JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

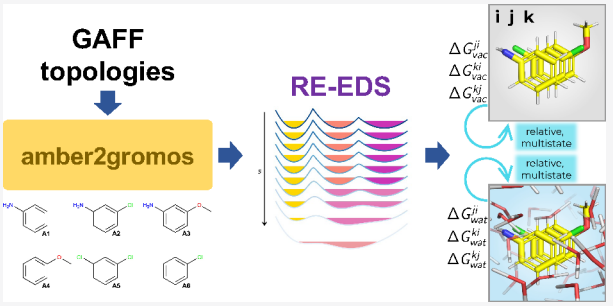
Replica-Exchange Enveloping Distribution Sampling Using Generalized AMBER Force-Field Topologies: Application to Relative Hydration Free-Energy Calculations for Large Sets of Molecules

Salomé R. Rieder, Benjamin Ries, Kay Schaller, Candide Champion, Emilia P. Barros, Philippe H. Hünenberger,* and Sereina Riniker*

Cite This: *J. Chem. Inf. Model.* 2022, 62, 3043−3056 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Free-energy differences between pairs of end-states can be estimated based on molecular dynamics (MD) simulations using standard pathway-dependent methods such as thermodynamic integration (TI), free-energy perturbation, or Bennett's acceptance ratio. Replica-exchange enveloping distribution sampling (RE-EDS), on the other hand, allows for the sampling of multiple end-states in a single simulation without the specification of any pathways. In this work, we use the RE-EDS method as implemented in GROMOS together with generalized AMBER force-field (GAFF) topologies, converted to a GROMOS-compatible format with a newly developed GROMOS++ program *amber2gromos*, to compute relative hydration free energies for a series of benzene derivatives. The results obtained with RE-EDS are compared to the experimental data as well as calculated values from the literature. In addition, the estimated free-energy differences in water and in vacuum are compared to values from TI calculations carried out with GROMACS. The hydration free energies obtained using RE-EDS for multiple molecules are found to be in good agreement with both the experimental data and the results calculated using other free-energy methods. While all considered free-energy methods delivered accurate results, the RE-EDS calculations required the least amount of total simulation time. This work serves as a validation for the use of GAFF topologies with the GROMOS simulation package and the RE-EDS approach. Furthermore, the performance of RE-EDS for a large set of 28 end-states is assessed with promising results.

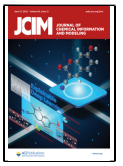


INTRODUCTION

In recent years, free-energy calculations (either absolute or relative) based on classical molecular dynamics (MD) simulations have started to play an increasingly important role in the field of computer-aided drug design.[1−9] There exist many well-established pairwise free-energy methods such as thermodynamic integration (TI),[10] free-energy perturbation (FEP),[11] Bennett's acceptance ratio (BAR),[12] and multistate BAR (MBAR).[13] Approaches such as multisite λ-dynamics[14−16] and enveloping distribution sampling (EDS)[17,18] enable the calculation of pairwise free-energy differences for multiple end-states from a single simulation. While λ-dynamics uses the coupling parameter λ as a dynamic variable to connect the end-states, EDS is a pathway-independent method that samples a reference state "enveloping" all end-states. Recently, replica-exchange EDS (RE-EDS)[19−21] and accelerated EDS (A-EDS)[22,23] have been developed as extensions of EDS to simplify the parameter optimization and improve the performance of EDS. Both methods are implemented in the GROMOS software package.[24]

Apart from the free-energy method, the quality of the underlying force field is crucial for the accuracy of free-energy calculations and of MD simulations in general.[25−28] In the past years, various tools have been developed to automate the otherwise laborious task of topology generation for small molecule ligands, such as antechamber,[29−31] the automated topology builder (ATB),[32,33] the fragment-based CombiFF,[34,35] general automated atomic model parametrization (GAAMP),[36,37] LigParGen,[38] open force field (SMIRNOFF and OpenFF),[39,40] ParamChem,[41−43] PRODRG,[44] R.E.D.,[45] or SwissParam.[46] MD simulation engines such as AMBER,[47−49] CHARMM,[50,51] GROMACS,[52,53] GROMOS,[24,54] or OpenMM[55,56] require specific file formats to describe the system topology and coordinates. In addition, there are also small differences in the functional form of the force fields or in the units used by different MD engines.[28] In many cases, tools are already available to translate between some of the different

Received: April 1, 2022
Published: June 8, 2022



file formats, enabling, e.g., the use of AMBER topologies with GROMACS.[57−61]

The calculation of (absolute or relative) hydration free energies serves as a straightforward test case to assess and compare the quality of different free-energy methods and to validate force fields.[62,63] Thanks to databases such as FreeSolv,[64,65] the Minnesota solvation database,[66] or the ATB server,[32,33] ample reference data is available for both experimental results as well as calculated values obtained with different force fields and free-energy methods. Furthermore, the calculation of hydration free energies is computationally far less expensive than, for example, that of binding free energies.

In this work, a newly introduced GROMOS++[67] program *amber2gromos*, developed by the authors of this study, is described. It translates a topology from the AMBER prmtop[68] file format to a GROMOS topology, enabling users of the GROMOS MD engine to simulate systems with the AMBER or generalized AMBER (GAFF[30]) force fields. Extension to the OpenFF[39,40] family of force fields is straightforward. In the following, the underlying differences between the AMBER and GROMOS force fields are discussed, and the necessary conversions are described in detail. The correctness of the topology conversion is validated by comparison of single-molecule simulations in vacuum using GROMACS or GROMOS. Furthermore, two sets of small benzene derivatives are assembled from the FreeSolv[64,65] database: a small set of six molecules, labeled A, and a larger set of 28 molecules, labeled B. For these molecules, relative hydration free energies are calculated with TI[10] in GROMACS (set A only) and with RE-EDS[19−21] in GROMOS (sets A and B). The results are compared to each other and to the experimental and calculated values reported in the FreeSolv[64,65] database.

THEORY

Differences between the AMBER and GROMOS Force Fields. The use of an automation tool such as AmberTools[49] (i.e., *antechamber*[29−31] and *tleap*) simplifies the process of topology generation for small organic molecules considerably. In order to use GAFF[30] topologies in GROMOS, they have to be converted to a GROMOS-compatible file format. There are several differences between the AMBER/GAFF and GROMOS[69−71] force fields.

First, GAFF is an all-atom force field, whereas most GROMOS (compatible) force fields use united atoms (i.e., implicit hydrogens) for the aliphatic CH_n groups, to reduce the computational cost.[28] The GROMOS force fields are usually parametrized with the simple point-charge (SPC)[72] water model, whereas the AMBER force-field family is parametrized with the TIP3P[73] water model.[28] A minor difference is the use of different units, e.g., nm, degrees, and kJ mol^{-1} in GROMOS versus Å, radians, and kcal mol^{-1} in AMBER.[49,74]

Second, there are several differences in the potential-energy function, i.e., the functional form of the force fields. Here, the subscripts "*A*" (AMBER) and "*G*" (GROMOS) are used to distinguish the terms/parameters of the two force-field families. In AMBER, harmonic bond stretching and bond-angle bending terms are used[28,30,49]

$$V_{A,i}^{\mathrm{bond,harm}}(d_i) = K_{A,i}^{b,\mathrm{harm}}(d_i - d_{0,i})^2 \tag{1}$$

$$V_{A,i}^{\mathrm{angle,harm}}(\theta_i) = K_{A,i}^{a,\mathrm{harm}}(\theta_i - \theta_{0,i})^2 \tag{2}$$

where d_i is the distance between two bonded atoms, $K_{A,i}^{b,\mathrm{harm}}$ is the harmonic bond force constant, $d_{0,i}$ is the equilibrium distance, θ_i is the angle formed by three bonded atoms, $K_{A,i}$ is the harmonic angle force constant, and $\theta_{0,i}$ is the reference bond angle. In GROMOS, harmonic bond stretching and bond-angle bending are also implemented. However, quartic bond stretching and cosine-harmonic bond-angle bending are used by default to increase computational efficiency.[28] They are defined as follows[75]

$$V_{G,i}^{\mathrm{bond,harm}}(d_i) = \frac{1}{2}K_{G,i}^{b,\mathrm{harm}}(d_i - d_{0,i})^2 \tag{3}$$

$$V_{G,i}^{\mathrm{angle,harm}}(\theta_i) = \frac{1}{2}K_{G,i}^{a,\mathrm{harm}}(\theta_i - \theta_{0,i})^2 \tag{4}$$

$$V_{G,i}^{\mathrm{bond,quart}}(d_i) = \frac{1}{4}K_{G,i}^{b,\mathrm{quart}}(d_i^2 - d_{0,i}^2)^2 \tag{5}$$

$$V_{G,i}^{\mathrm{angle,cos}}(\theta_i) = \frac{1}{2}K_{G,i}^{a,\mathrm{cos}}(\cos(\theta_i) - \cos(\theta_{0,i}))^2 \tag{6}$$

with the parameters being defined analogously to the AMBER parameters. It is important to note that for AMBER/GAFF, the factor 1/2 in the harmonic bond stretching and bond-angle bending equations is already included in the force constants $K_{A,i}^{b,\mathrm{harm}}$ and $K_{A,i}^{a,\mathrm{harm}}$ (compare eqs 1 and 2 with eqs 3 and 4).[30,49] The harmonic force constants can be converted to the quartic and cosine-harmonic force constants, respectively, as[75]

$$K_{G,i}^{b,\mathrm{quart}} = \frac{K_{G,i}^{b,\mathrm{harm}}}{2d_{0,i}^2} \tag{7}$$

$$K_{G,i}^{a,\mathrm{cos}} = \frac{2k_\mathrm{B}T_\mathrm{eff}}{\left[\cos\left(\theta_{0,i} + \left(\frac{k_\mathrm{B}T_\mathrm{eff}}{K_{G,i}^{a,\mathrm{harm}}}\right)^{1/2}\right) - \cos\theta_{0,i}\right]^2} + \frac{2k_\mathrm{B}T_\mathrm{eff}}{\left[\cos\left(\theta_{0,i} - \left(\frac{k_\mathrm{B}T_\mathrm{eff}}{K_{G,i}^{a,\mathrm{harm}}}\right)^{1/2}\right) - \cos\theta_{0,i}\right]^2} \tag{8}$$

where k_B is the Boltzmann constant, and T_eff is an effective absolute temperature for the conversion (e.g., 300 K).[75] Another difference between AMBER and GROMOS is the potential-energy function used for the out-of-plane distortions. In AMBER, the same function is used for both proper and improper dihedral changes[28,30,49]

$$V_{A,i}^{\mathrm{tors/imp}}(\theta_i) = K_{A,i}^{\mathrm{tors/imp}}[1 + \cos(m\theta_i - \theta_{0,i})] \tag{9}$$

In contrast, GROMOS uses different functional forms for proper and improper dihedral changes[28,75]

$$V_{G,i}^{\mathrm{tors}}(\theta_i) = K_{G,i}^{\mathrm{tors}}[1 + \cos(m\theta_i - \theta_{0,i})] \tag{10}$$

$$V_{G,i}^{\mathrm{imp}}(\xi_i) = \frac{1}{2}K_{G,i}^{\mathrm{imp}}(\xi_i - \xi_{0,i})^2 \tag{11}$$

The improper term is also used for out-of-tetrahedron distortions around the CH_1 united atom. Both force-field families use the Lennard-Jones functional form for the van der Waals interactions[30,49,75]

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

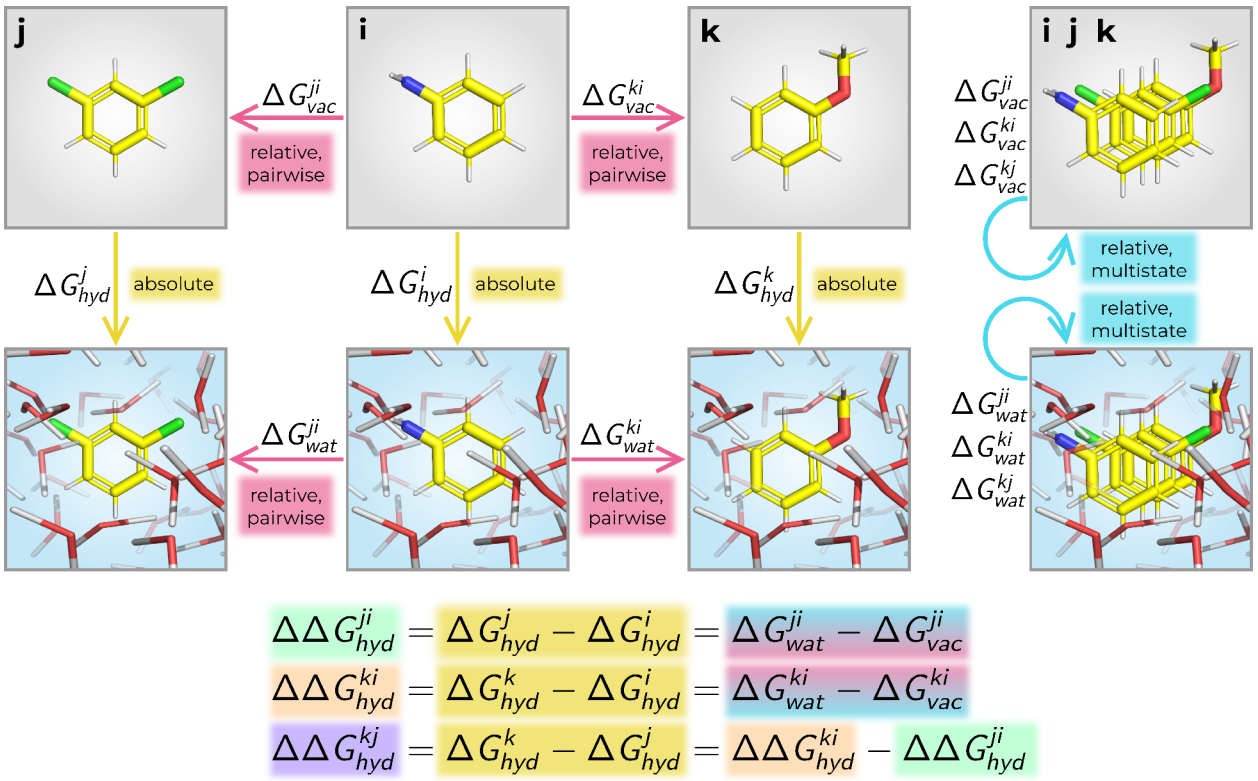


Figure 1. Thermodynamic cycle to calculate relative hydration free energies $\Delta\Delta G_{\mathrm{hyd}}$ for three small molecules i (aniline), j (1,3-dichlorobenzene), and k (anisole). Here, $\Delta G^{i}_{\mathrm{hyd}}$ is the hydration free energy of molecule i, $\Delta G^{ji}_{\mathrm{vac}}$ is the free-energy difference between molecules i and j in vacuum, $\Delta G^{ji}_{\mathrm{wat}}$ is the free-energy difference between the two molecules in water, and $\Delta\Delta G^{ji}_{hyd}$ is the hydration free-energy difference between the two molecules (relative hydration free energy). The free-energy difference between two molecules can be calculated from multiple pairwise simulations (as shown on the left, e.g., with TI) or from one simulation with multiple molecules (as shown on the right, e.g., with RE-EDS).

$$V^{\mathrm{vdW}}_{A,ij} = \left(\frac{A_{ij}}{r^{12}_{ij}} - \frac{B_{ij}}{r^{6}_{ij}}\right) \tag{12}$$

$$V^{\mathrm{vdW}}_{G,ij} = \left(\frac{C_{12,ij}}{r^{12}_{ij}} - \frac{C_{6,ij}}{r^{6}_{ij}}\right) \tag{13}$$

where r_{ij} is the (minimum-image) distance between atoms i and j, A_{ij} and $C_{12,ij}$ are the repulsion coefficients, and B_{ij} and $C_{6,ij}$ are the dispersion coefficients. There are, however, differences in the combination rules used (geometric in GROMOS and Lorentz−Berthelot in AMBER[28]) and the handling of third-neighbor interactions. In AMBER/GAFF, the Lennard-Jones 1,4-interactions are scaled by a factor 1/2,[30,49] whereas GROMOS force fields contain a special set of parameters[76] (CS_{12} and CS_6) for such interactions, typically involving a reduced repulsion coefficient. Using a scaling factor or reduced interaction parameters for third-neighbor interactions avoids having a too large repulsion in *gauche* conformations (relative to trans conformations).[75] In AMBER/GAFF, electrostatic 1,4-interactions are also scaled by a factor 1/1.2.[30,49] Such a scaling is not applied in GROMOS, although in some cases, third neighbors are excluded completely, for example for atoms that are in or attached to an aromatic ring.[76] Furthermore, in AMBER/GAFF, the factor[68]

$$k_e^{1/2} = \left(\frac{1}{4\pi\varepsilon_0}\right)^{1/2} = 18.2223\ [\mathrm{kcal\ \AA\ mol^{-1}\ e^{-2}}]^{1/2} \tag{14}$$

is included in the atomic charges for computational efficiency. Here, k_e is Coulomb's constant, and ε_0 is the permittivity of vacuum.

Relative Hydration Free Energies. The hydration free energy quantifies the free-energy change when a molecule is transferred from gas to water.[62,77] In this work, relative hydration free energies $\Delta\Delta G_{\mathrm{hyd}}$ are used to compare different free-energy methods against each other and against experimental values (Figure 1). For three molecules i, j, and k, it holds that[78]

$$\Delta\Delta G^{ji}_{\mathrm{hyd}} = \Delta G^{j}_{\mathrm{hyd}} - \Delta G^{i}_{\mathrm{hyd}} = \Delta G^{ji}_{\mathrm{wat}} - \Delta G^{ji}_{\mathrm{vac}} \tag{15}$$

$$\Delta\Delta G^{ki}_{\mathrm{hyd}} = \Delta G^{k}_{\mathrm{hyd}} - \Delta G^{i}_{\mathrm{hyd}} = \Delta G^{ki}_{\mathrm{wat}} - \Delta G^{ki}_{\mathrm{vac}} \tag{16}$$

$$\Delta\Delta G^{kj}_{\mathrm{hyd}} = \Delta\Delta G^{ki}_{\mathrm{hyd}} - \Delta\Delta G^{ji}_{\mathrm{hyd}} \tag{17}$$

where $\Delta G^{i}_{\mathrm{hyd}}$ is the hydration free energy of molecule i, $\Delta G^{ji}_{\mathrm{vac}}$ is the free-energy difference between molecules i and j in vacuum, $\Delta G^{ji}_{\mathrm{wat}}$ is the free-energy difference between the two molecules in water, and $\Delta\Delta G^{ji}_{hyd}$ is the hydration free-energy difference between the two molecules (relative hydration free energy).

In classical MD simulations, hydration free energies are typically calculated with so-called alchemical free-energy methods.[64,79−81] Such methods transform a molecule (or its interaction with the environment) into another one via nonphysical pathways.[64] The following sections give a brief overview of the two free-energy methods used in the present study.

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

Thermodynamic Integration (TI). TI is a well-established method to calculate free-energy differences.[10] For two end-states A and B, and using a linear coupling scheme, the potential energy of the system is defined as

$$V(\mathbf{r};\lambda) = (1-\lambda)V_A(\mathbf{r}) + \lambda V_B(\mathbf{r}) \tag{18}$$

At $\lambda = 0$ and $\lambda = 1$, the potential energy corresponds to that of end-state A and end-state B, respectively. This defines a λ-dependent path between the two end-states. After carrying out independent simulations at discrete λ-values between 0 and 1, the free-energy difference between states A and B can be estimated as[10]

$$\Delta G_{BA} = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \tag{19}$$

Replica-Exchange Enveloping Distribution Sampling (RE-EDS). RE-EDS is a multistate free-energy method,[19−21] which combines Hamiltonian replica exchange[82,83] (RE) with enveloping distribution sampling (EDS).[17,18] In EDS, a reference state V_R is defined based on N end-states as

$$V_R(\mathbf{r};s,\mathbf{E}^R) = -\frac{1}{\beta s}\ln\left[\sum_{i=1}^{N} e^{-\beta s(V_i(\mathbf{r})-E_i^R)}\right] \tag{20}$$

where s is the smoothness parameter ($s > 0$), $\mathbf{E}^R$ is a vector of energy offsets, and $\beta = 1/(k_B T)$.

At high s-values (∼1.0), due to the negative exponent, the end-state with the lowest value of $V_i(\mathbf{r}) - E_i^R$ contributes the most to the sampling of V_R. As s decreases, it "smoothes" the potential-energy landscape such that all end-states contribute to the reference state, leading to an unphysical intermediate situation,[19] referred to as "undersampling".[84] The energy offsets control the contribution of the end-states to the reference-state potential. Optimal energy offsets ensure equal weights of all end-states in the reference state.

The free-energy difference between any pair of end-states can then be obtained from a single simulation as[17,18]

$$\Delta G_{BA} = -\frac{1}{\beta}\ln\frac{\langle e^{-\beta(V_B-V_R)}\rangle_R}{\langle e^{-\beta(V_A-V_R)}\rangle_R} \tag{21}$$

For EDS simulations, it is essential to choose an optimal s-value along with optimal energy offsets to achieve adequate sampling of all end-states, which is nontrivial for more than two end-states.[84] RE-EDS enhances the sampling and simplifies the parameter optimization by simulating several replicas with different s-values in parallel and performing replica exchanges between them.[19]

In recent studies, RE-EDS was successfully used to calculate relative binding and hydration free energies for molecules containing relatively large structural changes (i.e., R-group modifications and core-hopping transformations such as ring opening/closing and ring size changes).[21,85]

METHODS

Implementation of *amber2gromos*. The *amber2gromos* program is a novel C++[86] tool integrated into the GROMOS++ package of programs.[67] It converts AMBER topologies to a GROMOS-compatible file format. For a given AMBER topology, the program parses the input topology file, converts the force-field parameters, and outputs a GROMOS topology. The conversion steps of the program are outlined in the Supporting Information. An example of an AMBER topology for aniline and the resulting GROMOS topology translated with *amber2gromos* can be found in the Supporting Information in Listings 1 and 2, respectively.

The conversion process described in the Supporting Information generates a valid GROMOS topology from a given AMBER topology. However, there is still a difference between AMBER and GROMOS in the handling of the 1,4-electrostatic interactions (i.e., scaling by a factor 1/1.2[30,49] in AMBER). As 1,4-electrostatic interactions are not scaled in the GROMOS force fields, the scaling option is not supported within a GROMOS topology. Therefore, some changes had to be made to the GROMOS[24] source code. A new block type "AMBER" was added to the GROMOS input file, which contains a switch (0 = off, 1 = on) for the scaling of the electrostatic third-neighbor interactions together with the scaling parameter (e.g., 1.2, for scaling by 1/1.2). When the switch is off, the scaling parameter is set to 1.0 so that no scaling is applied. When a reaction field (RF) correction[87] is used in GROMOS to account for long-range electrostatic interactions, these interactions are calculated as[28,75]

$$V^{\mathrm{ele}} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0}\left[\frac{1}{r_{ij}} - \frac{C_{RF} r_{ij}^2}{2R_{RF}^3} - \frac{1-0.5C_{RF}}{R_{RF}}\right] \tag{22}$$

where q_i and q_j are the charges of atoms i and j, ε_0 is the permittivity of vacuum, and R_{RF} is the cutoff distance.[28,75,76,87] Here, C_{RF} is a constant characterizing the effect of the RF continuum given by[75,87]

$$C_{RF} = \frac{(2\varepsilon_{cs}-2\varepsilon_{RF})(1+\kappa_{RF}R_{RF}) - \varepsilon_{RF}(\kappa_{RF}R_{RF})^2}{(\varepsilon_{cs}+2\varepsilon_{RF})(1+\kappa_{RF}R_{RF}) + \varepsilon_{RF}(\kappa_{RF}R_{RF})^2} \tag{23}$$

where ε_{cs} is the relative permittivity of the medium in which the simulation is performed, ε_{RF} is the RF permittivity, and κ_{RF} is the inverse Debye screening length.[75] The electrostatic interactions in the GROMOS[24,88] source code were changed to

$$V^{\mathrm{ele}} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0}\left[\frac{s_{ij}}{r_{ij}} - \frac{C_{RF} r_{ij}^2}{2R_{RF}^3} - \frac{1-0.5C_{RF}}{R_{RF}}\right] \tag{24}$$

where s_{ij} corresponds to the AMBER scaling factor when i and j are third neighbors, and 1.0 otherwise. Note that in this expression, only the direct Coulombic interactions are scaled.[88,89]

As the AMBER force fields do not use charge groups, by default, all the atoms in a molecule/residue are assigned to the same charge group for the simulations with RF electrostatics[87] in GROMOS (which is only appropriate for small molecules). In addition, there is also the option to consider each atom as defining its own charge group (resulting in an atom-based cutoff). The user can also prepare a so-called charge-group file, defining which (consecutive) atoms should be assigned to the same charge group. There are currently ongoing efforts to combine RE-EDS with the shifting-function based scheme developed by Kubincóva et al.[90] such that an atom-based cutoff can be employed in the RF electrostatics without cutoff artifacts. This will become the default choice when using the AMBER (and OpenFF) force fields in GROMOS in the near future, mitigating the requirement to define charge groups altogether.

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

RE-EDS Pipeline. Recently, Ries et al.[21] presented an improved pipeline to perform relative free-energy calculations with RE-EDS. The RE-EDS pipeline can be divided into three main phases: parameter exploration, parameter optimization, and production. The parameter exploration phase is used to generate relevant configurations for all end-states, to determine a lower bound for the *s*-values that ensures undersampling, and to obtain initial estimates for the energy offsets. During the parameter optimization, the distribution of *s*-values and the energy offsets E_i^R are refined such that all end-states are sampled nearly equally at $s = 1$. Finally, a production run is performed to calculate the free-energy differences between all pairs of end-states simultaneously. The entire workflow can be executed using the Python3[91] *reeds* module.[92] In this work, the RE-EDS pipeline was applied to the calculation of relative hydration free energies of benzene derivatives.

Data Sets. The main goals of this work were to validate the topology conversion with *amber2gromos* and the accompanying changes in the GROMOS MD engine, as well as to investigate the performance of RE-EDS for systems with many end-states. To this end, two sets of benzene derivatives with experimental hydration free energies in the FreeSolv[64,65] database were selected. This test system was chosen due to the common benzene core, the relatively small size (which limits deviations due to sampling issues), and the availability of calculated and experimental reference data. The potential of RE-EDS to handle larger perturbations has already been demonstrated in previous publications.[21,85]

Set A: Six Benzene Derivatives. As a first test set, six benzene derivatives were selected (Figure 2). The number of

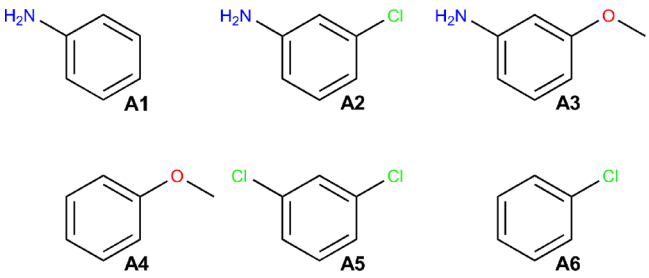


Figure 2. Set A consists of six benzene derivatives, selected from the FreeSolv[64,65] database. The molecule indices, the FreeSolv identifiers, the SMILES strings, and the names of the molecules can be found in Table S1 in the Supporting Information.

end-states was deliberately kept small to efficiently test the implementation. The mol2 and frcmod files provided by FreeSolv[64,65] were used to generate AMBER topologies using *tleap* (AmberTools16).[49] The topologies were then converted to GROMOS format using *amber2gromos* and to GROMACS[53,93] format using ParmEd.[57] The force-field parameters of the original AMBER topologies and of the generated GROMOS/GROMACS topologies were compared manually. In addition, a single energy evaluation was performed for the individual molecules in vacuum/water using both GROMOS and GROMACS. It was verified that the covalent and nonbonded energy terms calculated by the two different MD engines were nearly identical. Longer MD simulations of 5 ns were then carried out for each molecule in vacuum to compare properties such as the system temperature and the different energy terms.

Next, the free-energy differences in vacuum/water were calculated for all 15 molecule pairs from RE-EDS simulations containing six end-states. Complementary single-topology pairwise TI calculations were performed using GROMACS. The relative hydration free energies $\Delta\Delta G_{\mathrm{hyd}}^{ji} = \Delta G_{\mathrm{wat}}^{ji} - \Delta G_{\mathrm{vac}}^{ji}$ were calculated from the RE-EDS and TI calculations. They were compared to the relative hydration free energies $\Delta\Delta G_{\mathrm{hyd}}^{ji} = \Delta G_{\mathrm{hyd}}^{j} - \Delta G_{\mathrm{hyd}}^{i}$ obtained from the experimental and calculated hydration free energies reported in the FreeSolv[64,65] database.

Set B: 28 Benzene Derivatives. To investigate the performance of RE-EDS for a larger number of end-states, the set of benzene derivatives was increased to 28 (Figure 3). This extended test set serves as a proof of principle that RE-EDS can be used to calculate free-energy differences for systems with larger numbers of end-states. It is currently the largest set of end-states considered in an RE-EDS simulation. Previous studies involved systems with five,[21] six,[85] nine,[20] and ten[19,85] end-states. The free-energy differences were calculated for all 378 molecule pairs in vacuum/water from RE-EDS simulations containing 28 end-states. Analogous to set A, the relative hydration free energies from the RE-EDS calculation were compared to the hydration free energies reported in the FreeSolv[64,65] database.

To further investigate the performance of RE-EDS for such a large set of end-states, set B was subdivided into two smaller subsets, Ba and Bb. Subset Ba consisted of molecules B1−B14, and subset Bb consisted of molecule B1 together with molecules B15−B28. For the two subsets, RE-EDS simulations were carried out in vacuum/water to calculate the relative hydration free energies for all end-state pairs in both sets. The relative hydration free energies between the molecule pairs *j*-*k* that were not in the same subset were calculated via molecule B1, which was present in both subsets, as

$$\Delta\Delta G_{\mathrm{hyd}}^{kj} = \Delta\Delta G_{\mathrm{hyd}}^{k,B1} - \Delta\Delta G_{\mathrm{hyd}}^{j,B1} \tag{25}$$

Simulation Details. The AMBER topologies were generated using AmberTools16[49] and the GAFF 1.7[30] force field with the mol2 and frcmod files provided in the FreeSolv[64,65] database as a starting point. The atomic charges were generated using the AM1-BCC[94,95] approach. The input files for the single-topology TI calculations in GROMACS were prepared with FESetup.[96] The input files for the GROMOS RE-EDS simulations were prepared using *amber2gromos* as well as the GROMOS++[67] programs *pdb*2g96, *red_top*, and *prep_eds*. The molecules were aligned for the RE-EDS simulations using the RDKit[97] (details in Figures S2 and S8 in the Supporting Information). Since the coordinates of all molecules are present separately in the system, the molecules are, in principle, able to drift away from each other during a simulation. To ensure that the molecules remain well-aligned during the whole simulation, atomic distance restraints were applied. The pairwise distance restraints were generated with *RestraintMaker*.[85] *RestraintMaker* chooses reference distances r_0 between restrained atoms according to the input alignment. For some molecule pairs, the ring atoms did not perfectly overlap in the initial alignment. The reference distances assigned by *RestraintMaker* were manually set to 0 for those pairs. Four atoms of each molecule were restrained to four atoms of two other molecules, forming a chain of pairwise distance restraints. For set B with 28 molecules, the chain arrangement allowed for relatively large deviations between the

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

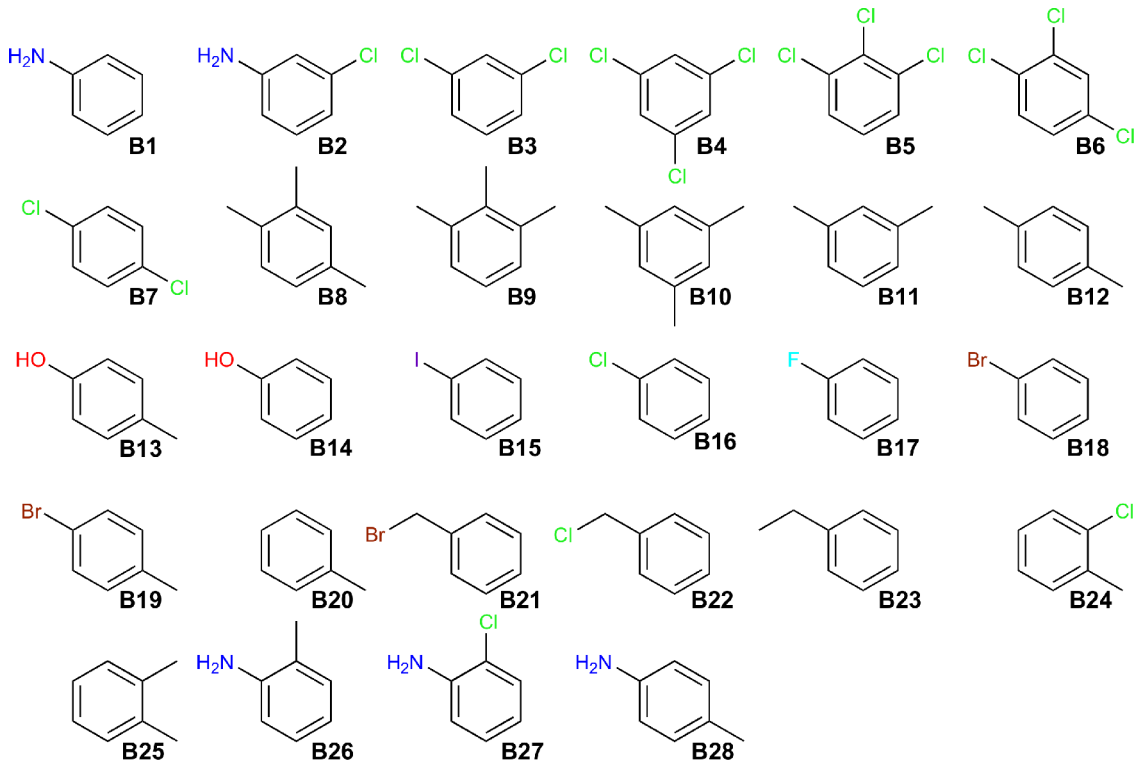


Figure 3. Set B consists of 28 benzene derivatives, selected from the FreeSolv[64,65] database. Set B was further subdivided into subset Ba (B1−B14) and subset Bb (B1 along with B15−B28). The molecule indices, the FreeSolv identifiers, the SMILES strings, and the names of the molecules can be found in Table S2 in the Supporting Information.

molecules furthest apart in the chain. Therefore, additional distance restraints were manually added for four molecule pairs on opposite sides of the chain. The workflow to generate the input files for RE-EDS simulations with GAFF[30] parameters is illustrated in Figure 4.

The RE-EDS simulations were performed with a modified version of GROMOS[24,98] 1.5.0 and the open-source Python3[91] *reeds* module.[92] The TI simulations were performed in GROMACS[53,93] version 2016.6. For the simulations in water, the TIP3P water model[73] was used. A single cutoff radius of 1.2 nm was used for the calculation of the nonbonded interactions. The integration time step was set to 2 fs, and the pairlist was updated every five steps. Long-range nonbonded interactions were calculated in GROMOS using a reaction-field correction[87] with ε_{RF} = 1 for the simulations in vacuum and ε_{RF} = 78.5 for the simulations in water.[99,100] In GROMACS, the long-range nonbonded interactions were calculated using a plain cutoff in vacuum and smooth particle mesh Ewald (SPME)[101−103] in water with a grid spacing of 0.1 nm and an interpolation of order 6. To maintain the temperature at 298.15 K and the pressure at 0.06102 kJ mol^{-1} nm^{-3} (≈1 atm), a Berendsen thermostat and barostat[104] were used in GROMOS for the simulations in water. For the GROMACS TI calculations and the RE-EDS calculations in vacuum, the leapfrog stochastic dynamics integrator was used, so that no temperature scaling was necessary. In the TI calculations, the pressure in water was kept constant at 1.01325 bar (≈1 atm) using a Parrinello−Rahman barostat.[105] All bonds were constrained with the LINCS algorithm[106] (for the TI calculations, 12th order) or the SHAKE algorithm[107] (for the RE-EDS calculations, relative tolerance 10^{-4}), respectively, and harmonic bond-angle bending was employed. In the RE-EDS simulations in GROMOS, the force constant for the distance restraints was set to 5000 kJ/(mol·nm^2).[85]

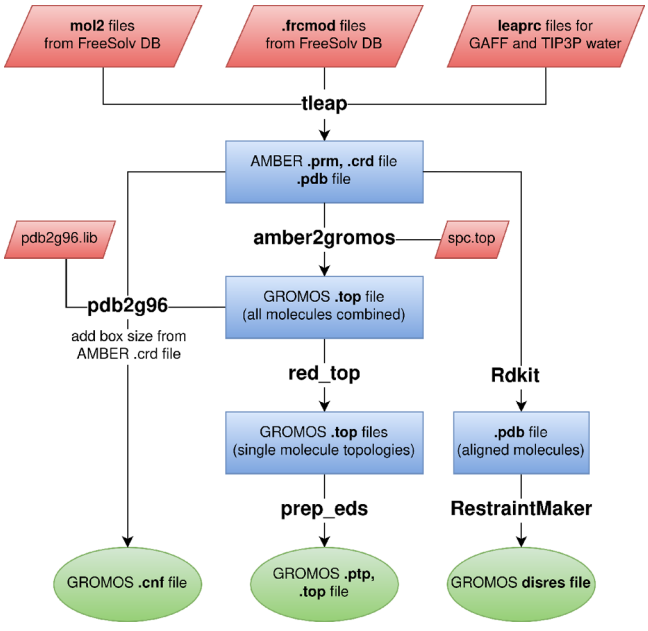


Figure 4. Schematic illustration of the RE-EDS input file preparation. The input files (topology, perturbed topology, coordinates, and distance restraints) for the RE-EDS simulations in GROMOS were created from the frcmod and mol2 files of the FreeSolv[64,65] database. This workflow can easily be extended to also perform the molecule parametrization (i.e., to generate mol2 and frcmod files) using *antechamber* and *parmchk*.[29−31]

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

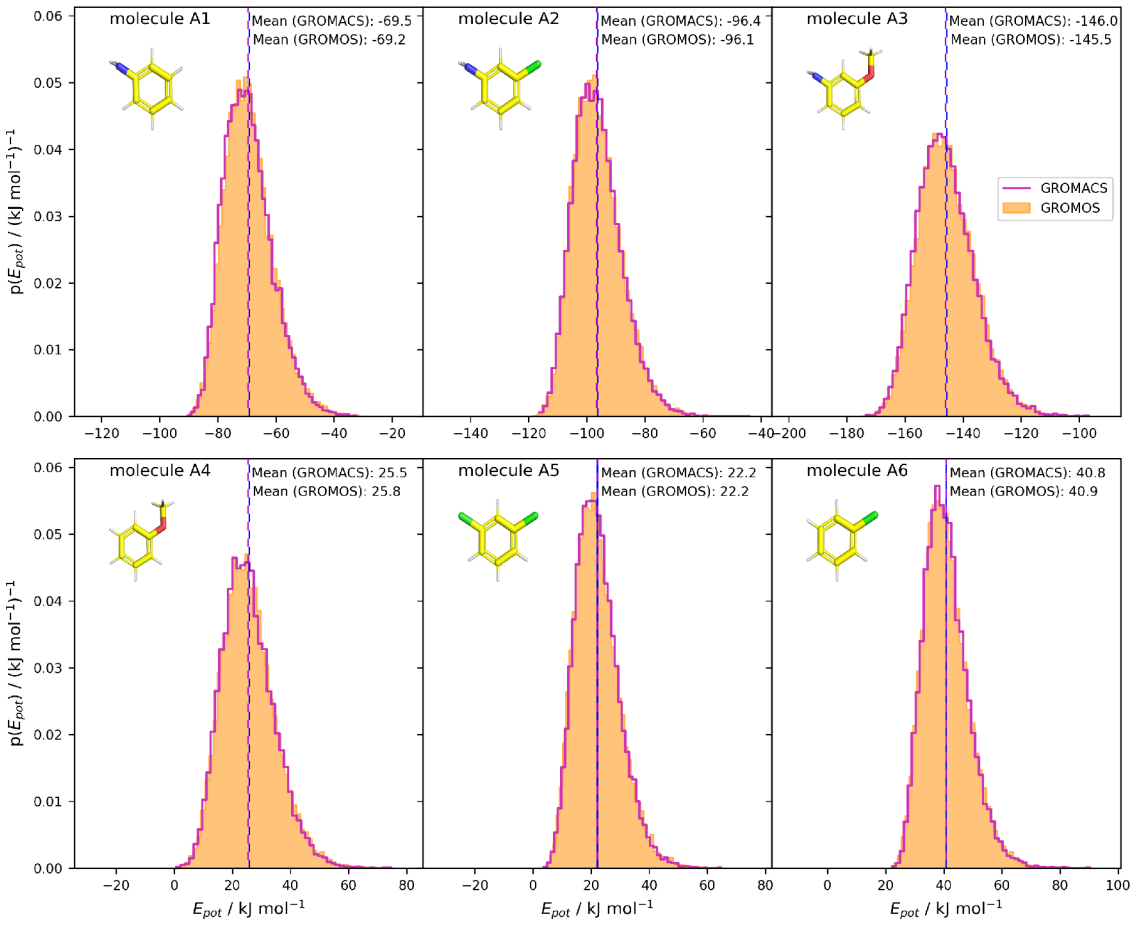


Figure 5. Potential-energy distributions of single-molecule simulations of set A based on 5 ns vacuum simulations in GROMOS[24] (orange bars) and GROMACS[53] (pink lines). The topologies for the simulations are based on the AMBER topologies taken from the FreeSolv[64,65] database. The GROMACS topologies were translated with ParmEd,[57] and the GROMOS topologies were converted with *amber2gromos*. Energies were written every 100 time steps (i.e., every 200 fs), and the first 1.25 ns of the simulations were discarded as equilibration.

The calculated hydration free energies reported in the FreeSolv[64,65] database were obtained from alchemical MBAR[13] simulations in GROMACS with 20 λ-values, each 5 ns long. In the first five intermediate states, the electrostatic interactions were modified, and in the last 15 states, the Lennard-Jones terms were changed.[65] In this work, the relative hydration free energies for set A were obtained from pairwise TI simulations and from RE-EDS simulations in vacuum/water. For set B, they were determined from RE-EDS simulations in vacuum/water. Here, it is important to note that the three methods used different pathways of the thermodynamic cycle to calculate the relative hydration free energies. Considering Figure 1, the MBAR calculations[64,65] correspond to the yellow pathways, the TI calculations correspond to the pink pathways, and the multistate RE-EDS calculations correspond to the blue pathways.

The input files for the RE-EDS simulations can be found at https://github.com/rinikerlab/reeds/tree/main/examples/systems/benzenes_amber2gromos.

Set A. To obtain the pairwise free-energy differences in vacuum/water with TI calculations in GROMACS, 21 λ-values were used in vacuum, and 27 λ-values were used in water. In vacuum, they were spread in steps of 0.05 between 0 and 1. In water, they were spread in steps of 0.05 between 0.1 and 0.9 and more densely in steps of 0.02 around the extreme values (i.e., between 0 and 0.1 and between 0.9 and 1). The values were spread more densely around the extreme values for the simulations in water to smooth discontinuities in the $\langle \partial V/\partial \lambda \rangle$-curves. Such discontinuities were mainly observed between 0.0 and 0.1 but also occurred between 0.9 and 1.0 for molecule pairs A1 - A2, A1 - A3, and A2 - A3 (Figure S3 in the Supporting Information). The masses of the end-state atoms were not perturbed but kept to the masses of the first end-state. After a short steepest-descent energy minimization with maximally 5000 steps, the systems were equilibrated for 0.5 ns. The free-energy differences were calculated from five independent production runs in vacuum/water. The production runs were 5 ns long at each λ-point.

To generate relevant configurations of all end-states for the RE-EDS calculations, six EDS simulations in vacuum/water of 2 ns length were carried out at $s = 1$. The energy offsets were biased toward one of the end-states (molecules) in each of the simulations by setting one energy offset to 500 kJ mol^{-1} and the others to -500 kJ mol^{-1}. The optimized configurations were used for the starting-state mixing (SSM) approach.[21] Simultaneously, 21 EDS simulations of 0.2 ns length with s-values logarithmically distributed between 1 and 10^{-5} were used to determine the lower bound of s for the RE-EDS simulations, which were set to 0.0178 in vacuum and 0.01 in water. RE-EDS simulations of 0.8 ns length were carried out to estimate the energy offsets with 11 replicas in vacuum and 12 replicas in water. Next, the s-distribution was optimized to

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

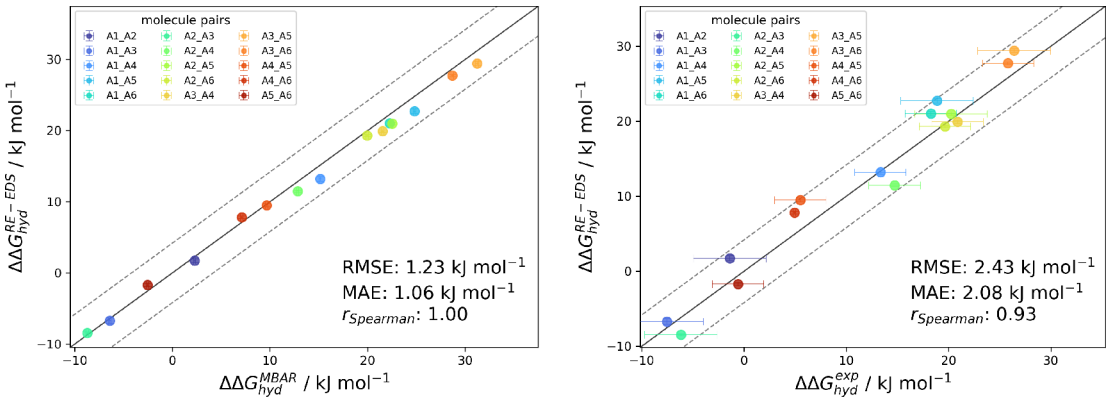


Figure 6. Comparison of the relative hydration free energies of set A: RE-EDS ($\Delta\Delta G_{hyd}^{RE\text{-}EDS}$) versus MBAR ($\Delta\Delta G_{hyd}^{MBAR}$) (left) and RE-EDS versus experiment ($\Delta\Delta G_{hyd}^{exp}$) (right) as reported by the FreeSolv[64,65] database. The gray diagonal lines correspond to perfect alignment within ±4.184 kJ mol^{-1} (±1 kcal mol^{-1}). The RE-EDS results were averaged over five independent production runs in vacuum/water, and the errors of the ΔG values correspond to the standard deviation over the five repeats. The error estimate of the $\Delta\Delta G$ values was calculated via Gaussian error propagation. The numerical values are provided in Table S3 in the Supporting Information. A plot of the deviations from experiment for the different methods is shown in Figure S7 in the Supporting Information. All pairwise comparisons between the different simulation methods and the experimental results are provided in Figure S5.

achieve frequent round trips. Following the SSM approach,[21] the *s*-values were redistributed to include a replica with $s = 1$ and optimized initial coordinates for each end-state (in total 12 replicas in vacuum, 13 in water). In vacuum, four replicas were added after one *s*-optimization step of 0.5 ns length. In water, two *s*-optimization steps of 0.5 and 1.0 ns lengths, respectively, were used, adding in total eight replicas. To achieve good sampling of all end-states, the initial energy offsets in water were rebalanced over two 0.5 ns RE-EDS simulations. Finally, the free-energy differences were calculated from five independent production runs of 0.5 ns in vacuum (11 replicas) and water (16 replicas). For the production runs, the additional replicas with $s = 1$, which were added during the optimization phase, were removed again.

Set B. For set B, 28 EDS simulations of 2 ns length were carried out to generate optimized configurations. The setup for the lower bound search was analogous to set A, and the values were set to 0.01 in vacuum and 0.0056 in water. For the energy offset estimation, 34 replicas were used in vacuum, and 35 replicas were used in water with 0.8 ns length. The *s*-optimization steps were analogous to set A, adding four replicas in vacuum and eight replicas in water. For set B, rebalancing was also required in vacuum. Both in vacuum and in water, four rebalancing steps were used, of 0.5 ns length each. The production run was 1 ns long in vacuum (34 replicas) and 2 ns long in water (39 replicas). As for set A, the obtained free-energy differences were averaged over five independent production runs.

To generate optimized coordinates, 14 EDS simulations were performed for subset Ba, and 15 EDS simulations were performed for subset Bb. The lower bounds for *s* were 0.01 for subset Ba in vacuum, 0.032 for subset Bb in vacuum, and 0.01 for both subsets in water. For the energy offset estimation, 20 (Ba, vacuum), 19 (Bb, vacuum), 20 (Ba, water), and 21 (Bb, water) replicas were used. For both subsets in vacuum, one *s*-optimization step was sufficient. For subset Ba in water, two *s*-optimization steps of 0.5 and 1.0 ns, respectively, were necessary, while for subset Bb in water, only one *s*-optimization step was needed. In vacuum, three energy offset rebalancing steps were carried out for both subsets. In water, four rebalancing steps were used for subset Ba, and three steps were used for subset Bb. Finally, the production runs were 1 ns long in vacuum (20 replicas for Ba, 19 replicas for Bb) and 2 ns long in water (24 replicas for Ba, 21 replicas for Bb). Again, the obtained free-energy differences were averaged over five independent production runs.

Analysis. The analysis of the simulations was carried out using GROMOS++[67] and PyGromosTools.[108] The following Python packages were used for visualization and analysis: Matplotlib,[109] mpmath,[110] NumPy,[111] Pandas,[112] SciPy,[113] and Seaborn.[114] For all sets/subsets, the root-mean-square error (RMSE), the mean absolute error (MAE), and the Spearman[115] correlation coefficient were calculated between the different simulation methods and the experimental values.

RESULTS

Validation of *amber2gromos*. For the molecules of set A, the manual topology comparison showed that the topologies generated by *amber2gromos* were almost identical to the GROMACS topologies generated by ParmEd[57] (apart from differences in units, functional forms, and slight numerical differences). Apart from negligible numerical differences, the potential-energy terms of the 0^{th} integration step in vacuum and water were identical for the simulations in GROMOS and in GROMACS.

After this initial validation, simulations of 5 ns length were performed in vacuum. The distributions of different energy terms as well as the system temperature were compared. They were all qualitatively similar, with almost identical mean values (Figure 5 and Figure S1 in the Supporting Information).

Calculation of Relative Hydration Free Energies for Set A. For set A, the 15 pairwise free-energy differences in vacuum/water obtained from the RE-EDS calculations were first compared to the ones from the TI calculations in GROMACS. With an RMSE of 1.9 kJ mol^{-1} (vacuum) and 2.0 kJ mol^{-1} (water) and an MAE of 1.6 kJ mol^{-1} (vacuum) and 1.7 kJ mol^{-1} (water), the results agreed well. The Spearman correlation coefficient was 1.00 for the vacuum and the water simulations (Figure S4 in the Supporting Information).

Next, the relative hydration free energies from the different sources (i.e., TI in GROMACS, RE-EDS in GROMOS, MBAR in GROMACS,[65] and experimental[64,65]) were compared. The

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

results from all three simulation methods agreed well with the other calculated results, as well as with the experimental values (Figure 6 and Table 1). The RMSEs against the experimental

Table 1. Set A: Overview of Statistical Metrics (RMSE, MAE, and Spearman Correlation Coefficients) with Respect to the Experimental Results and Total Simulation Time for the Different Free-Energy Methods[a]

	$\Delta\Delta G_{hyd}^{MBAR}$[64,65]	$\Delta\Delta G_{hyd}^{TI}$	$\Delta\Delta G_{hyd}^{RE\text{-}EDS}$
RMSE [kJ mol^{-1}]	3.1 ± 0.4	2.4 ± 0.4	2.4 ± 0.3
MAE [kJ mol^{-1}]	2.7 ± 0.3	2.0 ± 0.4	2.1 ± 0.3
$r_{Spearman}$	0.94	0.93	0.94
$t_{preparation}$ [ns]	18	360	101.3
$t_{production}$ [ns]	600	3600	13.5

[a]The RE-EDS and TI results were averaged over five independent production runs in vacuum/water, and the errors of the ΔG values correspond to the standard deviation over the five repeats. The error estimate of the $\Delta\Delta G$ values was calculated *via* Gaussian error propagation. The uncertainties of the RMSE and MAE values were estimated from the distribution of RMSE and MAE when a random selection of up to four molecules was removed from the calculations (5000 repetitions). The accumulated simulation time is split into preparation (pre-processing, equilibration) and production time. Table S3 in the Supporting Information contains the complete table.

results were 2.4 kJ mol^{-1} (TI), 2.4 kJ mol^{-1} (RE-EDS), and 3.1 kJ mol^{-1} (MBAR). The corresponding MAEs were 2.0 kJ mol^{-1} (TI), 2.1 kJ mol^{-1} (RE-EDS), and 2.7 kJ mol^{-1} (MBAR), respectively. The calculated relative hydration free energies also had a high correlation with the experimental results, with Spearman correlation coefficients ranging between 0.93 and 0.94. The full details are provided in Table S3 in the Supporting Information.

While all three free-energy methods achieve comparable and accurate results, the RE-EDS calculations require by far the lowest accumulated simulation time. The total simulation time (preprocessing and production) for the RE-EDS calculations was about 115 ns. For the TI calculations, the total simulation time (equilibration and production) was 3960 ns. This could, of course, be reduced by calculating only the minimal number of required pairwise free-energy differences (i.e., $N-1$, which is five for set A). The production time could likely also be reduced to a total of 2−3 ns without affecting the convergence significantly. Both measures would reduce the total simulation time required for the TI calculations to about 600−840 ns. However, even then, the total simulation time would still be more than five times longer than for the RE-EDS calculations. The calculated values reported in FreeSolv[64,65] required 618 ns total simulation time (equilibration and production). Also here, the simulation time was chosen with convergence in mind and could probably be reduced. However, even with 2−3 ns production runs, the total simulation time would still be about 2−3 times the simulation time required by RE-EDS. Plots of the convergence of the free-energy calculations with RE-EDS as well as the λ-curves for the TI simulations can be found in Figures S6 and S3 in the Supporting Information.

Calculation of Relative Hydration Free Energies for Set B. For set B, we found again excellent agreement between the results obtained from the RE-EDS simulations and the calculated and experimental values reported in the FreeSolv[64,65] database (Figure 7 and Table 2). The RMSE against the experimental results was 2.6 kJ mol^{-1} for RE-EDS and 2.0 kJ mol^{-1} for MBAR.[64,65] The corresponding MAEs were 2.2 kJ mol^{-1} and 1.6 kJ mol^{-1}, respectively. The correlation with experiment was high for both methods, with $r_{Spearman}^{RE\text{-}EDS}$ = 0.89 and $r_{Spearman}^{MBAR}$ = 0.92. The agreement between the two simulation methods was also good, with an RMSE of 1.0 kJ mol^{-1}, an MAE of 0.7 kJ mol^{-1}, and a Spearman correlation coefficient of 0.99. For RE-EDS, molecule pairs B1 - B4, B4 - B8, B4 - B9, B4 - B23, B4 - B25, B6 - B8, and B6 - B9 showed deviations above 5.5 kJ mol^{-1}, the largest for molecule pair B4 - B9 with 6.7 kJ mol^{-1} (Figure S12 in the Supporting Information). For MBAR, molecule pairs B1 - B4, B1 - B6, B4 - B8, B4 - B25, B4 - B27, and B4 - B28 deviated by more than 4.3 kJ mol^{-1} from experiment. Here, the highest absolute deviation was observed for molecule pair B1 - B4 with 4.9 kJ mol^{-1}. The Spearman correlation coefficient of the absolute deviations from experiment for the two simulation methods was relatively high at 0.88, indicating that some of the

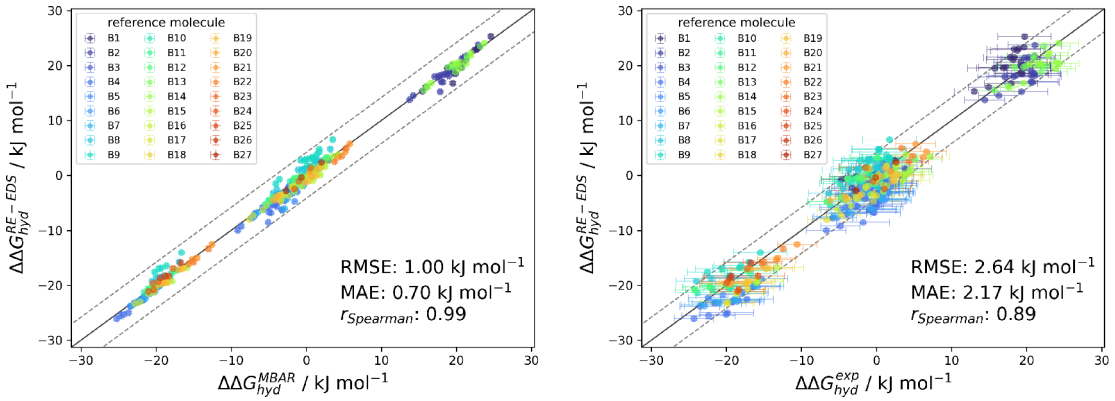


Figure 7. Comparison of the relative hydration free energies of set B: RE-EDS ($\Delta\Delta G_{hyd}^{RE\text{-}EDS}$) versus MBAR ($\Delta\Delta G_{hyd}^{MBAR}$) (left) and RE-EDS versus experiment ($\Delta\Delta G_{hyd}^{exp}$) (right) as reported by the FreeSolv[64,65] database. The gray diagonal lines correspond to perfect alignment within ±4.184 kJ mol^{-1} (±1 kcal mol^{-1}). The $\Delta\Delta G_{hyd}^{ii}$ values are colored according to end-state i (i.e., the "reference molecule" for the calculation). The RE-EDS results were averaged over five independent production runs in vacuum/water, and the errors of the ΔG values correspond to the standard deviation over the five repeats. The error estimate of the $\Delta\Delta G$ values was calculated via Gaussian error propagation. The numerical values are provided in Table S4 in the Supporting Information. A plot of the deviations from experiment for the different methods is shown in Figure S12 in the Supporting Information. All pairwise comparisons between the different simulation methods and the experimental results are provided in Figure S9.

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

Table 2. Set B: Overview of Statistical Metrics (RMSE, MAE, and Spearman Correlation Coefficients) with Respect to the Experimental Results and Total Simulation Time for the Different Free-Energy Methods[a]

	$\Delta\Delta G_{hyd}^{MBAR}$[64,65]	$\Delta\Delta G_{hyd}^{RE\text{-}EDS}$	$\Delta\Delta G_{hyd}^{RE\text{-}EDS,subsets}$
RMSE [kJ mol^{-1}]	2.0 ± 0.2	2.6 ± 0.3	2.4 ± 0.3
MAE [kJ mol^{-1}]	1.6 ± 0.2	2.2 ± 0.2	2.0 ± 0.2
$r_{Spearman}$	0.92	0.89	0.90
$t_{preparation}$ [ns]	84 ns	549 ns	514 ns
$t_{production}$ [ns]	2800 ns	112 ns	129 ns

[a]The RE-EDS and TI results were averaged over five independent production runs in vacuum/water, and the errors of the ΔG values correspond to the standard deviation over the five repeats. The error estimate of the $\Delta\Delta G$ values was calculated *via* Gaussian error propagation. For RE-EDS, both the results for the full set B and for the combined subsets Ba and Bb are reported. The uncertainties of the RMSE and MAE values were estimated from the distribution of RMSE and MAE when a random selection of up to 26 molecules was removed from the calculations (5000 repetitions). The accumulated simulation time is split into preparation (preprocessing, equilibration) and production time. Table S4 in the Supporting Information contains the complete table.

deviations might be related to shortcomings in the force field or in the experimental determination.

To investigate the efficiency and accuracy of RE-EDS free-energy calculations for a larger number of molecules, the RE-EDS results of set B were compared to the ones obtained from RE-EDS simulations of subsets Ba and Bb (Figure 8 and Table 2). With an RMSE against the experimental values[64,65] of 2.4 kJ mol^{-1}, an MAE of 2.0 kJ mol^{-1}, and a Spearman correlation coefficient of 0.90, the combined results from the two separate RE-EDS pipelines (i.e., Ba and Bb) were marginally more accurate. The agreement with the MBAR results[64,65] was slightly higher with an RMSE of 0.8 kJ mol^{-1} and an MAE of 0.5 kJ mol^{-1} (Figure S9 in the Supporting Information). The RMSE between the RE-EDS results for set B versus subsets Ba and Bb was 0.4 kJ mol^{-1} with perfect correlation ($r_{Spearman}$ = 1.00). The slightly higher agreement of the results obtained from the two subsets with the experimental and MBAR results indicates that there is a small "diffusion effect" for this system when more molecules are added to the simulation. As more end-states are added to a system, the number of frames where an end-state contributes maximally to the reference state decreases. Additionally, more *s*-values are required to obtain frequent round trips and more energy offset rebalancing iterations are needed to obtain approximately equal sampling of all the end-states.

Also here, the relatively small simulation time required for the RE-EDS calculations can be highlighted. The total simulation time for the RE-EDS simulations of set B was about 661 ns, compared to 2884 ns for MBAR.[64,65] If pairwise TI simulations in vacuum/water had been used as for set A, the total simulation time would have been between 7128 ns (minimal 27 pairs = $N-1$) and 99627 ns (all pairs). The total simulation time for the RE-EDS pipelines of subsets Ba and Bb combined was about 643 ns. This is slightly shorter than the simulation length for the full set B, mainly due to the fact that for subset Bb in water, only four instead of eight replicas were added during the *s*-optimization, and one less rebalancing step was required for both subsets in vacuum and for Bb in water. Plots of the convergence of the free-energy calculations with RE-EDS can be found in Figures S10 and S11 in the Supporting Information.

CONCLUSION AND OUTLOOK

In this work, the GROMOS++ program *amber2gromos* was introduced to convert topologies from the AMBER to the GROMOS file format. An overview of the differences between AMBER and GROMOS force fields was presented together with a description of the conversion of the AMBER topology parameters to GROMOS topology parameters and the necessary slight modification to the GROMOS source code. A workflow was outlined to prepare topology, coordinate, and distance restraint input files for RE-EDS free-energy calculations with GAFF parameters in GROMOS. The extension of this workflow to the OpenFF family of force fields is straightforward.

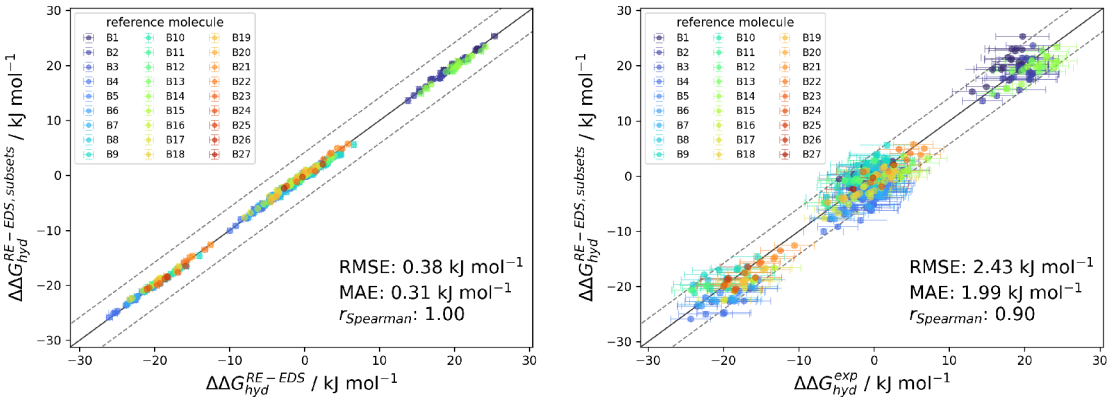


Figure 8. Comparison of the relative hydration free energies with RE-EDS of subsets Ba and Bb ($\Delta\Delta G_{hyd}^{RE\text{-}EDS,subsets}$) compared to the full set B ($\Delta\Delta G_{hyd}^{RE\text{-}EDS}$) (left) and compared to experiment ($\Delta\Delta G_{hyd}^{exp}$) (right) as reported by the FreeSolv[64,65] database. The gray diagonal lines correspond to perfect alignment within ±4.184 kJ mol^{-1} (±1 kcal mol^{-1}). The $\Delta\Delta G_{hyd}^{ji}$ values are colored according to end-state *i* (i.e., the "reference molecule" for the calculation). The RE-EDS results were averaged over five independent production runs in vacuum/water, and the errors of the ΔG values correspond to the standard deviation over the five repeats. The error estimate of the $\Delta\Delta G$ values was calculated via Gaussian error propagation. The numerical values are provided in Table S4 in the Supporting Information. A plot of the deviations from experiment for the different methods is shown in Figure S12 in the Supporting Information. All pairwise comparisons between the different simulation methods and the experimental results are provided in Figure S9.

https://doi.org/10.1021/acs.jcim.2c00383
J. Chem. Inf. Model. 2022, 62, 3043−3056

Two sets of benzene derivatives were selected from the FreeSolv database with six (set A) and 28 molecules (set B). Set A was used to validate the implementation of *amber2gromos* and the related source-code changes to the GROMOS MD engine. The generated GROMOS topologies for the six benzene derivatives were compared to GROMACS topologies generated by ParmEd from the same AMBER topologies. Single-molecule simulations in vacuum performed in GROMOS and in GROMACS showed nearly identical energy and temperature distributions.

Finally, relative hydration free energies were calculated for both sets. For set A, both TI and RE-EDS simulations were carried out in vacuum/water to estimate the 15 pairwise free-energy differences. These results were compared to the relative hydration free energies obtained from experiment as well as the hydration free energies reported in the FreeSolv database (calculated with MBAR). Overall, an excellent agreement was observed between the different free-energy methods and with experiment. While all methods delivered highly accurate results, the RE-EDS calculations required the least amount of total simulation time. The system size was increased to 28 molecules in set B to challenge the RE-EDS pipeline. Again, the results agreed well with the ones from MBAR and with the experimental values.

To test if it is more efficient to use a large set of molecules or two subsets with a shared molecule, set B was divided into two subsets: Ba (molecules B1−B14) and Bb (molecule B1 and molecules B15−B28). While both the results and simulation time of the two RE-EDS approaches were almost identical, smaller subsets may offer some advantages in practice. RE-EDS simulations are, in principle, highly parallelizable, as large parts of both the replicas and the interactions within the replicas can be carried out independently with relatively infrequent communication. Nevertheless, as more molecules/replicas are added to the system, the wall-clock time of the simulations increases (more interactions to calculate, larger communication overhead, more replicas). Using two subsets decreased the elapsed real-time of the RE-EDS pipeline. Further research will be needed to determine optimal splits of data sets into subsets as well as the choice of the common molecule(s). The aim will be to find a balance between avoiding a diffusion effect from too many end-states in one system and error propagation due to too small subsets or suboptimal common molecule(s).

Overall, it has been shown that hydration free-energy calculations with RE-EDS and GAFF parameters executed in GROMOS accurately reproduce both experimental values and results obtained with different free-energy estimators and MD engines. While the molecules of the chosen data sets were relatively small and contained a well-defined common benzene core, previous studies successfully used RE-EDS to calculate binding and hydration free energies for molecule sets involving larger structural changes such as R-group modifications, ring opening/closing, and ring size changes. In future work, GAFF parametrized topologies will be used to perform binding free-energy calculations with RE-EDS in GROMOS.

DATA AND SOFTWARE AVAILABILITY

The input files for the RE-EDS simulations can be found at https://github.com/rinikerlab/reeds/tree/main/examples/systems/benzenes_amber2gromos. The GROMOS software package and the GROMOS++ package of programs can be downloaded for free at http://gromos.net/, AmberTools can be downloaded at https://ambermd.org/AmberTools.php, and GROMACS can be downloaded at http://www.gromacs.org/. The Python code for PyGromosTools and the RE-EDS pipeline are freely available at https://github.com/rinikerlab. *amber2gromos* will be part of the next scheduled release of GROMOS.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00383.

Details of topology conversion with *amber2gromos*, example of *amber2gromos* translation, details of data sets, and additional figures and tables for single-molecule simulations in vacuum and for calculation of relative hydration free energies (PDF)

AUTHOR INFORMATION

Corresponding Authors

Philippe H. Hünenberger − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0002-9420-7998; Email: phil@igc.phys.chem.ethz.ch

Sereina Riniker − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0003-1893-4031; Email: sriniker@ethz.ch

Authors

Salomé R. Rieder − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0001-8036-651X

Benjamin Ries − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0002-0945-8304

Kay Schaller − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland*

Candide Champion − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0001-6982-1569

Emilia P. Barros − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0001-9755-260X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00383

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support by the Swiss National Science Foundation (Grant no. 200021-178762 and no. 200021-175944).

REFERENCES

(1) Michel, J.; Foloppe, N.; Essex, J. W. Rigorous Free Energy Calculations in Structure-Based Drug Design. *Mol. Inform.* **2010**, *29*, 570−578.

(2) Borhani, D. W.; Shaw, D. E. The Future of Molecular Dynamics Simulations in Drug Discovery. *J. Comput. Aided Mol. Des.* **2012**, *26*, 15−26.

(3) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035−4061.

(4) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911−2937.
(5) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of "Alchemical Perturbation" in Medicinal Chemistry. *J. Med. Chem.* **2018**, *61*, 638−649.
(6) Armacost, K. A.; Riniker, S.; Cournia, Z. Novel Directions in Free Energy Methods and Applications. *J. Chem. Inf. Model.* **2020**, *60*, 1−5.
(7) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153−4169.
(8) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D., Jr; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(9) Song, L. F.; Merz, K. M., Jr Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5308−5318.
(10) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(11) Zwanzig, R. W. High-Temperature Equation of State by a Pertubation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(12) Bennett, C. H. Efficient Estimation of Free Energy Differences From Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(13) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
(14) Knight, J. L.; Brooks, C. L., III Multisite λ Dynamics for Simulated Structure-Activity Relationship Studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.
(15) Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L., III Automated, Accurate, and Scalable Relative Protein-Ligand Binding Free-Energy Calculations Using λ Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 7895−7914.
(16) Hayes, R. L.; Buckner, J.; Brooks, C. L., III BLaDE: A Basic Lambda Dynamics Engine for GPU-Accelerated Molecular Dynamics Free Energy Calculations. *J. Chem. Theory Comput.* **2021**, *17*, 6799−6807.
(17) Christ, C. D.; van Gunsteren, W. F. Enveloping Distribution Sampling: A Method to Calculate Free Energy Differences From a Single Simulation. *J. Chem. Phys.* **2007**, *126*, 184110.
(18) Christ, C. D.; van Gunsteren, W. F. Multiple Free Energies From a Single Simulation: Extending Enveloping Distribution Sampling to Nonoverlapping Phase-space Distributions. *J. Chem. Phys.* **2008**, *128*, 174112.
(19) Sidler, D.; Schwaninger, A.; Riniker, S. Replica Exchange Enveloping Distribution Sampling (RE-EDS): A Robust Method to Estimate Multiple Free-Energy Differences From a Single Simulation. *J. Chem. Phys.* **2016**, *145*, 154114.
(20) Sidler, D.; Cristòfol-Clough, M.; Riniker, S. Efficient Round-Trip Time Optimization for Replica-Exchange Enveloping Distribution Sampling (RE-EDS). *J. Chem. Theory Comput.* **2017**, *13*, 3020−3030.
(21) Ries, B.; Normak, K.; Weiss, R. G.; Rieder, S.; de Barros, E.; Champion, C.; König, G.; Riniker, S. Relative Free-Energy Calculations for Scaffold Hopping-Type Transformations with an Automated RE-EDS Sampling Procedure. *J. Comput. Aided Mol. Des.* **2022**, *36*, 117−130.
(22) Perthold, J. W.; Oostenbrink, C. Accelerated Enveloping Distribution Sampling: Enabling Sampling of Multiple End States While Preserving Local Energy Minima. *J. Phys. Chem. B* **2018**, *122*, 5030−5037.
(23) Perthold, J. W.; Petrov, D.; Oostenbrink, C. Toward Automated Free Energy Calculation with Accelerated Enveloping Distribution Sampling (A-EDS). *J. Chem. Inf. Model.* **2020**, *60*, 5395−5406.
(24) Schmid, N.; Christ, C. D.; Christen, M.; Eichenberger, A. P.; van Gunsteren, W. F. Architecture, Implementation and Parallelization of the GROMOS Software for Biomolecular Simulation. *Comput. Phys. Commun.* **2012**, *183*, 890−903.
(25) Hünenberger, P. H.; van Gunsteren, W. F. *Potential Energy Surfaces*; Springer: 1999; pp 177−214, DOI: 10.1007/978-3-642-46879-7_4.
(26) Nerenberg, P. S.; Head-Gordon, T. New Developments in Force Fields for Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129−138.
(27) Monticelli, L.; Tieleman, D. P. Force Fields for Classical Molecular Dynamics. *Methods Mol. Biol.* **2013**, *924*, 197−213.
(28) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565−578.
(29) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: An Accessory Software Package for Molecular Mechanical Calculations. *Abstr. Pap. Am. Chem. Soc.*; 2001; Vol. *222*, p U403.
(30) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(31) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(32) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026−4037.
(33) Stroet, M.; Caron, B.; Visscher, K. M.; Geerke, D. P.; Malde, A. K.; Mark, A. E. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *J. Chem. Theory Comput.* **2018**, *14*, 5834−5845.
(34) Oliveira, M. P.; Andrey, M.; Rieder, S. R.; Kern, L.; Hahn, D. F.; Riniker, S.; Horta, B. A. C.; Hunenberger, P. H. Systematic Optimization of a Fragment-Based Force Field Against Experimental Pure-Liquid Properties Considering Large Compound Families: Application to Saturated Haloalkanes. *J. Chem. Theory Comput.* **2020**, *16*, 7525−7555.
(35) Oliveira, M. P.; Hünenberger, P. H. Systematic Optimization of a Fragment-Based Force Field Against Experimental Pure-Liquid Properties Considering Large Compound Families: Application to Oxygen and Nitrogen Compounds. *Phys. Chem. Chem. Phys.* **2021**, *23*, 17774−17793.
(36) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on ab initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543−3556.
(37) Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D., Jr; Roux, B. Optimized Lennard-Jones Parameters for Druglike Small Molecules. *J. Chem. Theory Comput.* **2018**, *14*, 3121−3131.
(38) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Res.* **2017**, *45*, W331−W336.
(39) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076−6092.
(40) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0 - the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.
(41) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom

Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(42) Vanommeslaeghe, K.; MacKerell, A. D., Jr Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.
(43) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.
(44) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: A Tool for High-Throughput Crystallography of Protein-Ligand Complexes. *Acta Crystallogr. D* **2004**, *60*, 1355−1363.
(45) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. RED Server: A Web service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res.* **2011**, *39*, W511−W517.
(46) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: AFast Force Field Generation Tool for Small Organic Molecules. *J. Comput. Chem.* **2011**, *32*, 2359−2368.
(47) Weiner, P. K.; Kollman, P. A. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *J. Comput. Chem.* **1981**, *2*, 287−303.
(48) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198−210.
(49) Case, D. A.; Walker, R. C.; Cheatham, C.; Simmerling, T. E.; Roitberg, A.; Merz, K. M.; Luo, R.; Darden, T.; Wang, J.; Duke, R. E.; Roe, D. R.; LeGrand, S.; Swails, J.; Cerutti, D.; Monard, G.; Sagui, C.; Kaus, J.; Betz, R.; Madej, B.; Lin, C.; Mermelstein, D.; Li, P.; Onufriev, A.; Izadi, S.; Wolf, R. M.; Wu, X.; Götz, A. W.; Gohlke, H.; Homeyer, N.; Botello-Smith, W. M.; Xiao, L.; Luchko, T.; Giese, T.; Lee, T.; Nguyen, H. T.; Nguyen, H.; Janowski, P.; Omelyan, I.; Kovelnko, A.; Kollman, P. A. *AMBER 2016 Reference Manual*; University of California: San Francisco, 2016.
(50) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(51) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(52) Bekker, H.; Berendsen, H. J. C.; Dijkstra, E. J.; Achterop, S.; van Drunen, R.; van der Spoel, D.; Sijbers, A.; Keegstra, H. GROMACS: A Parallel Computer for Molecular Dynamics Simulations. *Physics Computing 92* **1993**, 252−256.
(53) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(54) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, Switzerland, 1996; pp 1−1042.
(55) Eastman, P.; Pande, V. OpenMM: AHardware-Independent Framework for Molecular Simulations. *Comput. Sci. Eng.* **2010**, *12*, 34−39.
(56) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
(57) Swails, J.; Hernandez, C.; Mobley, D. L.; Nguyen, H.; Wang, L.-P.; Janowski, P. *ParmEd*; 2010. https://github.com/ParmEd/ParmEd (accessed 2022-05-05).
(58) da Silva, A. W. S.; Vranken, W. F. ACPYPE-Antechamber Python Parser Interface. *BMC Res. Notes* **2012**, *5*, 367.
(59) Lee, T. *gromos2amber*. https://github.com/ATB-UQ/gromos2amber (accessed 2022-05-05).
(60) Vermaas, J. V.; Hardy, D. J.; Stone, J. E.; Tajkhorshid, E.; Kohlmeyer, A. TopoGromacs: Automated Topology Conversion from CHARMM to GROMACS within VMD. *J. Chem. Inf. Model.* **2016**, *56*, 1112−1116.
(61) Hedges, L. O.; Mey, A. S.; Laughton, C. A.; Gervasio, F. L.; Mulholland, A. J.; Woods, C. J.; Michel, J. BioSimSpace: An Interoperable Python Framework for Biomolecular Simulation. *J. Open Source Softw.* **2019**, *4*, 1831.
(62) Guthrie, J. P. A Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.
(63) Kashef Ol Gheta, S.; Oliveira, M. P.; Rieder, S. R.; Horta, B. A. C.; Acree, W. E.; Hünenberger, P. H. Evaluating Classical Force Fields against Experimental Cross-Solvation Free Energies. *J. Chem. Theory Comput.* **2020**, *16*, 7556−7580.
(64) Mobley, D. L.; Guthrie, J. P. FreeSolv: ADatabase of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput. Aided Mol. Des.* **2014**, *28*, 711−720.
(65) 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. *J. Chem. Eng. Data* **2017**, *62*, 1559−1569.
(66) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database*, version 2012.
(67) Eichenberger, A. P.; Allison, J. R.; Dolenc, J.; Geerke, D. P.; Horta, B. A. C.; Meier, K.; Oostenbrink, C.; Schmid, N.; Steiner, D.; Wang, D.; Wang, D.; van Gunsteren, W. F. The GROMOS++ Software for the Analysis of Biomolecular Simulation Trajectories. *J. Chem. Theory Comput.* **2011**, *7*, 3379−3390.
(68) Swails, J. M. Free Energy Simulations of Complex Biological Systems at Constant pH. Ph.D. thesis, University of Florida, 2013.
(69) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: the GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(70) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions: 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(71) Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J. Chem. Theory Comput.* **2016**, *12*, 3825−3850.
(72) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331−342, DOI: 10.1007/978-94-015-7658-1_21.
(73) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(74) *GROMOS Software for (Bio)Molecular Simulation*; 2021; Volume 4. http://gromos.net (accessed 2022-05-05).
(75) *GROMOS Software for (Bio)Molecular Simulation*; 2021; Volume 2. http://gromos.net (accessed 2022-05-05).
(76) *GROMOS Software for (Bio)Molecular Simulation*; 2021; Volume 6. http://gromos.net (accessed 2022-05-05).
(77) Klimovich, P. V.; Mobley, D. L. Predicting Hydration Free Energies Using All-Atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput. Aided Mol. Des.* **2010**, *24*, 307−316.

(78) 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. *J. Chem. Phys.* **1988**, *89*, 3742−3746.
(79) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime Time? *Annu. Rep. Comput. Chem.* **2007**, *3*, 41−59.
(80) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569−1582.
(81) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632−2647.
(82) Hansmann, U. H. E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(83) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(84) Riniker, S.; Christ, C. D.; Hansen, N.; Mark, A. E.; Nair, P. C.; van Gunsteren, W. F. Comparison of Enveloping Distribution Sampling and Thermodynamic Integration to Calculate Binding Free Energies of Phenylethanolamine N-Methyltransferase Inhibitors. *J. Chem. Phys.* **2011**, *135*, 024105.
(85) Ries, B.; Rieder, S. R.; Rhiner, C.; H, H. P.; Riniker, S. RestraintMaker: A Graph-Based Approach to Determine Distance Restraints for Free-Energy Calculations with Dual Topology. *J. Comput. Aided Mol. Des.* **2022**, *36*, 175.
(86) Stroustrup, B. *The C++ Programming Language*, 4th ed.; Addison-Wesley: 2013.
(87) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102*, 5451−5459.
(88) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Kräutler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS Software for Biomolecular Simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719−1751.
(89) Bergdorf, M.; Peter, C.; Hünenberger, P. H. Influence of Cut-Off Truncation and Artificial Periodicity of Electrostatic Interactions in Molecular Simulations of Solvated Ions: A Continuum Electrostatics Study. *J. Chem. Phys.* **2003**, *119*, 9129−9144.
(90) Kubincová, A.; Riniker, S.; Hünenberger, P. H. Reaction-Field Electrostatics in Molecular Dynamics Simulations: Development of a Conservative Scheme Compatible with an Atomic Cutoff. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26419−26437.
(91) van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; 2009.
(92) Ries, B.; Rieder, S. R.; Champion, C.; Barros, E. P.; Riniker, S. *rinikerlab/reeds: An Automatized RE-EDS Sampling Procedure (v1.0)*. 2021. https://github.com/rinikerlab/reeds (accessed 2022-05-05).
(93) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(94) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(95) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(96) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485−2490.
(97) Landrum, G.; Tosco, P.; Kelley, B.; Riniker, S.; Ric; gedeck; Vianello, R.; Schneider, N.; Dalke, A.; N, D.; Eisuke, K.; Cole, B.; Turk, S.; Swain, M.; Alexander, S.; Cosgrove, D.; Vaucher, A.; W'ojcikowski, M.; Jones, G.; Probst, D.; Godin, G.; Scalfani, V. F.; Pahl, A.; Francois, B.; JLVarjo; strets123; JP; DoliathGavin; Sforna, G.; Jensen, J. H. *rdkit/rdkit: 2021_03_2 (Q1 2021) Release*; 2021. https://doi.org/10.5281/zenodo.4750957 (accessed 2022-05-05).
(98) van Gunsteren, W. F.; Daura, X. *gromos.net*; 2021. http://www.gromos.net/ (accessed 2022-01-06).
(99) Glättli, A.; Daura, X.; van Gunsteren, W. F. Derivation of an Improved Simple Point Charge Model for Liquid Water: SPC/A and SPC/L. *J. Chem. Phys.* **2002**, *116*, 9811−9828.
(100) Riniker, S.; Kunz, A.-P. E.; van Gunsteren, W. F. On the Calculation of the Dielectric Permittivity and Relaxation of Molecular Models in the Liquid Phase. *J. Chem. Theory Comput.* **2011**, *7*, 1469−1475.
(101) Ewald, P. P. Ewald Summation. *Ann. Phys.* **1921**, *369*, 1−2.
(102) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(103) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(104) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(105) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196.
(106) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(107) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(108) Lehner, M. T.; Ries, B.; Rieder, S. R.; Riniker, S. *rinikerlab/PyGromosTools: PyGromosTools_V2 (v2.0)*; 2021. https://zenodo.org/record/4849409#.Yp0gTKjMLIU (accessed 2022-05-05).
(109) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(110) Johansson, F. *mpmath: A Python Library for Arbitrary-Precision Floating-Point Arithmetic (Version 0.18)*; 2013. http://mpmath.org/ (accessed 2022-06-05).
(111) Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22−30.
(112) McKinney, W. Data Structures for Statistical Computing in Python. *Proc. of the 9th Python in Science Conf.* **2010**, *445*, 51−56.
(113) Virtanen, P.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261−272.
(114) Waskom, M. L. seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021.
(115) Spearman, C. The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **1904**, *15*, 72−101.