sire/SOMD. One key difference is in the softcore functions employed[43,44] used in each code as summarised in section 1f the SI. Softcore functions are used to avoid numerical and thus stability problems of the conventional van der Waals and Coulombic potentials[45] 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 analysis problems even when simulations run.[43–45]

Another difference is how the code scales force field parameters ("parameter scaling") and/or the total energy ("energy scaling").[27] 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[25,46] 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 linear scaling the 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 i.e. also for atoms that have an equivalent in the other state but 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 changing bond lengths involving constraints in the perturbed group. There is only one global λ for parameter transformation. Separated protocols (see below) must be emulated through careful construction of topologies by keeping force field parameters

that are not being modified constant between different files.

CHARMM The PERT module duplicates the topology similar to `sander` but the definitions for mapped atoms are given 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 1n the SI). The code can handle constraints of changing bond lengths in the perturbed group but this may cause wrong results with PSSP softcores (Stefan Boresch, private communication). There is only one global λ 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. The softcore potential applies to dummy atoms only determined from atoms having zero vdW parameters. 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 λs 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). Bond and angle terms are parameter–scaled while the dihedral term is energy–scaled. The softcore potential applies to dummy atoms only. The code cannot handle constraints of changing bond lengths in the perturbed group. There is only one global λ for parameter scaling. Separated protocols (see below) must be emulated through careful construction of the patch file. Sire/SOMD is Sire[38] employing OpenMM[39] for MD simulation.

2.2 RAFE Setup

The setup for all relative free energy simulations has been carried out with the tool FESetup[47] in version 1.2. FESetup is a perturbed topology writer for AMBER, CHARMM, GROMACS, Sire/SOMD and also NAMD[48] (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. This means atom type to atom type conversion and the only current limit is that rings are required to be preserved.[28] In this way we achieve a maximal single topology description: 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 will have the same amount of total atoms. 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. 47.

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 λ. 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.

The ethane $\rightarrow$ methanol transformation is traditionally regarded as a standard test for RAFE simulations.[15,49] The other transformations are centred around mutations from and to methane. 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. 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". Results from both mutations will be shown and discussed.

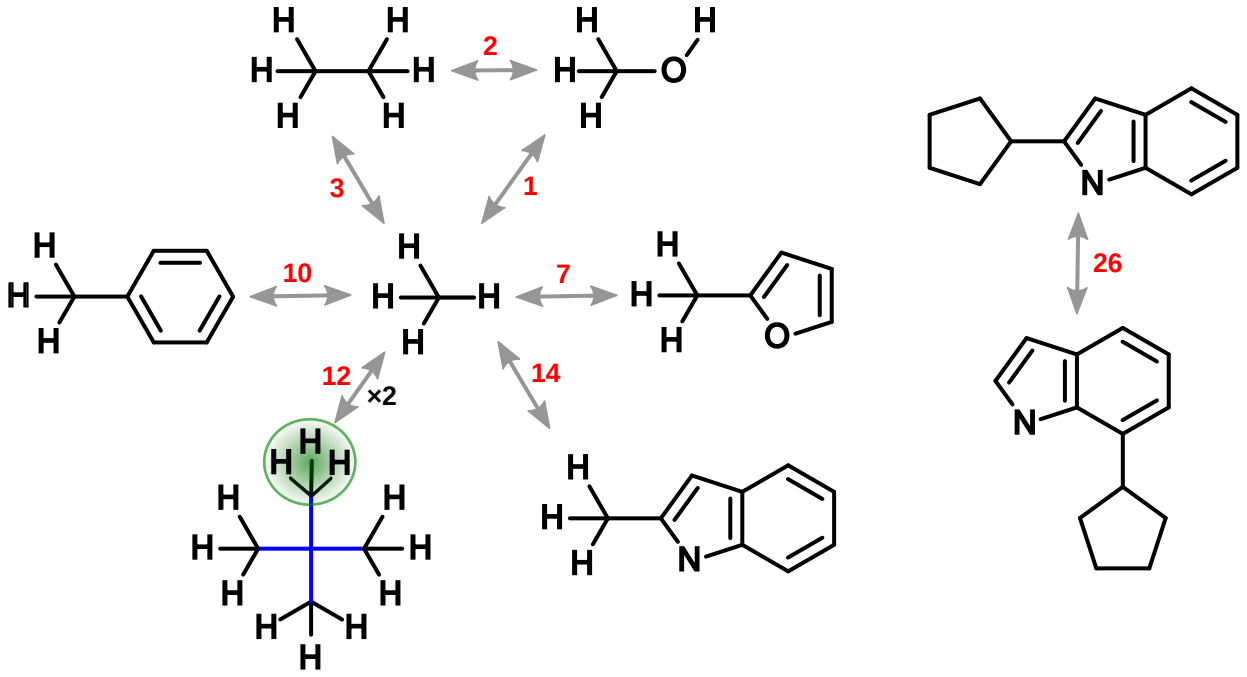

Figure 2: The thermodynamic cycles considered in this study. To compute the free energy of hydration all pair–wise transformations have to be carried out once in solution and once in vacuum. Green and blue colours in neopentane show two alternative mappings for methane. The numbers in red denote the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

2.3 RAFE 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 MD software employs a wide range of methods and algorithms that may not be available in the 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.[50,51] The force field was chosen to be GAFF[52] (version 1.8), utilising AM1/BCC charges[53,54] for the solute and TIP3P[55] for the solvent. Charges were computed with the `antechamber` program and missing bonded and vdW terms were gener-

ated 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. 56,57.

While the MD packages principally allow a "one–step" transformation[58] that is both van der Waals and Coulombic contributions vary simultaneously, it can be(is? ref?) more efficient to carry out a separated protocol. 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[43,44] (see section 1n 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 2n the SI.

All simulations were started from the simulation box which has been created with FESetup.[47] 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 λ 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 λ 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 Sire protocol below for an alternative

approach). The temperature was controlled through a Langevin thermostat with a friction constant of 2.0 ps − 1 and pressure rescaling through a Monte Carlo barostat with 100 steps between isotropic volume change attempts.

Sire/SOMD All simulations were carried out with Sire/OpenMM 6.3 (revision 16.1). 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 "unperturbed H protocol". An atom–based Barker–Watts reaction field[59] with a dielectric constant of 82.0 was adopted for the solution phase simulations. Furthermore, a cutoff of 12 Å was set for the non–bonded interactions and periodic boundary conditions were imposed. Temperature control was achieved with the Andersen thermostat[60] with a coupling constant of $10\,\mathrm{ps}^{-1}$. A Monte Carlo barostat assured pressure control, attempting isotropic box edge scaling every 25 time steps. The reaction field was not employed in vacuum and a Coulombic potential without cutoff was used. This aspect led to an inconsistent description of the intramolecular electrostatic interactions of the solute in the solvated and vacuum phases. Thus, to enable a meaningful estimation of the free energy change a free energy correction term G_{c} was evaluated to treat intramolecular Coulombic interactions consistently between the water and vacuum legs of the thermodynamic cycle (see Fig 1) as detailled in Ref. 61. The G_{c} term was obtained via post-processing of the end state trajectories of each water phase simulation, using the Zwanzig relationship:[62]

$$G_{\mathrm{c}} = -\beta^{-1} \ln\langle exp\left[-\beta(U_{\mathrm{ic,nc}}(\mathbf{r}) - U_{\mathrm{ic,sim}}(\mathbf{r}))\right]\rangle_{\mathrm{sim}} \tag{1}$$

where $U_{\mathrm{ic,nc}}(\mathbf{r})$ is the solute intramolecular electrostatic potential that depends on the coordinates $\mathbf{r}$ of the solute and is given by Coulomb's law. $U_{\mathrm{ic,sim}}(\mathbf{r})$ is the intramolecular electrostatic potential term as computed in the simulation with a BWRF cutoff.

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 a 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. We used the Langevin integrator implemented in GROMACS with a default friction coefficient of 1.0 $\text{ps}/\text{m}_{\text{atom}}$. Absolute hydration free energies were calculated using a protocol used in previous works.[22,42] A Parrinello–Rahman barostat with $\tau_p = 10$ ps and compressibility equal to 4.5×10^{-5} bar^{-1}. We used two methods to calculate electrostatic interactions: Particle Mesh Ewald (PME) and Reaction Field with a dielectric of 78.3, as implemented in the software. We set the non–bonded cutoff to 10.0 Å, with a switch starting at 9.0 Å. PME calculations were of order 6 and had a tolerance of 1.0×10^{-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.

6.make figure explaining the protocols described here.

2.4 Analysis

Various estimators have been proposed to obtain the free energy from AFE simulations. Early work by Kirkwood[63] expresses the free energy as an ensemble average of the derivative of the Hamiltonian with respect to the coupling parameter λ. The method is now known as

Thermodynamic Integration (TI). Zwanzig devised the exponential formula[62] (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 with the configuration of one end state being used for both end states assuming that this configuration is representative for either state. As the phase–space overlap needs to be sufficiently large[64,65] the EXP approach typically needs intermediates, controlled by λ. However, it has been shown that EXP has an asymmetric bias depending on the directionality of λ[66] and that the Bennett Acceptance Ratio (BAR) method[11] is considerably more effective in obtaining an accurate result.[67] 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 incorporating more than just the immediate λ neighbours and, in fact, all other λs. This approach has been called multi–state BAR (MBAR)[68] method. MBAR has been shown to have the lowest variance of any known estimator.[69]

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 2 computes the free energy as

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\mathscr{H}(\mathbf{q}, \mathbf{p}; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{2}$$

where $\mathscr{H}(\mathbf{q}, \mathbf{p}; \lambda)$ 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 λ. The angle brackets denote the ensemble average of the gradient of the Hamiltonian with respect to λ at a given λ value. An AFE simulation is typically carried out in a series of equilibrium simulations at discrete values of λ 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[70] for 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 RAFE 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 $\Delta\Delta G_{\mathrm{hydr}}$ 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 pmplementation–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

$$\mathrm{err}(\Delta\Delta G_{\mathrm{hydr}}) = \frac{\sigma}{\sqrt{n}} \tag{3}$$

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

$$\mathrm{err}(\mathrm{combined}) = \sqrt{\sum_i \sigma_i^2}. \tag{4}$$

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

3 Results

7.what protocols did not work

3.1 AMBER

Using AMBER for RAFE simulations has revealed several problems with the implementation. Some bugs were identified and have been fixed for AMBER16 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 (Ross Walker, private communication). A disadvantage of `sander` is that it cannot be used to simulate the λ end points[71] such that the TI gradients need to be extrapolated (minimum and maximum allowed λs 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.[71]

We also found that, in contrast to the other three codes, AMBER cannot correctly reproduce relative free energies in a 1–step protocol i.e. when all force field parameters are scaled simultaneously (see Table S1. This appears to be a problem when more than a few dummy atoms are involved while the 1–step protocol works for the smaller transformations (refer to Figure 2). The separated RAFE protocol and absolute free energies, however, are very close to the other MD packages as demonstrated in Table 1.

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 RAFE Setup and Figure 1). As shown in Figure S3he methane end-state exhibits incorrect distances (which are too long) 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 S3emonstrates 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 form explicit dummy atom simulations and results from absolute transformations. Table 1 lists the free energies for these three approaches together with the SEM. SHAKE was explicitly deactivated for all bonds in the perturbed region in these protocols. In addition, the table also shows selected results for transformations with SHAKE enabled

Table 1: Comparing AMBER results for simulations with implicit and explicit dummy atoms, and results from absolute transformation. A few select cases with SHAKE enabled and a time step of 2 fs are shown in addition. Simulations were carried out with the separated protocol. Signs of the backward transformation have been reverted to correspond to the forward transformation.

transformation		implicit $\Delta\Delta G$	explicit $\Delta\Delta G$	absolute ΔG	SHAKE[a] $\Delta\Delta G$
ethane	methane	0.02 ± 0.01	-0.13 ± 0.02	-0.02 ± 0.01	
methane	ethane	0.00 ± 0.03	-0.19 ± 0.03		
methanol	methane	6.19 ± 0.01	6.20 ± 0.02	6.20 ± 0.01	
methane	methanol	6.20 ± 0.03	-6.15 ± 0.01		
ethane	methanol	-6.20 ± 0.01	-6.27 ± 0.01	-6.22 ± 0.01	-6.20
methanol	ethane	-6.20 ± 0.01	-6.25 ± 0.01		
toluene	methane	3.24 ± 0.02	3.39 ± 0.02	3.19 ± 0.01	3.29
methane	toluene	3.42 ± 0.03	3.52 ± 0.03		
neopentane[b]	methane	0.32 ± 0.04	-0.03 ± 0.06		0.37
methane[b]	neopentane	0.25 ± 0.03	-0.07 ± 0.03	-0.13 ± 0.02	
neopentane[c]	methane	-0.13 ± 0.01	-0.12 ± 0.02		
methane[c]	neopentane	-0.13 ± 0.03	-0.12 ± 0.03		
2-methylfuran	methane	3.09 ± 0.01	3.10 ± 0.01	2.96 ± 0.02	
methane	2-methyfuran	3.10 ± 0.03	3.15 ± 0.03		
2-methylindole	methane	8.78 ± 0.03	8.78 ± 0.04	8.72 ± 0.01	
methane	2-methylindole	9.14 ± 0.02	9.13 ± 0.03		
2–cyclopentanylindole[d]	7–cyclopentanylindole	0.36 ± 0.03	0.63 ± 0.06	0.39 ± 0.04	
7–cyclopentanylindole[d]	2–cyclopentanylindole	0.34 ± 0.05	0.50 ± 0.03		

[a] implicit dummy atom protocol with $\delta t = 2$ fs and SHAKE on all H–bonds except perturbed bonds.
[b] central mapping.
[c] terminal mapping.
[d] partial re/discharge i.e. only the charges of the appearing and the disappering 5–rings are switched.

for all bonds to hydrogens except those bonds that change bond length during transformation. These 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 RAFE simulations (cmp. SOMD simulations). 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 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 1 for the closed cycle ethane→ methanol → methane → 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 $\Delta\Delta G_{\mathrm{hydr}}$ of $(0.069 \pm 0.041)\,\mathrm{kcal\,mol^{-1}}$ and for the explicit dummy simulation the error is $(-0.016 \pm 0.047)\,\mathrm{kcal\,mol^{-1}}$.

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 → methane with central mapping differs from the result with terminal mapping by about $0.4\,\mathrm{kcal\,mol^{-1}}$ (cf. Table 1). 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 S2. A discrepancy of consistently 0.2–$0.4\,\mathrm{kcal\,mol^{-1}}$ is evident from every λ step of the vdW plus bonded transformation with both implicit and explicit dummy atoms).

3.2 CHARMM

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

9.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

10.investigate why methane ethane and ethane methane differ so much from the other packages

11.Make a figure that illustrates the choices of relative transformation

12.Run split protocol simulations with gaff1.8 *and* soft-core coulomb

13.make figures XX1 and XX2 (I think they should probably go in the SI)

GROMACS has some run input options which can simplify the procedure for setting up free energy calculations; however, these same options can also provide a source of error if used naively. Additionally, running the simulations depends on understanding how these features are used because they can be a source of errors if misused. 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-moltype` should not be used in relative transformations in GROMACS because relative calculations rely on turning off interactions between only part of a molecule and the rest of the system, not the entire molecule. `couple-lambda0` and `couple-lambda1` control the interactions of the molecule specified by `couple-moltype` with its surroundings. For instance, if `couple-lambda0` is set to `vdw-q` and `couple-lambda1` to `none`, the calculation will be set in such a way that the tagged molecule is fully coupled to its surrounding by its electrostatic and van der Waals interactions, while it will be completely decoupled from its surroundings in the final state, effectively acting as an ideal gas molecule. These should be the default choices in absolute free energy simulations; there is no need to specify both end states in the topology file. In relative free energy calculations, however, `couple-lambda0` and `-lambda1` are not necessary because `couple-moltype` should be set to `none`. `couple-intramol` specifies if you want

to modify intramolecular interactions along the alchemical path and should be set to `yes` in relative transformations. Setting this to `no` would mean intramolecular interactions are maintained at their existing values while the free energy transformation is conducted, leading to an incomplete transformation and introducing errors in computed free energies.

Here, we choose absolute hydration free energies as our standard point of comparison because these are easily calculated with considerable precision[22] 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[72] (see Fig 7 in reference), further supporting this view. Thus, we compared 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 stages the transformation in a van der Waals parameter change followed by a simultaneous bonded terms and charges modifications, or vice-versa. In the framework of the split routine, windows without a reasonably strong counterbalancing Lennard-Jones component are subject to very large electrostatic forces that will often lead to simulations which crash with standard protocols.[73,74] Thus we find that for stable and effective relative free energy calculations, particle deletion processes require electrostatic terms to be turned off before van de Waals and bonded parameters are removed/modified. Insertion processes, in their turn, demand bonded and van der Waals parameters to be changed before charges are turned on. The relative free energies of processes $A \rightarrow B$ and $B \rightarrow A$ tend to be equivalent if the protocol of the former is the reverse of the latter, i.e., both end states should be connected by the same path in λ-space.

Transformations having simultaneous particle insertion or deletion, such as 2–cyclopentanylindole $\rightarrow$ 7–cyclopentanylindole and its reverse, require an adapted split protocol. The simulation is divided into two stages: a deletion stage, in which part of the molecule is transformed

into dummy atoms, and an insertion stage, in which dummy atoms which were initially present are transformed into an interacting part of the molecule. Each stage requires its own alchemical free energy calculation and they differ from the aforementioned deletion and insertion protocols in such a way that we first turned off the charges on atoms being deleted, then modified the bonded and van der Waals terms, then turned back on the charges on any atoms which were being inserted. This modification allows a smaller insertion stage of 16 windows in which only charges and bonded terms of the new molecular part are modified.

14.make a figure illustrating the idea of the previous paragraph

Table 2 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. A few noteworthy exceptions are the differences between the unified and split results of methane $\rightarrow$ methanol, ethane $\rightarrow$ methanol, and their reverse processes. In the former, $\langle \partial \text{H}/\partial\lambda \rangle$ plots show a sharp decrease near $\lambda = 1.0$, while similar behavior is not observed in the latter (Figure XX1).

Investigation of the phase space overlap plots (Figure XX2) obtained from `alchemical_analysis.py` show in the unified case a poor overlap between the last state and its neighbors, which can explain the difference in free energy.

15.we speculate it is an soft-core coulomb issue, but I'm not confident enough to write it down before I run some gaff1.8 simulations using soft-core coulomb in the split protocol (I had that information for gaff 1.7). It will be possible to compare electrostatic components of $\Delta G_{A\rightarrow B}$ then.

Ethane $\rightarrow$ methane and methane $\rightarrow$ ethane split and unified results are quite different from their corresponding absolute results, and from other MD packages. Neopentane $\rightarrow$ methane transformations, and their reverse processes also disagree with the absolute protocol, but similar differences can be seen in AMBER.

We 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.

Table 2: Relative hydration free energies obtained from GROMACS simulations in $kcal \cdot mol^{-1}$.

		Split [a]				Unified [b]				Absolute [c]			
		GROMACS (RF)		GROMACS (PME)		GROMACS (RF)		GROMACS (PME)		GROMACS (RF)		GROMACS (PME)	
Transformations		$\Delta\Delta G$	SEM	$\Delta\Delta G$	SEM	$\Delta\Delta G$	SEM	$\Delta\Delta G$	SEM	$\Delta\Delta G$	SEM	$\Delta\Delta G$	SEM
methane	ethane	-0.20	0.02	-0.18	0.01	-0.19	0.02	-0.17	0.01	0.06	0.01	0.04	0.01
ethane	methane	0.19	0.02	0.18	0.02	0.248	0.003	0.236	0.002				
ethane	methanol	-6.05	0.01	-6.11	0.01	-7.04	0.01	-7.13	0.02	-5.83	0.01	-5.98	0.01
methanol	ethane	6.05	0.00	6.12	0.00	7.25	0.01	7.32	0.01				
methane	methanol	-6.21	0.05	-6.24	0.05	-7.21	0.02	-7.29	0.02	-5.77	0.01	-5.95	0.01
methanol	methane	6.39	0.01	6.42	0.01	7.53	0.02	7.61	0.01				
methane	2-methylfuran	-3.13	0.01	-3.08	0.01	-3.14	0.01	-3.10	0.01	-2.87	0.01	-2.95	0.01
2-methylfuran	methane	3.08	0.01	3.03	0.01	3.14	0.01	3.09	0.01				
methane	toluene	-3.25	0.01	-3.20	0.01	-3.33	0.01	-3.31	0.01	-2.97	0.01	-3.16	0.01
toluene	methane	3.27	0.03	3.26	0.02	3.30	0.01	3.29	0.01				
methane	2-methylindole	-8.79	0.03	-8.80	0.01	-8.67	0.03	-8.73	0.04	-8.44	0.02	-8.79	0.02
2-methylindole	methane	8.74	0.02	8.76	0.03	8.76	0.01	8.83	0.03				
methane	neopentane	-0.21	0.02	-0.15	0.05	-0.24	0.01	-0.08	0.01	0.18	0.01	0.14	0.01
neopentane	methane	0.36	0.02	0.31	0.03	0.29	0.02	0.22	0.03				
methane	neopentane2	-0.05	0.02	0.04	0.02	-0.03	0.01	0.04	0.02	0.18	0.01	0.14	0.01
neopentane2	methane	-0.03	0.01	-0.04	0.01	-0.05	0.01	-0.07	0.01				
2-cyclopentanylindole	7-cyclopentanylindole	-0.07	0.02	-0.03	0.03	-0.09	0.05	-0.1	0.1	-0.02	0.05	0.02	0.02
7-cyclopentanylindole	2-cyclopentanylindole	0.12	0.06	0.20	0.04	0.12	0.08	0.2	0.1				

[a] results obtained from alchemical transformations with electrostatic and bonded scaling separate from vdW parameter change.
[b] results obtained from alchemical transformation with all parameters scaling together.
[c]results obtained from absolute free energy calculations.

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 hydrogen bond constraining algorithms such as SHAKE or LINCS. Successful simulations seem to require turning off the constraint and decreasing the time step.

3.4 Sire/SOMD

Figure 3 compares relative free energy of hydration $\Delta\Delta G$ with relative $\Delta\Delta G$ estimations from absolute free energy calculations, according to the best protocol for Sire/SOMD. A very good agreement is present between both calculations, highlighting consistency and reproducibility of results in Sire/SOMD itself, with a correlation index R^2=0.99 and a mean unsigned error MUE=0.10 kcal mol^{-1}

16.we could write a 95% confidence interval MUE: $0.08 < 0.11 < 0.14\,\mathrm{kcal\,mol^{-1}}$

To achieve reproducibility in Sire/SOMD and among all the other codes, the role of constraints was crucial and a new kind of constraint was employed: unperturbed hydrogen bond constraint. Here, all the unperturbed hydrogen atoms'bonds are constrained, while the perturbed hydrogen atoms, namely atoms which undergo a transformation, are unconstrained and their masses are equal to the heavier mass of the two end states. In this way symmetry between forward and backward transformations is guaranteed, and a time step of 2 fs can be employed.

17.maybe add figure to SI explaining how these constraints are used

This new constraint overcomes issues encountered with Sire/SOMD starting protocol. Usually, RAFE calculations were performed with all bonds constrained, in order to have time step longer than 2 fs,[75] enhancing sampling with a reasonable computational cost. However, from a technical point of view, a constraint in Sire/SOMD sets the bond energy to zero. This may be a valid approximation for small and rigid molecules, but a systematic offset is introduced which necessarily increases with every additional perturbation introduced. Indeed, looking at neopentane → methane (centrally mapped) this bias gives

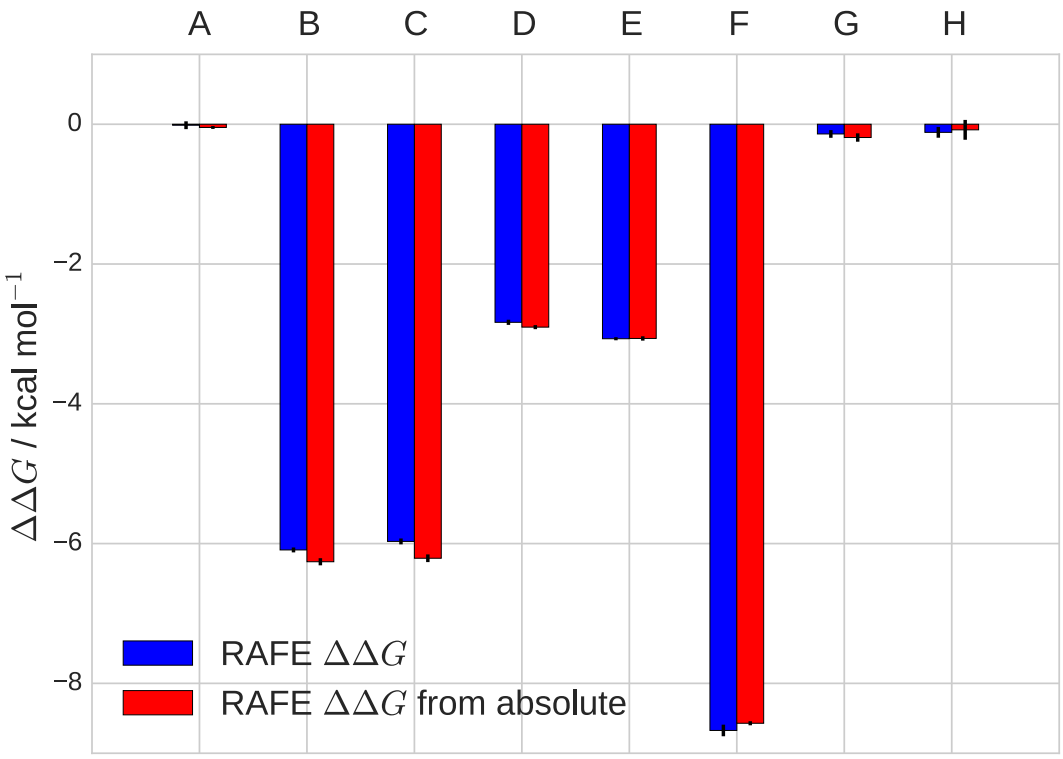

Figure 3: Relative free energy of hydration $\Delta\Delta G$, computed with RAFE calculations, compared with $\Delta\Delta G$ derived from absolute free energy calculations for A: methane to ethane, B: ethane to methanol, C: methane to methanol, D: methane to 2-methylfuran, E: methane to toluene, F: methane to 2-methylindole, G: methane to neopentane, H: 2-cyclopentanylindole to 7-cyclopentanylindole.

a RAFE $\Delta\Delta G$=$(2.04 \pm 0.01)\,\mathrm{kcal\,mol^{-1}}$ in contrast to ΔG=$(-0.19 \pm 0.06)\,\mathrm{kcal\,mol^{-1}}$ from the absolute calculations, as tab. S5nd fig. S5how. Furthermore, neopentane $\rightarrow$ methane (terminally mapped) shows a $\Delta\Delta G$ =$(0.59 \pm 0.01)\,\mathrm{kcal\,mol^{-1}}$. As the free energy must be independent of the chosen path this discrepancy denotes a problem with using constraints on transformed atoms. This problem has been discussed by Pearlman[76] and Boresch,[27,49] as a missing bond-stretching term, which can be estimated as:

$$\Delta G = RT \ln\left(\frac{r_\mathrm{i}}{r_\mathrm{f}}\right) \tag{5}$$

where r_i and r_f are the initial and final bond lengths, respectively. Although for small bond length changes this error may be negligible, it becomes pivotal for perturbations of hydrogen atoms to carbon, nitrogen or oxygen, where the bond length changes of several tenths of an Å.

After the implementation of the our new constraint scheme, a comparison with simulations without constraints has been run. Figure S5hows that both protocols attain the same results with a final MUE = 0.09 kcal mol^{-1}.

Tab 3 shows the MUE between Sire/SOMD, GROMACS employing reaction field (GROMACS RF), GROMACS with PME (GROMACS PME) and AMBER (alternatively: Tab 4 shows the MUE comparison with 95% of confidence interval between Sire/SOMD, GROMACS with reaction field (GROMACS RF), GROMACS with PME (GROMACS PME) and AMBER). As regards reaction field methods, Sire/SOMD implements a Barker Watts reaction field (BWRF) with a atom based cutoff, as explained in subsection 2.3, while GROMACS adopts a BWRF with a charge group cutoff. Despite this technical difference, Sire/SOMD and GROMACS RF produce comparable results with a MUE of 0.18 kcal mol^{-1}. In]the methane $\rightarrow$ neopentane transformation is still matter of problems. In this case, Sire/SOMD is able to maintain consistency between central and terminal mapping, as shown in tab S4 while GROMACS RF displays a $\Delta\Delta G$ of (-0.20 ± 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. As regards PME methods, Sire/SOMD presents a MUE = 0.18 kcal mol^{-1} and MUE = 0.23 kcal mol^{-1} with GROMACS PME and AMBER respectively, achieving a very good agreement between $\Delta\Delta G$ predictions.

18.it may be good to say somewhere how much was the MUE with the initial standard protocol

19.GROMACS RF in vacuum uses a dielectric of 1.0, this causes problems for Sire vacuum simulations

Overall, the Sire/SOMD free energy estimations are in good agreement with the other MD packages, as the MUE suggests. Reaction field and PME results are in good agreement. In particular, the unperturbed hydrogen bonds constraint allows a timestep of 2 fs

and Sire/SOMD RAFE simulations can be carried out in one single step.

Table 3: MUE comparison between Sire/SOMD, GROMACS with Reaction Field (GROMACS RF), GROMACS with PME (GROMACS PME) and AMBER in terms of MUE (kcal mol^{-1}.

Package	Sire/SOMD	Gromcs RF	GROMACS PME	AMBER
Sire/SOMD	0.00	0.18	0.18	0.23
GROMACS RF	0.18	0.00	0.04	0.14
GROMACS PME	0.18	0.04	0.00	0.14
AMBER	0.23	0.14	0.14	0.00

Table 4: MUE with 95% confidence interval between Sire/SOMD, GROMACS with Reaction Field (GROMACS RF), GROMACS with PME (GROMACS PME) and AMBER in terms of MUE (kcal mol^{-1}.

Package	Sire/SOMD	Gromcs RF	GROMACS PME	AMBER
Sire/SOMD	0.02< 0.04< 0.05	0.17< 0.19< 0.21	0.16< 0.18< 0.20	0.21< 0.23< 0.25
GROMACS RF	0.17< 0.18< 0.19	0.01< 0.02< 0.03	0.04< 0.05< 0.06	0.13< 0.14< 0.15
GROMACS PME	0.17< 0.18< 0.19	0.04< 0.05< 0.06	0.01< 0.02< 0.03	0.14< 0.15< 0.16
AMBER	0.22< 0.23< 0.24	0.13< 0.14< 0.15	0.14< 0.15< 0.16	0.01< 0.02< 0.03

4 Discussion

20.recommended protocols

21.protocols to avoid

22.lessons learned

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

24.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?

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

26.further investigation required: binding RAFEs?

ToDo

Acknowledgement

HHL is supported through an EPSRC provided SLA, funding the core support of CCPBioSim. 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) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *The Journal of Physical Chemistry* **1994**, *98*, 1487–1493.

(26) 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.

(27) 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.

(28) 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.

(29) Shobana, S.; Roux, B.; Andersen, O. S. Free Energy Simulations: Thermodynamic Reversibility and Variability. *The Journal of Physical Chemistry B* **2000**, *104*, 5179–5190.

(30) 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.

(31) 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.

(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) 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.

(35) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.

(36) 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.

(37) 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.

(38) Woods, C.; Mey, A. S.; Calabro, G.; Michel, J. Sire molecular simulations framework. 2016.

(39) 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.

(40) 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.

(41) 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.

(42) 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.

(43) 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.

(44) 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.

(45) 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.

(46) Cross, A. J. Influence of hamiltonian parameterization on convergence of kirkwood free energy calculations. *Chemical Physics Letters* **1986**, *128*, 198 – 202.

(47) 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.

(48) 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.

(49) 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.

(50) 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.

(51) 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.

(52) 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.

(53) 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.

(54) 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.

(55) 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.

(56) 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.

(57) 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.

(58) 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.

(59) Barker, J.; Watts, R. Monte Carlo studies of the dielectric properties of water-like models. *Molecular Physics* **1973**, *26*, 789–792.

(60) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* **1980**, *72*, 2384–2393.

(61) 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.

(62) Zwanzig, R. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

(63) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *Journal Of Chemical Physics* **1935**, *3*, 300–313.

(64) 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*.

(65) 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*.

(66) 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.

(67) 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.

(68) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*.

(69) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*, 124105.

(70) 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.

(71) 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.

(72) 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.

(73) Pitera, J. W.; Gunsteren, W. F. v. A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations. *Mol. Simul.* **2002**, *28*, 45–65.

(74) Anwar, J.; Heyes, D. M. Robust and Accurate Method for Free-Energy Calculation of Charged Molecular Systems. *J. Chem. Phys.* **2005**, *122*, 224117.

(75) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of chemical theory and computation* **2015**, *11*, 1864–1874.

(76) 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