

# Replica-Exchange Enveloping Distribution Sampling: Calculation of Relative Free Energies in GROMOS

Salomé R. Rieder<sup>§</sup>, Benjamin Ries, Candide Champion, Emilia P. Barros, Philippe H. Hünenberger, and Sereina Riniker\*

<sup>§</sup>SCS-DSM Award for best virtual poster in Computational Chemistry

**Abstract:** Free-energy calculations based on molecular dynamics (MD) simulations are playing an increasingly important role for computer-aided drug design and material discovery in recent years. Free-energy differences between pairs of end-states can be estimated using well-established methods such as thermodynamic integration (TI) or Bennett's acceptance ratio (BAR). An attractive alternative is the recently developed replica-exchange enveloping distribution sampling (RE-EDS) method, which enables estimating relative free-energy differences between multiple molecules from a single simulation. Here, we provide an introduction to the principles underlying RE-EDS and give an overview of the RE-EDS pipeline. In addition, we provide a description of the two complementary tools *RestraintMaker* and *amber2gromos*. We briefly discuss the findings of three recent applications of RE-EDS to calculate relative binding or hydration free energies. In all three studies, good agreement was found between the results obtained using RE-EDS and experimental values as well as values obtained using other free-energy methods.

**Keywords:** Computational chemistry · Free energy · RE-EDS



From top left to bottom right: S. R. Rieder, B. Ries, C. Champion, E. P. Barros, P. H. Hünenberger, S. Riniker

**Salomé R. Rieder** obtained her Bachelor's and Master's degree in Computational Science and Engineering with a specialization in fluid dynamics and computational chemistry from ETH Zurich. For her Master's thesis she was awarded the ETH Medal

\*Correspondence: Prof. S. Riniker, E-mail: [sriniker@ethz.ch](mailto:sriniker@ethz.ch)  
Laboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 2,  
CH-8092 Zurich

in 2019. She has been working as a PhD candidate in the CSMS and CCG research groups since 2018. Her research is focused on the CombiFF force-field parameterization project and on the development and application of the RE-EDS multistate free-energy method.

## 1. Introduction

Molecular dynamics (MD) simulation is a well-established tool to investigate the properties of molecular systems *in silico*. In classical MD simulations, Newton's equations of motion are integrated numerically to propagate the particles of a system in time. The accuracy of MD calculations relies critically on the quality of the underlying force field, *i.e.* the functional form and parameters that determine the covalent and non-bonded interactions between the particles.

Free-energy calculations are an important and challenging task in the field of MD simulation, specifically for computer-aided drug design and material discovery. The free energy of a state determines the probability that the system will adopt this specific state.<sup>[1]</sup> For example, the (relative) binding free energy calculated in simulations is directly related to the experimental binding affinities of ligands to a protein.<sup>[2]</sup> Persistent methodological, software and hardware advances continue to make the calculation of binding free energies more efficient and accurate. Furthermore, the automation of free-energy pipelines renders the setup of such calculations less error prone. However, they can still be quite time-consuming. Owing to much smaller system sizes, the calculation of solvation/hydration free energies of small organic molecules is much faster, making them a good test case to compare different free-energy methods and force fields.

Here, we provide an overview of the replica-exchange enveloping distribution sampling (RE-EDS)<sup>[3,4]</sup> pipeline,<sup>[5]</sup> as well as of the two complementary tools *RestraintMaker*<sup>[6]</sup> and *amber2gromos*.<sup>[7]</sup> We also briefly discuss the results of three recent studies

investigating the performance of RE-EDS, *RestraintMaker* and *amber2gromos* to calculate relative binding or hydration free energies.

## 2. RE-EDS in a Nutshell

### 2.1 Calculation of Free-energy Differences

In the thermodynamic cycle, two equivalent pathways can be taken to obtain the relative hydration free energy  $\Delta\Delta G_{hyd}^{BA}$  of two end-states *A* and *B*, either *via* absolute hydration free energies or *via* pairwise free-energy differences in water and in vacuum (Eqn. (1))

$$\Delta\Delta G_{hyd}^{BA} = \Delta G_{hyd}^B - \Delta G_{hyd}^A = \Delta G_{wat}^{BA} - \Delta G_{vac}^{BA}. \quad (1)$$

Thermodynamic integration (TI)<sup>[8]</sup> and (multistate) Bennett's acceptance ratio (MBAR)<sup>[9,10]</sup> are two well-established pathway-dependent methods to estimate free-energy differences from MD simulations using a coupling parameter  $\lambda$ . They can be used, for example, to estimate the free-energy difference of a molecule upon a change in the environment, or the relative free-energy difference of a pair of molecules. Enveloping distribution sampling (EDS),<sup>[11,12]</sup> on the other hand, is a pathway-independent multi-state free-energy method, *i.e.* the free-energy difference of any end-state pair in the system can be calculated from a single simulation. In EDS, a reference state  $V_R$  combining  $N$  end-states is defined in Eqn. (2) (Fig. 1)

$$V_R(\mathbf{r}; s, \mathbf{E}^R) = \frac{-1}{\beta s} \ln \left[ \sum_{i=1}^N e^{-\beta s (v_i(\mathbf{r}) - E_i^R)} \right], \quad (2)$$

where  $s > 0$  is the smoothness parameter,  $\mathbf{E}^R$  is a set of energy offsets, and  $\beta = 1/(k_B T)$  with  $k_B$  being the Boltzmann constant and  $T$  the absolute temperature. The energy offsets govern the contribution of the end-states to the reference potential energy. At high  $s$ -values, the reference state is dominated by the end-state with the lowest value of  $V_i(\mathbf{r}) - E_i^R$ . As the smoothness parameter decreases towards zero, it flattens the potential-energy landscape such that each end-state contributes to the reference state simultaneously (Fig. 1). The free-energy difference between any pair of end-states in the system can then be calculated as,

$$\Delta G_{BA} = \frac{-1}{\beta} \ln \frac{\left\langle e^{-\beta(V_B - V_R)} \right\rangle_R}{\left\langle e^{-\beta(V_A - V_R)} \right\rangle_R}. \quad (3)$$

To obtain accurate free-energy estimates with EDS, an optimal choice of the  $s$ -value and of the energy offsets is essential in practice. To mitigate the necessity for an optimal  $s$ -value, replica-exchange EDS (RE-EDS) was introduced,<sup>[3,4]</sup> combining EDS with Hamiltonian replica exchange.<sup>[13,14]</sup> In RE-EDS, the sampling of the end-states is enhanced by simulating multiple replicas with different  $s$ -values and attempting replica exchanges at fixed intervals (Fig. 2). Three 'sampling regions' can be distinguished: physical sampling ( $s \approx 1$ ) with typically one dominant end-state, an intermediate region where more end-states start to contribute to the sampling, and a so-called 'undersampling' region with all end-states contributing roughly equally.<sup>[5,15]</sup>

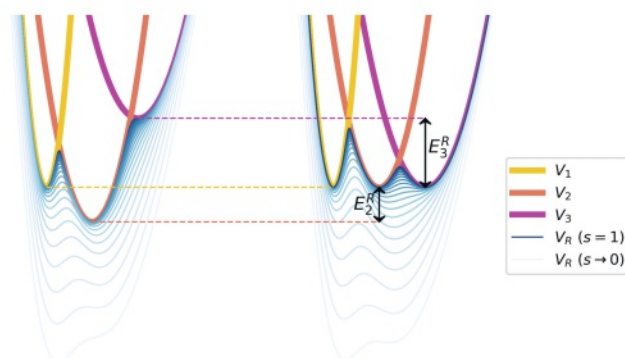


Fig. 1. Schematic illustration of the EDS reference potential-energy function  $V_R$  with the effect of the energy offsets  $E_i^R$  and of the  $s$ -values ( $0 < s \leq 1$ ) highlighted. The energy offsets modify the contribution of the different end-states to the reference state, and the  $s$ -values reduce the energy barriers.

### 2.2 RE-EDS Pipeline

Recently, Ries *et al.*<sup>[5]</sup> published an improved RE-EDS pipeline consisting of three main phases: parameter exploration, parameter optimization, and production. During the parameter exploration, a favorable configuration is generated for each end-state to use as initial coordinates for a starting state mixing (SSM) approach. Further, a lower bound for the  $s$ -values is determined and initial energy offsets are estimated. The second phase involves optimizing the  $s$ -distribution (number and location of replicas) and rebalancing of the energy offsets such that all end-states are sampled roughly equally at  $s = 1$ . Finally, the relative free-energy differences of all end-state pairs in the system are calculated from the production run. The workflow is controlled by the Python3 *reeds* module, freely available at <https://github.com/rinikerlab/reeds/>.

### 2.3 RestraintMaker

In free-energy calculations, the end-states can be represented as 'single', 'hybrid', or 'dual topology'.<sup>[6]</sup> In (RE-)EDS, a dual topology representation is typically used, *i.e.* the coordinates of all ligands are present separately in the system, independent from each other. Consequently, the molecules might drift away from each other during a simulation. One strategy to prevent this drift is the application of (atomic) distance restraints to the common core of the molecules. As the number of end-states in the system and/or the size of the ligands grows, the choice

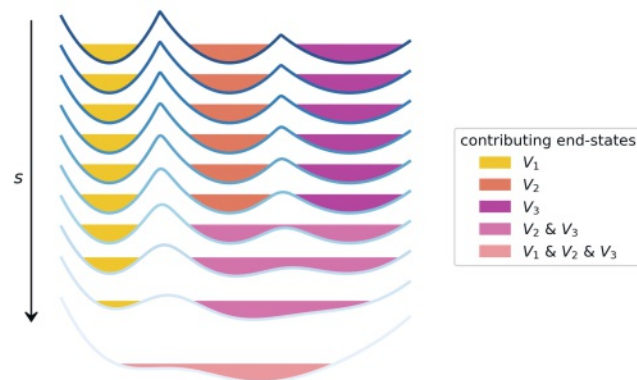


Fig. 2. Schematic illustration of the replicas of a RE-EDS simulation. In RE-EDS, multiple replicas of EDS simulations with successively lower  $s$ -values are run in parallel. After a fixed number of steps, a Hamiltonian replica exchange is attempted between neighboring replicas. This ensures that there is approximately equal sampling of all end-states at  $s = 1$ .

of appropriate distance restraints becomes increasingly non-trivial. To address this, the Python3 program *RestraintMaker* was developed by Ries *et al.*,<sup>[6]</sup> which is freely available at <https://github.com/rinikerlab/restraintmaker/>. Given a set of pre-aligned molecules, *RestraintMaker* employs a greedy algorithm to select locally optimal distance restraints for dual topology free-energy calculations, *e.g.* with RE-EDS or TI.

## 2.4 amber2gromos

A popular small molecule force field is the generalized AMBER force field (GAFF).<sup>[16]</sup> One of the advantages of GAFF is the availability of antechamber,<sup>[17]</sup> a tool to automatically parameterize molecular systems. GAFF was originally developed for the AMBER MD engine.<sup>[18]</sup> There are several differences between AMBER and GROMOS force fields, such as the file formats, the units, and the functional form.<sup>[19]</sup> As RE-EDS is currently implemented in the GROMOS MD engine,<sup>[20]</sup> the C++ program *amber2gromos*<sup>[7]</sup> was developed to convert GAFF (or generally AMBER) topologies into a format compatible with the GROMOS MD engine, facilitating the preparation of new systems of interest for (RE-EDS) simulations in GROMOS. *amber2gromos* will be part of the next scheduled release of the GROMOS++ package,<sup>[21]</sup> freely available at <http://gromos.net/>.

## 3. RE-EDS Case Studies

The performance of RE-EDS has been assessed in several studies.<sup>[3–7]</sup> In this section, we would like to highlight three recent publications.

### 3.1 Validation of the RE-EDS Pipeline: Relative Binding Free Energies of CHK1 Inhibitors

In a study by Ries *et al.*,<sup>[5]</sup> the improved RE-EDS pipeline was tested by estimating the relative binding free energies of five inhibitors of human checkpoint kinase 1 (CHK1).<sup>[22]</sup> The five selected ligands contained R-group modifications as well as different core-hopping transformations such as ring size change and ring opening/closing. Two RE-EDS pipelines were compared, one with a single set of starting coordinates (labeled 1SS) and one with SSM. The relative binding free energies obtained were then compared to experimental values<sup>[22]</sup> and to reported results obtained

with FEP+<sup>[23]</sup> and QligFEP<sup>[24]</sup> (Fig. 3, left). With an RMSE of 3.3 kJ/mol against the experimental values, the accuracy of the results obtained with RE-EDS SSM was considerably better than with RE-EDS 1SS (RMSE = 4.8 kJ/mol). The accuracy of RE-EDS SSM was found to be comparable to that of FEP+ and QligFEP. With a total simulation time of about 50 ns, QligFEP was the most efficient of the three approaches, followed by the RE-EDS pipeline with about 160 ns, requiring only about a quarter of the simulation time of FEP+. One of the main advantages of RE-EDS is the fact that all transformations are sampled explicitly, avoiding the error propagation that occurs when only selected pairs are calculated with pathway dependent methods.

### 3.2 Validation of RestraintMaker: Relative Hydration Free Energies of Small Molecules

In another study,<sup>[6]</sup> relative hydration free energies were calculated for two sets of small molecules in order to validate the quality of the distance restraints selected by *RestraintMaker*. The first set (labeled A) consisted of six molecules and the second set (labeled B) contained ten molecules. The selected molecules contained a small ring core and involved R-group changes and core-hopping transformations, *e.g.* benzene to cyclohexane. Dual topology TI and RE-EDS simulations were carried out in vacuum/water in GROMOS to estimate the relative hydration free energies. *RestraintMaker* was used to select atomic distance restraints. For the multistate RE-EDS calculations, each molecule was restrained to two other molecules, forming a cyclic chain of pairwise restrained molecules. The obtained relative hydration free energies were compared to experimental values as well as to the results obtained from TI calculations as reported on the ATB server<sup>[24]</sup> (Fig. 3, center). For a reasonable force constant (5000 kJ mol<sup>-1</sup> nm<sup>-2</sup>), the molecules remained well-aligned during the simulations and the sampling behavior was not (or only negligibly) affected by the employed distance restraints. For both selected sets, the results obtained with the different free energy estimators agreed well with each other and with the experimental values. The simulation time required to obtain converged free-energy estimates with RE-EDS was shorter than for the dual topology TI calculations by about a factor 5 for set A and a factor 3.5 for set B.

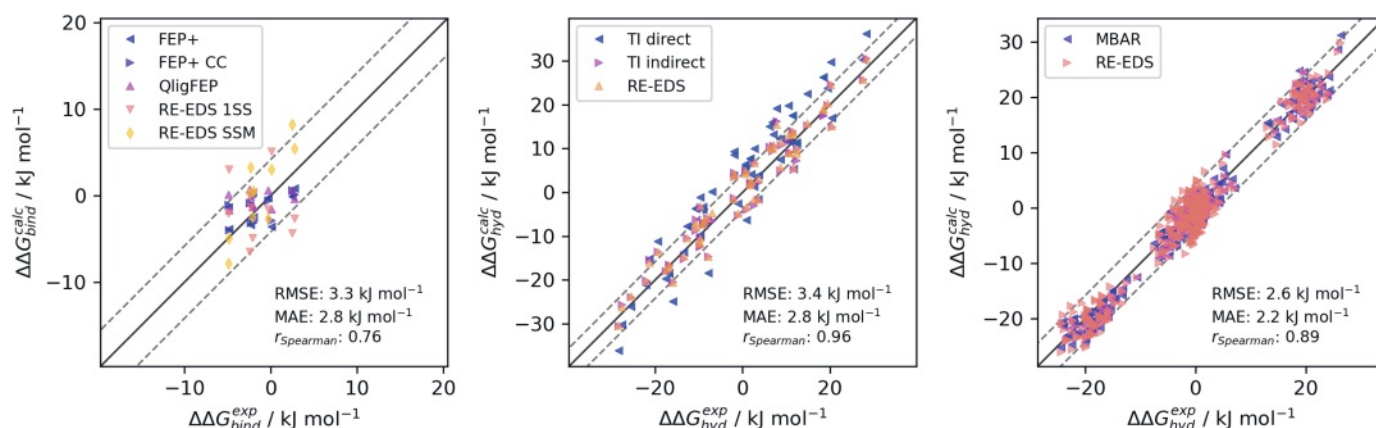


Fig. 3. Comparison of relative binding free energies (left) or hydration free energies (middle, right) obtained with different free-energy methods to experimental values. The underlying datasets and simulation setups are discussed in detail in refs. [5–7]. The gray line corresponds to perfect agreement with  $\pm 4.184$  kJ/mol ( $\pm 1$  kcal/mol). The reported metrics (RMSE, MAE,  $r_{\text{Spearman}}$ ) compare the RE-EDS results to the experimental values. Overall, there is a good agreement between RE-EDS and both experiment and other free-energy estimators. (Left): relative binding free energies of five CHK1 inhibitors obtained with FEP+ with/without cycle closure (CC),<sup>[23]</sup> QligFEP,<sup>[24]</sup> and RE-EDS (SSM and 1SS)<sup>[5]</sup> compared to experimental results.<sup>[22]</sup> (Middle): relative hydration free energies of two sets of small molecules with a ring core obtained with TI<sup>[25]</sup> (direct *via* absolute free-energy calculations), TI<sup>[6]</sup> (indirect *via* relative free-energy calculations), and RE-EDS<sup>[6]</sup> compared to experimental results.<sup>[25]</sup> (Right): relative hydration free energies for two sets of benzene derivatives obtained with MBAR<sup>[26,27]</sup> and RE-EDS<sup>[7]</sup> compared to experimental results.<sup>[26,27]</sup>



### 3.3 Validation of *amber2gromos*: Relative Hydration Free Energies of Small Benzene Derivatives

A third study<sup>[7]</sup> focused on the validation of the GAFF topologies converted with *amber2gromos* for RE-EDS simulations in GROMOS. Two sets of benzene derivatives were selected from the FreeSolv<sup>[26,27]</sup> database. A small set (labeled A) containing six molecules was used as an initial sanity check. To assess the accuracy of the relative free-energy differences in vacuum/water, complementary single topology TI simulations were carried out with the GROMACS<sup>[28]</sup> MD engine for set A. In order to explore the performance of multistate RE-EDS, a larger set (labeled B) with 28 molecules was selected. To investigate whether the performance is adversely affected by the larger number of end-states, set B was additionally divided into two subsets (Ba and Bb, with one molecule in common). For all sets, *amber2gromos* was used to generate the topologies from the input GAFF topologies, *RestraintMaker* to generate appropriate distance restraints, and the RE-EDS pipeline to carry out the simulations. The results were compared to the experimental and calculated (using MBAR) values reported in FreeSolv (Fig. 3, right). With an overall RMSE of 2.6 kJ/mol against experimental data and 1.1 kJ/mol against the MBAR results, the agreement was excellent. The relative hydration free energies obtained with RE-EDS from the two combined subsets Ba and Bb were slightly more accurate than the results for the full set B (RMSE of 2.5 kJ/mol versus 2.6 kJ/mol). As for the previously described study, the required simulation time was considerably shorter for RE-EDS than for TI (at least a factor 6 for set A) and MBAR (about a factor 5.5 for set A, 4.5 for set B).

### 4. Conclusions

We provided an overview of the RE-EDS pipeline and the complementary tools *RestraintMaker* and *amber2gromos*. Combined, they provide a convenient, robust, and largely automated workflow for the setup and execution of free-energy calculations with RE-EDS in GROMOS.

The results from three applications of RE-EDS to calculate relative binding or hydration free energies were briefly discussed. In all three studies, the relative free energies obtained with RE-EDS accurately reproduced experimental values, as well as results obtained with different free energy methods, even for a large number (e.g. 28) of end-states. A major advantage of the RE-EDS pipeline is the comparably small amount of total simulation time required.

### Acknowledgements

The authors gratefully acknowledge financial support by the Swiss National Science Foundation (Grant no. 200021-178762 and no. 200021-175944). We would like to thank the jurors of the Best Poster Presentation Award at the SCS Fall Meeting 2021, as well as the award's sponsor, DSM Nutritional Products.

- [8] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 300, <https://doi.org/10.1063/1.1749657>.
- [9] C. H. Bennett, *J. Comput. Phys.* **1976**, *22*, 2, [https://doi.org/10.1016/0021-9991\(76\)90078-4](https://doi.org/10.1016/0021-9991(76)90078-4).
- [10] M. R. Shirts, J. D. Chodera, *J. Chem. Phys.* **2008**, *129*, 124105 <https://doi.org/10.1063/1.2978177>.
- [11] S. Riniker, C. D. Christ, W. F. van Gunsteren, *J. Chem. Phys.* **2007**, *126*, 184110, <https://doi.org/10.1063/1.2730508>.
- [12] C. D. Christ, W. F. van Gunsteren, *J. Chem. Phys.* **2008**, *134*, 229901, <https://doi.org/10.1063/1.2913050>.
- [13] U. H. E. Hansmann, *Chem. Phys. Lett.* **1997**, *281*, 1, [https://doi.org/10.1016/S0009-2614\(97\)01198-6](https://doi.org/10.1016/S0009-2614(97)01198-6).
- [14] Y. Sugita, A. Kitao, Y. Okamoto, *J. Chem. Phys.* **2000**, *113*, 6042, <https://doi.org/10.1063/1.1308516>.
- [15] S. Riniker, C. D. Christ, H. S. Hansen, P. H. Hünenberger, C. Oostenbrink, D. Steiner, W. F. van Gunsteren, *J. Phys. Chem. B* **2011**, *115*, 46, <https://doi.org/10.1021/jp204303a>.
- [16] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 9, <https://doi.org/10.1002/jcc.20035>.
- [17] J. Wang, W. Wei, P. A. Kollman, D. A. Case, *J. Am. Chem. Soc.* **2001**, *123*, U403.
- [18] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 2, <https://doi.org/10.1002/wcms.1121>.
- [19] S. Riniker, *J. Chem. Inf. Model.* **2018**, *58*, 3, <https://doi.org/10.1021/acs.jcim.8b00042>.
- [20] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, W. F. van Gunsteren, *Comp. Phys. Comm.* **2012**, *183*, 890, <https://doi.org/10.1016/j.cpc.2011.12.014>.
- [21] A. P. Eichenberger, J. R. Allison, J. Dolenc, D. P. Geerke, B. A. C. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D. Wang, W. F. van Gunsteren, *J. Chem. Theory Comput.* **2011**, *7*, 10, <https://doi.org/10.1021/ct2003622>.
- [22] X. Huang, C. C. Cheng, T. O. Fischmann, J. S. Duca, X. Yang, M. Richards, G. W. Shipps Jr, *ACS Med. Chem. Lett.* **2012**, *3*, 2, <https://doi.org/10.1021/ml200249h>.
- [23] L. Wang, Y. Deng, Y. Wu, B. Kim, D. N. LeBard, D. Wandschneider, M. Beachy, R. A. Friesner, R. Abel, *J. Chem. Theory Comput.* **2017**, *13*, 1, <https://doi.org/10.1021/acs.jctc.6b00991>.
- [24] W. Jaspers, M. Esguerra, J. Åqvist, H. Gutiérrez-de-Terán, *J. Cheminf.* **2019**, *11*, 26, <https://doi.org/10.1186/s13321-019-0348-5>.
- [25] M. Stroet, B. Caron, K. M. Visscher, D. P. Geerke, A. K. Malde, A. E. Mark, *J. Chem. Theory Comput.* **2018**, *14*, 11, <https://doi.org/10.1021/acs.jctc.8b00768>.
- [26] D. L. Mobley, J. P. Guthrie, *J. Comput. Aided Mol. Des.* **2014**, *28*, 711, <https://doi.org/10.1007/s10822-014-9747-x>.
- [27] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, D. L. Mobley, *J. Chem. Eng. Data.* **2017**, *62*, 5, <https://doi.org/10.1021/acs.jced.7b00104>.
- [28] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX.* **2015**, *1*, 19–25, <https://doi.org/10.1016/j.softx.2015.06.001>.

### License and Terms



This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions (<https://chimia.ch/chimia/about>).

The definitive version of this article is the electronic one that can be found at <https://doi.org/10.2533/chimia.2022.327>

Received: January 27, 2022

- [1] C. Christ, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2009**, *31*, 8, <https://doi.org/10.1002/jcc.21450>.
- [2] Z. Courmia, B. Allen, W. Sherman, *J. Chem. Inf. Model.* **2017**, *57*, 12, <https://doi.org/10.1021/acs.jcim.7b00564>.
- [3] D. Sidler, A. Schwaninger, S. Riniker, *J. Chem. Phys.* **2016**, *145*, 154114, <https://doi.org/10.1063/1.4964781>.
- [4] D. Sidler, M. Cristófol-Clough, S. Riniker, *J. Chem. Theory Comput.* **2017**, *13*, 6, <https://doi.org/10.1021/acs.jctc.7b00286>.
- [5] B. Ries, K. Normak, R. G. Weiss, S. Rieder, E. P. Barros, C. Champion, G. König, S. Riniker, *J. Comput. Aided Mol. Des.* **2022**, online, <https://doi.org/10.1007/s10822-021-00436-z>.
- [6] B. Ries, S. R. Rieder, C. Rhiner, P. H. Hünenberger, S. Riniker, *J. Comput. Aided Mol. Des.* **2021**, <https://doi.org/10.1007/s10822-022-00445-6>.
- [7] S. R. Rieder, B. Ries, K. Schaller, C. Champion, E. P. Barros, P. H. Hünenberger, S. Riniker, *J. Chem. Inf. Model.* **2022**, submitted; preprint on <https://arxiv.org/abs/2204.01396>.