Accelerating Alchemical Free Energy Prediction Using a Multistate Method: Application to Multiple Kinases

Working Paper

Author(s):
Champion, Candide; Gall, René; Ries, Benjamin; Rieder, Salomé R.; Barros, Emilia P.; Riniker, Sereina

Publication date:
2023-09-14

Permanent link:
https://doi.org/10.3929/ethz-b-000640321

Originally published in:
ChemRxiv, https://doi.org/10.26434/chemrxiv-2023-kfh2s-v2

Funding acknowledgement:
212732 - Combining Molecular Dynamics and Machine Learning for Free Energy Calculation with Quantum-Mechanical Accuracy (SNF)

Accelerating Alchemical Free Energy Prediction Using a Multistate Method: Application to Multiple Kinases

Candide Champion[a], René Gall[a], Benjamin Ries[a,†], Salomé R. Rieder[a], Emilia P. Barros[a], Sereina Riniker[a]*

[a] Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland. *Email: sriniker@ethz.ch

[†] Current adress: Boehringer Ingelheim Pharma GmbH & Co KG, Medicinal Chemistry, Birkendorfer Str 65, 88397 Biberach an der Riss, Germany

Abstract

Alchemical free-energy methods based on molecular dynamics (MD) simulations have become important tools to identify modifications of small organic molecules that improve their protein binding affinity during lead optimization. The routine application of pairwise free-energy methods to rank potential binders from best to worst is impacted by the combinatorial increase of calculations to perform when the number of molecules to assess grows. To address this fundamental limitation, our group has developed replica-exchange enveloping distribution sampling (RE-EDS), a pathway-independent multistate method, enabling the calculation of alchemical free-energy differences between multiple ligands ($N > 2$) from a single MD simulation. In this work, we apply the method to a set of four kinases with diverse binding pockets, and their corresponding inhibitors (42 in total), chosen to showcase the general applicability of RE-EDS in prospective drug design campaigns. We show that for the targets studied, RE-EDS is able to model up to 13 ligands simultaneously with high sampling efficiency, leading to a substantial decrease in computational cost when compared to pairwise methods.

1 Introduction

Free energy is among the most fundamental quantities in chemistry. Any spontaneous chemical process (under constant pressure/temperature) is accompanied by a decrease in free energy. In the context of drug design, the properties of a drug candidate must satisfy specific criteria, which can be related to differences in free energy, e.g., binding affinity, selectivity, solubility, and passive membrane permeability [1]. During hit-to-lead optimization, accurate prediction of binding affinities allows to prioritize and reduce the number of compounds to synthesize [2, 3]. Approximate methods such as docking have been used for decades to estimate binding affinities [4], although they fail to account for entropic contributions [5], the role of specific water molecules in binding [6–8], *etc.*. Recent advances in computing power have enabled the routine application of rigorous free-energy methods based on molecular dynamics (MD) simulations such as free-energy perturbation (FEP) [9] or thermodynamic integration (TI) [10] in both academic and industrial settings [11, 12].

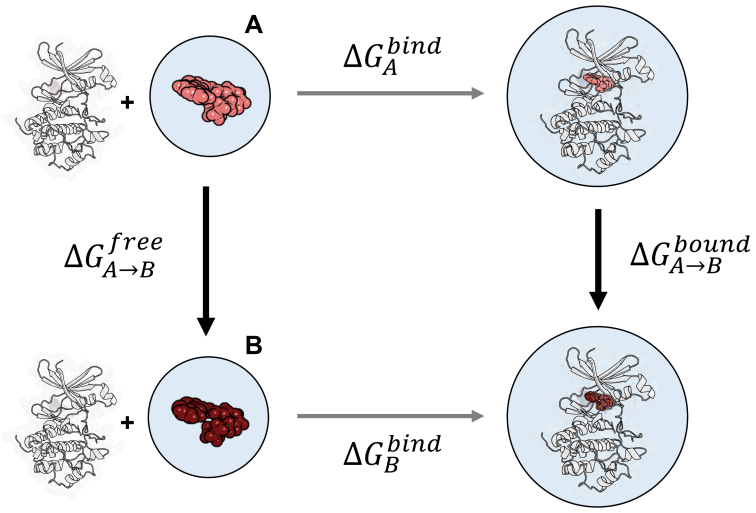

Figure 1: Thermodynamic cycle to estimate the relative binding free energy between inhibitors A and B. Simulating the alchemical pathway (bold black arrows) is computationally more feasible than the physical pathway (grey arrows). The relative binding free energy is obtained with Eq. (1).

While calculating the physical binding pathway of an inhibitor is possible [13, 14], it is computationally more feasible to estimate relative binding free energies (RBFE) [15] by performing "alchemical" transformations between analogous inhibitors (Figure 1 and Eq. (1)).

$$\begin{aligned}\Delta\Delta G_{A\to B} &= \Delta G_{A\to B}^{bound} - \Delta G_{A\to B}^{free} \\ &= \Delta G_B^{bind} - \Delta G_A^{bind}\end{aligned} \tag{1}$$

The main drawback of such pairwise methods is that the number of calculations to perform grows exponentially with the number of compounds to be screened [16], as accuracy relies on the simulation of many of the $\binom{n}{2}$ possible pairwise combinations [17]. To limit this combinatorial problem, automated workflows have been developed to estimate the smallest and/or most efficient perturbation map for a given set of ligands [17–20]. Despite these efforts, finding a general solution is no trivial task and remains an open question [17, 21, 22]. In this context, simulating many such alchemical transformations from the same MD simulation would thus decrease significantly the computational costs of the free-energy method, further improving their applicability in the drug design process.

The existing multistate free-energy methods can be grouped into two main families. The first set of methods is based on λ-dynamics [23–26], and the second group on enveloping distribution sampling (EDS) [27, 28]. In the former method, the system follows well defined pathways to transition from state to state, whereas no such constraints are imposed in EDS (i.e., it is "pathway independent"), which may provide additional flexibility for sampling. Transitions in alchemical space may be accelerated by coupling these methods to enhanced sampling algorithms, enabling a more rapid convergence of the free-energy differences. λ-dynamics based methods have been combined with the tailor-made adaptive landscape flattening algorithm [29], while EDS has been combined with well established algorithms, giving rise to replica-exchange EDS (RE-EDS) [30–34] and accelerated EDS [35, 36]. In this work, we aim to showcase the transferability of the RE-EDS method to multiple protein targets with distinct binding pockets. To that end, we have constructed a dataset of inhibitors for four different kinases, containing a wide range of chemical functional groups and scaffolds.

The kinase family of enzymes catalyzes the phosphorylation of proteins, contributing to the mediation of a wide range of critical biological processes such as metabolism, transcription, cytoskeletal rearrangement, apoptosis, and inter-cellular communication [37]. Deregulation of kinase function has been linked to various diseases such as autoimmune, cardiovascular, inflammatory, and nervous diseases as well as cancer [38, 39]. Over the past 20 years, drug discovery programs have focused heavily on

developing inhibitors targeting kinases, accounting for a significant part (20 – 33%) of drug discovery efforts worldwide [40]. While 68 FDA-approved drugs targeting more than 20 kinases already exist, research in the field remains very active, highlighted by the large number of additional compounds currently undergoing clinical trials [41]. In this work, we have applied the RE-EDS method to series of inhibitors of the following four kinases: checkpoint kinase 1 (CHK1) [42], NF-κB inducing kinase (NIK) [43], p21-activated kinase 1 (PAK) [44], and the proviral insertion in murine lymphoma kinase 1 (PIM) [45]. All four enzymes belong to the most common subclass of serine/threonine kinases [40]. All inhibitors examined in this work (Figure 4) bind in the hinge region connecting the terminal N- and C-lobes, in competition with the natural adenosine triphosphate (ATP) cofactor [46].

In summary, the main objective of the present work is to demonstrate the transferability of the RE-EDS method to multiple protein targets with distinct binding pockets, highlighting it as a viable and computationally less demanding alternative to conventional FEP and TI calculations. We present methodological improvements, which enable us to expand both the scope and number of alchemical perturbations that can be performed in a single RE-EDS simulation, further showcasing its suitability to calculate RBFE in prospective drug design campaigns. All simulations in this work were performed with two small-molecule force fields (GAFF [47] and OpenFF [48]), which allows us to discern deviations stemming from lack of sampling and from force-field inaccuracies. We will discuss the current limitations of RE-EDS and compare them to other state-of-the-art free-energy methods. While our study focuses on biologically relevant kinases, RE-EDS can of course also be applied to other protein families as well as solvation free energies (ΔG_{solv}) of small organic molecules [33, 34], or to investigate water thermodynamics in binding pockets [49].

2 Theory

RBFE calculations involve the simulation of the alchemical transformation of a molecule into another, and the adaptation of the environment in which this alteration is taking place (e.g., protein binding pocket). State-of-the-art methods such as FEP and TI introduce a coupling parameter λ to connect end states together. Independent simulations are then carried out at discrete λ-points along the path connecting the two end states, and free-energy differences can be obtained using the TI [10] or MBAR estimator [50, 51].

2.1 Replica-Exchange Enveloping Distribution Sampling (RE-EDS)

In EDS, the different end states are connected in a reference state R with the form [27, 28],

$$V_R(\mathbf{r}) = -\frac{1}{\beta s} \ln \sum_{i=1}^{N} \exp(-\beta s[V_i(\mathbf{r}) - E_i]), \tag{2}$$

where $\beta = (k_B T)^{-1}$, s is the smoothing parameter, and E_i the energy offset for each end state i. The force acting on a particle k is calculated using Eq. 3, which is simply a weighted contribution of the forces exerted by all end states (Eq. 4).

$$\mathbf{f}_k = -\frac{\partial V_R(\mathbf{r})}{\partial \mathbf{r}_k} = \sum_{i=1}^{N} w_i \left(-\frac{\partial V_i(\mathbf{r})}{\partial \mathbf{r}_k} \right) \tag{3}$$

$$w_i = \frac{\exp(-\beta s[V_i(\mathbf{r}) - E_i])}{\exp(-\beta s\, V_R(\mathbf{r}))} \tag{4}$$

Essentially, forces will be largest for the end state for which the current conformation is most compatible, and forces exerted by states in unfavourable conformations (e.g., overlap with the protein) are scaled to approximately zero, without requiring soft-core potentials [29, 52]. This reference potential-energy surface "envelops" the potential-energy surfaces of all end states, ensuring that all minima of the individual end states are minima of the reference state (Figure 2A). In order to sample the minima corresponding to

all end states in a reasonable amount of simulation time, two parameters were introduced to the EDS reference state [28]: the energy offsets (E_i), which enable equal sampling of all end states, and the smoothness parameter (s), which reduces energy barriers to enhance transitions. Finding appropriate reference-state parameters is not trivial, as the ideal energy offsets (at $s = 1$, assuming infinite simulation) correspond to the free-energy differences we seek to calculate [53]. Similarly, the s-values alter the shape of the reference potential-energy surface, and need to be chosen carefully [53, 54]. Excessive reduction of the s parameter leads to a distortion of the reference potential energy surface, such that its minima no longer maps to minima corresponding to physical conformations for any of the end states. We denote this unphysical regime "undersampling" (yellow curve in Figure 2B) [54].

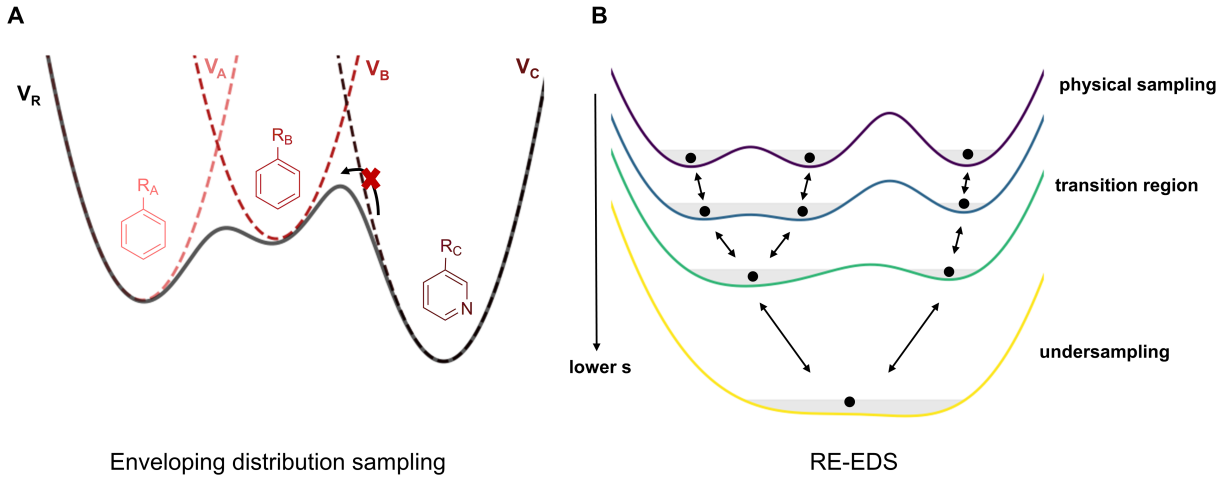

Figure 2: (**A**): "Enveloping" nature of the reference potential V_R (in dark grey), which encloses the minima of the potential-energy surfaces of three small molecules (colored dashed curves). If no energy offsets or s-values are used to scale minima or barriers heights, transitions between state C and B are unlikely (indicated with a red cross). (**B**): Effect of lowering the s-value on a reference potential V_R (all colored lines), with energy offsets chosen such that all minima are aligned at $s = 1$. Grey areas indicate regions of phase space that are sampled, and black dots and arrows represent possible replica-exchanges in RE-EDS.

In RE-EDS [30–32], multiple EDS simulations with different s-values are carried out in parallel, and exchange trials are performed between neighboring replicas every few steps. All sets of conformations (starting at distinct s-values) are thus able to 'visit' replicas at lower s-values that enable transitions between end states, as well as higher s-values where physically meaningful conformations are sampled. Free-energy differences between all pairwise combinations of end states i and j can then be recovered by employing the Zwanzig equation (Eq. 5) from the simulation data gathered at $s = 1$, or more optimally by combining data from all replicas with MBAR [51]. Coupling EDS to replica exchange was found to simplify the choice of appropriate s parameters [31], and affords enhanced transitions between states.

$$\Delta G_{i\to j} = -\frac{1}{\beta}\ln\frac{\langle\exp(-\beta[V_j - V_R])\rangle_R}{\langle\exp(-\beta[V_i - V_R])\rangle_R}\,. \tag{5}$$

Obtaining accurate free-energy differences thus relies on the ability to sample conformations relevant to all end states in the simulation at $s = 1$. Doing so depends crucially on the energy offsets and s-parameters, which are determined from a set of short iterative simulations prior to the production run (Figure 3). Ries *et al.* [32] have recently proposed an improved automated workflow to optimize these parameters. In the following paragraphs, we briefly describe each step of the RE-EDS pipeline. A more complete description and details are given in the Supporting Information (Section S2), and in previous publications [30–32].

First, starting coordinates have to be generated where the environment (solvent, protein) is well adapted for each of the end states. Starting from optimized coordinates was shown to improve the

accuracy of RE-EDS calculations [55], in particular when an experimental crystal structure is not available for all end states. This procedure, termed "end-state generation", consists of N independent EDS simulations, in which energy offsets are chosen to enforce sampling of one specific end state. The lowest energy conformation in each of these simulations is then used as input for all subsequent RE-EDS simulations ("starting state mixing", SSM). In parallel, the lower-bound search consists of very short EDS simulations performed at different s-values to determine when the system reaches undersampling. Then, a first guess for the energy offsets is estimated from free-energy differences in the undersampling regime (Eq. 5) of a RE-EDS simulation with default parameters (i.e., $E_i = 0$, logarithmically distributed s-parameters between 1.0 and the lower bound). Typically, a gap region with low exchange frequencies is observed during this step (see Figure S1), which can be crudely be filled prior to the optimization phase, making the latter more efficient (see details in Supporting Information Section S1).

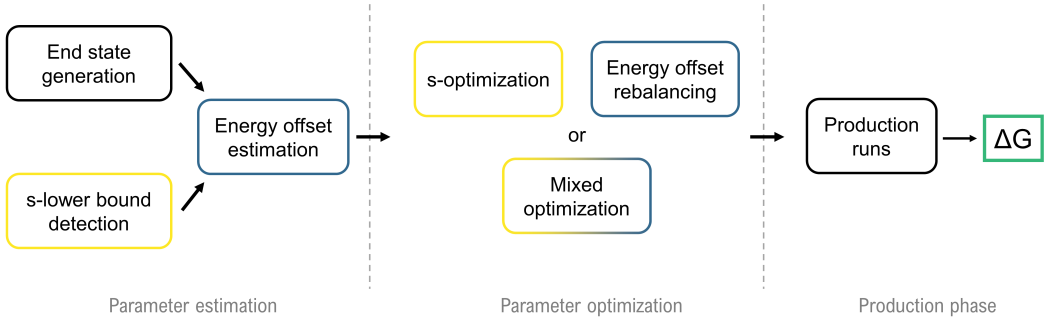

Figure 3: Schematic overview of the steps to calculate free-energy differences with RE-EDS. The steps during the parameter estimation and optimization phases are needed to acquire information about the system to improve the s-values (circled in yellow), or the energy offsets (circled in blue).

Subsequently, the distribution of s-values and energy offsets are further optimized in an iterative process, until specific criteria are met: no replica-exchange bottlenecks and approximately equal sampling of all end states at $s = 1$. Each iteration consists of a short RE-EDS simulation from which exchange probabilities and/or sampling distributions are extracted to guide the update of RE-EDS parameters. The parameters can be refined separately one after the other as proposed by Ries *et al.*(e.g., s-optimization followed by energy-offset optimization), or may be updated simultaneously ("mixed optimization" described in this study). Optimization of the s-distribution may be achieved by either adding replicas in regions of s-space with low exchange rates based on local criteria (N-LRTO algorithm [31]), or by redistributing the replicas in s-space to maximize the replica-exchange probability over the entire range of s-values (N-GRTO algorithm [31]). Note that the N-GRTO algorithm could also be used to add replicas, but this was not done in this study. If there are already some transitions between all replicas with the initial s-distribution, the N-GRTO algorithm is more robust and often requires less replicas than N-LRTO. The latter algorithm was developed mainly for cases with severe exchange bottlenecks, where N-GRTO fails. Optimization of the energy offsets is accomplished with the "energy-offset rebalancing" approach [32], which updates the energy offsets based on the current sampling of each end state (at $s = 1$). Sampling is determined with the "maximally contributing" [32] criterion, which assign the current frame to the end state which contributes most to the forces (largest w_i in Eqs. 3 and 4). Finally, the RE-EDS production phase is carried out with the parameters obtained, and the final free-energy differences are extracted from the simulation at $s = 1$ (or all s-values when using MBAR).

2.2 Comparison to Experiment

All simulated RBFE values were compared to experimental values (K_i or IC_{50}) by first converting them to absolute binding free energies (based on the known experimental binding free energies of the N inhibitors, Eq. 6) following the procedure introduced by Wang *et al.* [16]. We note that this conversion, which is simply a shift with respect to average values, does not alter the interpretation of the results in a prospective study (experimental affinities unknown) as the ranking of inhibitors and the differences

among them remain identical. We also followed recommendations from Mey *et al.* [56] to perform a bootstrapping analysis. This was performed by drawing new simulated binding free energies from normal distributions with standard deviations equal to those obtained from the five repeats started with different velocities. Note that the alignment with respect to the experimental values was not repeated after adding noise to the data based on the observed errors.

$$\Delta G_i^{sim} = \Delta\Delta G_{i\to R}^{sim} - \left(\frac{\sum \Delta\Delta G_{i\to R}^{sim}}{N} - \frac{\sum \Delta G_i^{exp}}{N} \right). \tag{6}$$

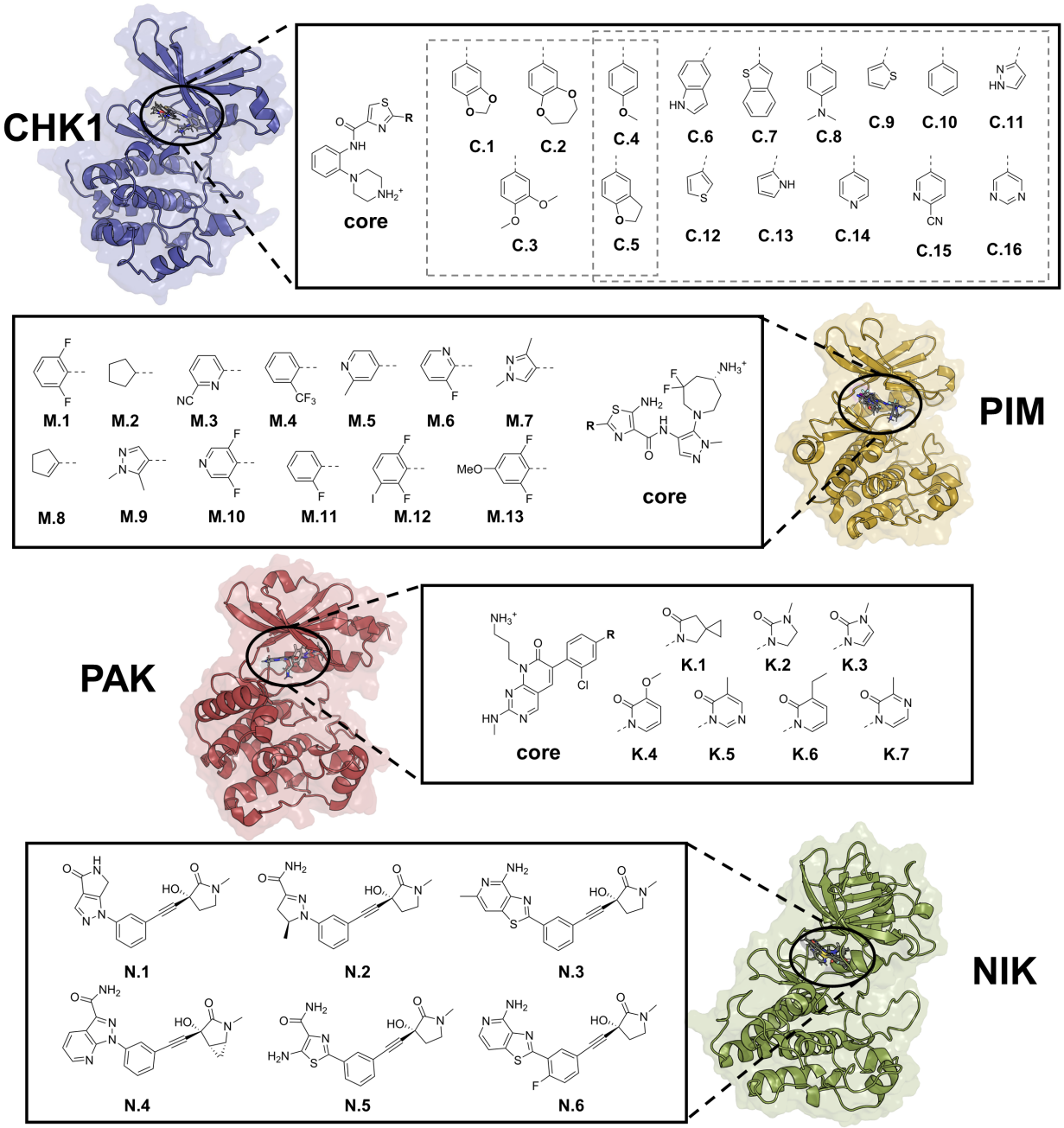

Figure 4: The four kinases and set of inhibitors used in this study. For CHK1, PIM, PAK, and NIK, the core is shown together with the different substituents and corresponding label. For CHK1, we studied two subsets separately, indicated by dashed lines. Molecules **C.4** and **C.5** are part of both subsets.

3 Methods

3.1 Datasets

The main focus of this work is to demonstrate the transferability of the RE-EDS methodology to different protein targets, as well as its capability to simulate a wide range of distinct chemical perturbations. Increasingly complex sets of inhibitors were chosen to showcase the generality of RE-EDS. Details of the assay conditions and crystal structure preparation are given for the different targets in the respective publications (NIK [57, 58], PAK [59, 60], PIM [61], CHK1 [62]).

We built sets of congeneric inhibitors for each target, with at least one crystal structure of the inhibitor-protein complex available, which could be used as starting structure. We also followed guidelines by Mey *et al.* [56] in the construction of the datasets to avoid known sources of error in free-energy calculations (e.g., small spread of experimental binding affinities, inhibitors with multiple protonation/tautomeric state at pH 7, unknown stereo-isomer in the experimental assay, etc.). For backward comparison, we simulated the same set of CHK1 inhibitors from our previous work (Ries *et al.* [32]), and in addition constructed a larger set of 13 ligands from the same experimental data [62]. The four kinases and the selected inhibitors are shown in Figure 4.

3.2 Topology Preparation and Parametrization

All initial topologies were prepared with the AmberTools [63] suite of programs. Unresolved residues in the crystal structure were added using SwissModel [64], and buried water molecules (if any) were retained. Solvent molecules were added ensuring a minimum distance of 1.2 nm between solute atoms and the edge of the box. All protein parameters were taken from the AMBER ff14SB force field [65], and the TIP3P water model was used [66]. The inhibitors were modeled with two different small-molecule force fields using AM1-BCC charges [67]: the general AMBER force field (GAFF) 1.8 [47], and the Open Force Field (OpenFF) version 2 ("Sage") [48, 68]. Topologies with both force fields, as well as initial coordinates were generated in the AMBER format and then translated into GROMOS format using the *amber2gromos* program [33].

Alchemical free-energy calculations may be performed with different representations of the perturbed region. We chose to represent this region with a "hybrid topology" [55, 69], where common core atoms between inhibitors are represented once, and all varying substituents represented separately. We opted for this representation to avoid hindering any motion of the common core with distance restraints, which are necessary to keep ligands overlayed with one another in RE-EDS simulations with the "dual topology" representation [32, 55]. Note that the same atom may have a different partial charge (and Lennard Jones parameters) according to the various end states in the hybrid topology representation. This allows to take into account electron withdrawing (or donating) effects of a perturbed group on the core region, while maintaining an integer net charge for each ligand, and without requiring any charge re-normalization protocol as in some λ-dynamics implementations [70].

Hybrid topologies were generated by combining the topologies of the individual inhibitors using the RDKit [71] and PyGromosTools [72] to select an appropriate core structure with a maximum common substructure (MCS) search and adding the different substituent atoms and parameters (bonded and non-bonded) to the initial topology. An example Jupyter notebook to build such a hybrid topology can be found in the GitHub repository (https://github.com/rinikerlab/reeds/blob/main/examples/input_preparation/hybrid_topology_maker.ipynb). For the NIK dataset, additional calculations using a dual topology representation (as in our previous work [32]) was investigated for comparison, imposing weak harmonic distance restraints ($k_{restraint} = 5000$ kJ mol^{-1} nm^{-2}) between selected core atoms of each inhibitor to prevent drifts (see Figure S3). The starting coordinates for the hybrid topologies were built by taking protein coordinates from a reference crystal structure and ligand coordinates from pre-aligned crystal structures if available (PAK, NIK), or by generating coordinates for the substituents using the constrained embedding functionality in the RDKit (CHK1, PIM).

3.3 Simulation Details

All simulations were performed using the GROMOS software package [73]. The program is freely available on www.gromos.net. The simulations were propagated using a leap-frog algorithm with an integration time step of 2 fs. All bond lengths were constrained with the SHAKE algorithm [74] and a relative tolerance of 10^{-4}. Non-bonded interactions were calculated up to an atomic cutoff of 1 nm (pairlist updated every step), and electrostatic contributions beyond the cutoff were modeled by a reaction field with a relative permittivity $\epsilon_{RF} = 66.7$. The simulations were carried out in the NpT thermodynamic ensemble by maintaining a temperature of 298 K and a pressure of 0.06102 kJ mol^{-1} nm^{-3} ($\approx$ 1 atm) with a Berendsen thermostat and barostat, respectively [75]. The translational and rotational motion of the center of mass of the box was removed every 1000 steps. Energies and coordinates of the system were recorded every 1000 steps (2 ps).

3.3.1 RE-EDS Simulations

RE-EDS simulations were performed by following the pipeline introduced by Ries *et al.* [32, 76]. The complete set of simulation details (lengths, initial conformations given, number of parameter optimization iterations, *etc.*) are given in Section S1 of the Supporting Information. Replica-exchange trials were carried out every 50 steps. After the parameter optimization phase, production runs of 5 ns were performed, of which the first 1 ns was discarded as equilibration. Five replicates with different initial velocities were performed, from which average values were obtained and standard deviations used to represent errors [77, 78]. To calculate the final free-energy differences, the data acquired at all s-values was incorporated using the MBAR procedure [51]. For this, the values of the reference potential-energy of each replica need to be re-evaluated with the reference potential parameters (s-values and/or energy offsets) of all other replicas, allowing us to optimally connect all thermodynamic states together. This re-evaluation can be simply done as a post-processing step. Free-energy differences with respect to the reference state at $s = 1$ ($\Delta G^{bound}_{i\rightarrow R}$ and $\Delta G^{free}_{i\rightarrow R}$) are then passed into Eq. (6).

3.3.2 TI Simulations

For the NIK dataset, we performed additionally TI calculations for comparison. Here, only the OpenFF was tested. TI calculations were performed using 21 evenly spaced λ-values between 0 and 1. Five replicates with different initial velocities were performed for 5 ns at each λ-value. Hamiltonian replica-exchange trials were performed every 50 steps between neighbouring λ-values to enhance sampling [79].

4 Results and Discussion

First, we present the modifications made to the RE-EDS pipeline and discuss the resulting gains in efficiency. Second, we examine the binding free energies calculated with RE-EDS for each dataset, and explain observed deviations. Subsequently, we discuss our results more generally, with an emphasis on the differences between the inhibitors/binding pockets studied, and how these relate to differences in the RE-EDS simulations. Finally, we compare the cumulative simulation time with state-of-the-art methods, and discuss future improvements expected to amplify the gains in computational cost provided by RE-EDS.

4.1 Improvements to the RE-EDS Pipeline

In this work, we build upon the pipeline described by Ries *et al.* [32] to set-up the system and optimize the reference-state parameters (Figure 3). Modifications to the workflow are proposed, which allowed us to fine tune and shorten the parameter optimization phase. First, we take advantage of the information regarding exchanges (or lack thereof) between replicas during the energy offset estimation step to pre-optimize the s-distribution. Typically, the default logarithmic s-distribution contains a "gap region" where the replica-exchange probability is low (see Figure S1 in the Supporting Information), especially

for simulations with the ligands in the protein binding pocket. Previously [31, 32], this problem was tackled using the N-LRTO algorithm to add replicas in the exchange bottleneck region until sufficient exchanges were obtained. By using the information about the location of the gap region from the energy offset estimation step and placing a large number of s-values in this gap region (uniformly on a log scale) prior to the s-optimization step, we were able to bridge the gap sufficiently to directly use the N-GRTO algorithm to optimize the s-distribution. We found that adding replicas in the gap region to reach a total of 32 replicas worked well for all systems studied here (see details in Supporting Information Section S1). This hyperparameter is analogous to the number of λ-values in FEP/TI simulations (often 21 values).

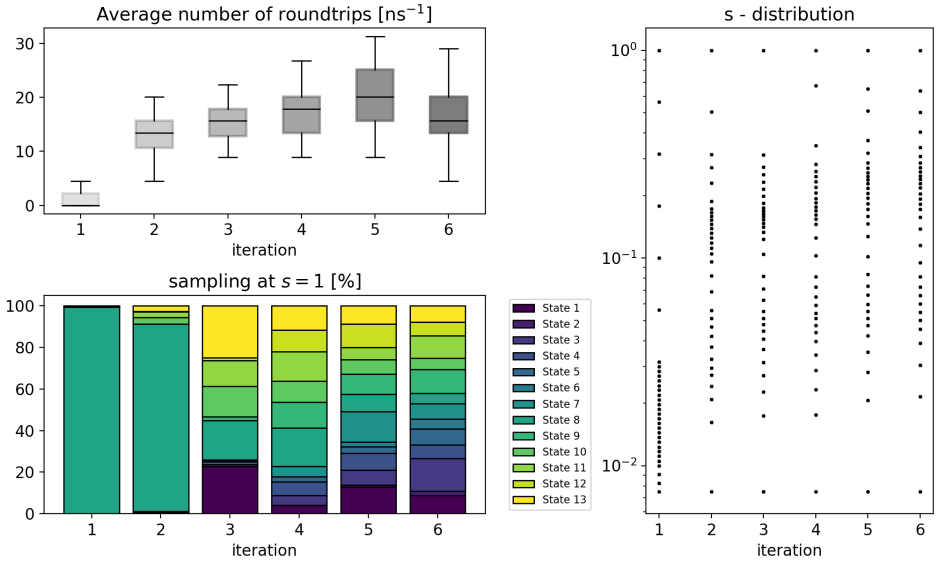

Figure 5: Mixed optimization of the RE-EDS parameters for the CHK1 dataset (ligands in water) with the GAFF force field. (Upper left): Average number of round trips made by all replicas. The error bars correspond to the standard deviations among the replicas. (Lower left): Sampling of the end states in the uppermost replica ($s = 1$). (Right): Distribution of the replicas in s-space.

We found that the optimal set of energy offsets may depend on the current s-distribution. As a result of this coupling of parameters, an optimized set of parameters (e.g., energy offsets), may require re-optimization upon changes to the other parameters (e.g., s-values). To facilitate the optimization of connected parameters, we explored the adjustment of both sets simultaneously (termed "mixed optimization"). This is particularly important if the initial guess for energy offsets leads to preferential sampling of only one end state, which we observed in one of our simulations (Figure 5). In this example, we noticed that the s-distribution changed significantly between iterations 1 and 3 where ligand **C.11** is predominantly sampled, and shifted again when the updated energy offsets led to more equal sampling of the end states between iterations 3 and 6. More generally, we found that the s-distribution converges close to an ideal solution within a few iterations (2 or 3, see Figure S2 in the Supporting Information), and the remaining iterations mostly involve fine-tuning of the energy offsets. This suggests that applying the N-GRTO algorithm onto an optimal s-distribution does not further modify it, supporting the robustness of the mixed optimization protocol.

In summary, we recommend optimization of the parameters via a mixed optimization protocol. In this work, the parameter optimization phase typically represents around 20% of the total simulation time (see discussion below). Note that the 500 ps used for each iteration was chosen based on our experience from previous simulations [31, 32]. The procedure in its current form may still be further refined, which is part of future work. Faster convergence towards optimal parameters may for example be obtained by performing an initial set of short optimization steps, followed by longer iterations to fine tune the parameters [25].

4.2 Binding Free Energies for the Kinase Inhibitors

The four different kinases will be discussed in order of increasing complexity in terms of the magnitude and number of alchemical perturbations as well as the flexibility of the binding site.

4.2.1 p21-Activated Kinase 1 (PAK)

The PAK dataset was constructed to highlight the transformations of 5-membered and 6-membered aromatic and non-aromatic heterocycles as might be explored in lead optimization. Note that the substituent modified here is in the buried region of the binding pocket, which is generally considered to be more challenging to simulate, as rearrangement of the protein pocket may be slow on the simulation time scale. However, all functional groups occupy a very similar volume and all ligands form the same key hydrogen bond between the carbonyl group oxygen of the modified substituent and Lys299 (Figure 6B). Overall, the binding free energies obtained with GAFF and OpenFF are in good agreement with experimental values, with a mean unsigned error (MUE) of 1.9 and 2.8 kJ mol^{-1} respectively, and all but one data point lie within the chemical accuracy limit [56] of 1 kcal mol^{-1} ($=$ 4.184 kJ mol^{-1}) (Figure 6A).

Table 1: Statistical metrics used to evaluate the accuracy of binding free energies calculated for the PAK dataset. Numbers reported corresponds to the mean, with upper and lower bounds of the 95% confidence interval in square brackets.

Force field	MUE [kJ mol^{-1}]	Kendall τ	Spearman ρ
GAFF	1.9 [1.80 – 2.01]	0.4 [0.33 – 0.46]	0.5 [0.45 – 0.60]
OpenFF	2.8 [2.66 – 2.89]	0.6 [0.57 – 0.65]	0.7 [0.71 – 0.76]

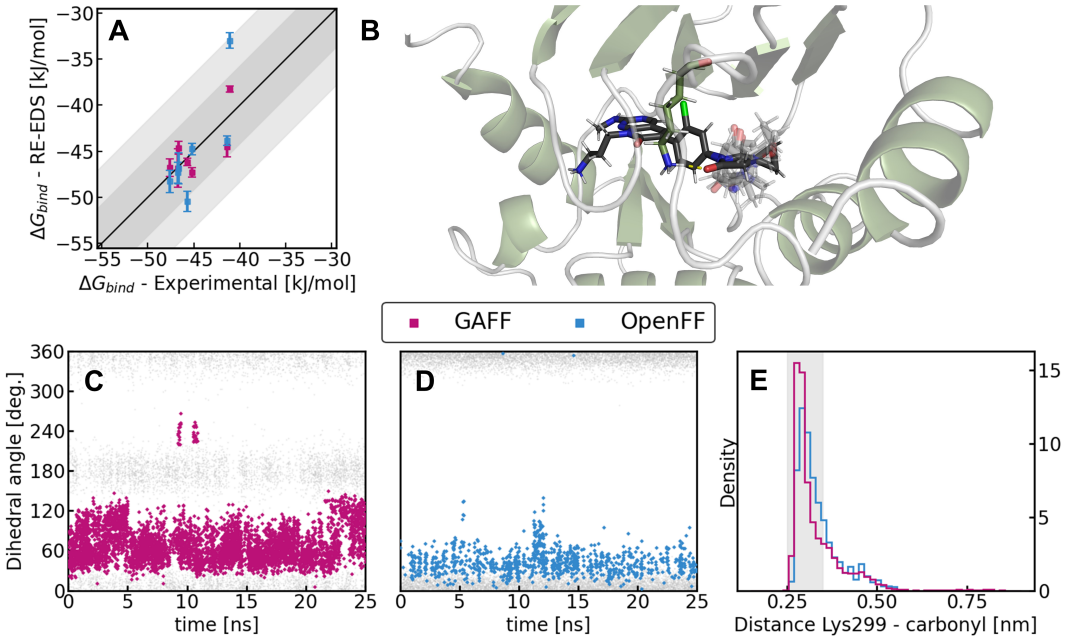

Figure 6: Results obtained for the PAK dataset. (**A**): Comparison of calculated and experimental binding free energies. Dark and light grey regions correspond to margins of error of 1 and 2 kcal mol^{-1}, respectively. (**B**): Binding pose of the RE-EDS system, in a conformation favorable for state 1. The core and perturbed functional group for state 1 are shown in black, whereas the non-sampled states are shown in light transparent black. The side chain of Lys299 is shown in dark green sticks. (**C** – **D**): Time series of the dihedral angle between the core and the perturbed substituent of ligand **K.1** in the simulations with GAFF (purple, **C**) and OpenFF (blue, **D**). Grey points correspond to the dihedral angle at all time steps, whereas colored points correspond to frames where ligand **K.1** is actively sampled in the reference state. (**E**): Probability distribution of the distance between the carbonyl oxygen of ligand **K.1** and the terminal nitrogen atom of Lys299 from frames in which ligand **K.1** is actively sampled. The grey region corresponds to distances compatible with the presence of the stabilizing hydrogen bond.

The largest deviation compared to experiment was observed for ligand **K. 1** with OpenFF. Inspection of the rotation around the bond connecting the modified substituent and the core in the simulations with GAFF and OpenFF (Figure 6C-D) showed some differences in dihedral-angle sampling between the force fields, which led to small differences for adopting the key hydrogen bond with Lys299 (Figure 6E), which occurs when the dihedral angle is around $120°$. Although the hydrogen bond was present in both simulations, the population differences may explain the larger deviation of ΔG_{bind} with OpenFF. This finding highlights the importance of accurate torsion parameters for binding free-energy calculations.

4.2.2 NF-κB Inducing Kinase (NIK)

The NIK dataset displays greater variety in structural modifications among the inhibitors with ring opening (e.g., ligand **N.1** compared to the others), fusion of rings (e.g., **N.1**, **N.3**, **N.4**, and **N.6**), and a second site of modification (e.g., **N.4** and **N.6**), adding different layers of complexity.

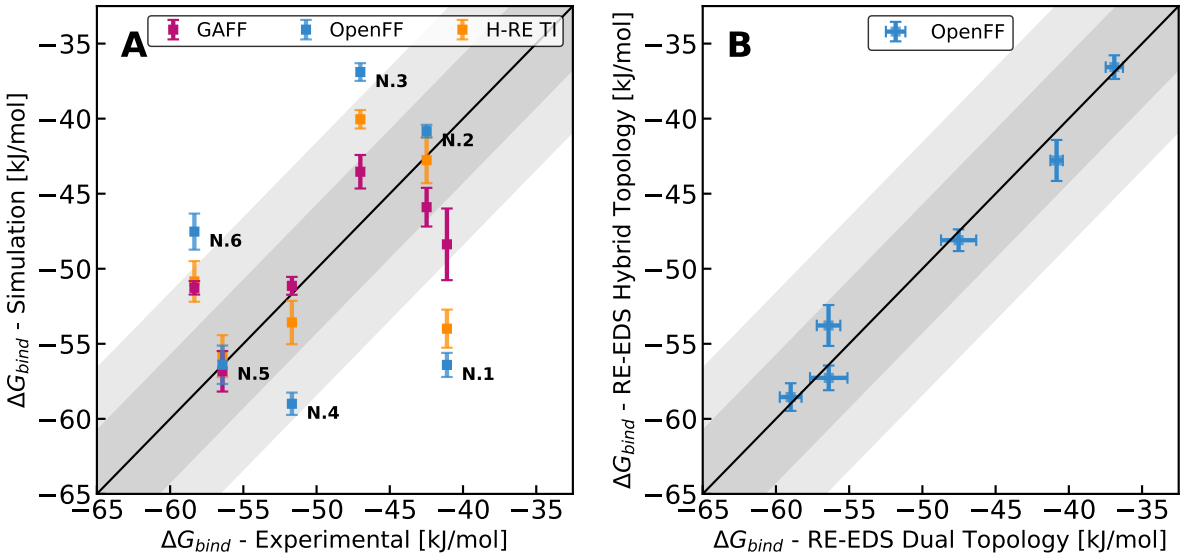

Figure 7: Results obtained for the NIK dataset: Comparison of calculated and experimental binding free energies. Dark and light grey regions correspond to margins of error of 1 and 2 kcal mol^{-1}, respectively. (**A**) Results from RE-EDS simulations with dual topologies were obtained with GAFF (purple) and OpenFF (blue), and reference H-RE TI calculations (orange) were performed for comparison with OpenFF. (**B**) Comparison of the results of RE-EDS simulations with the OpenFF (blue) using dual and hybrid topologies.

In the case of NIK, the results obtained with GAFF are within the margin of error with an MUE of 3.8 kJ mol^{-1}, in contrast to OpenFF with an MUE of 7.7 kJ mol^{-1} (Figure 7). The largest deviation compared to experiment with OpenFF was found for ligand **N.1** (15.3 kJ mol^{-1}). Detailed analysis of this cases pointed to insufficient sampling of the protein re-organization as the main source of error. First, we analyzed the binding poses of the six different ligands based on the crystal structures and identified the key interactions between protein and ligand. In all six crystal structures, the catalytically important DFG activation loops are well aligned (Figure S5 in the Supporting Information), indicating that the inhibitors interact in a similar fashion with this part of the binding pocket. On the other hand, the Arg410 side chain adopts various conformations in the crystal structures, some with and some without a hydrogen bond with either the ligand, Glu475, or Leu474. In the RE-EDS simulations, we find three key conformations of the protein with respect to Arg410, which we denote "open", "closed 1", and "closed 2" (Figure 8). The system adopted an open or closed conformation approximately 50% of the time with both GAFF and OpenFF. However, the "closed 2" form in which Arg410 interacts with Leu474 as well as the ligand occurs more prominently in the simulations with OpenFF (Figure 8D), which may explain the differences between the two force fields. As the crystal structures vary among the different inhibitors, it is not trivial to interpret them, but there is no evidence for the presence of this "closed 2" form (Figure S5).

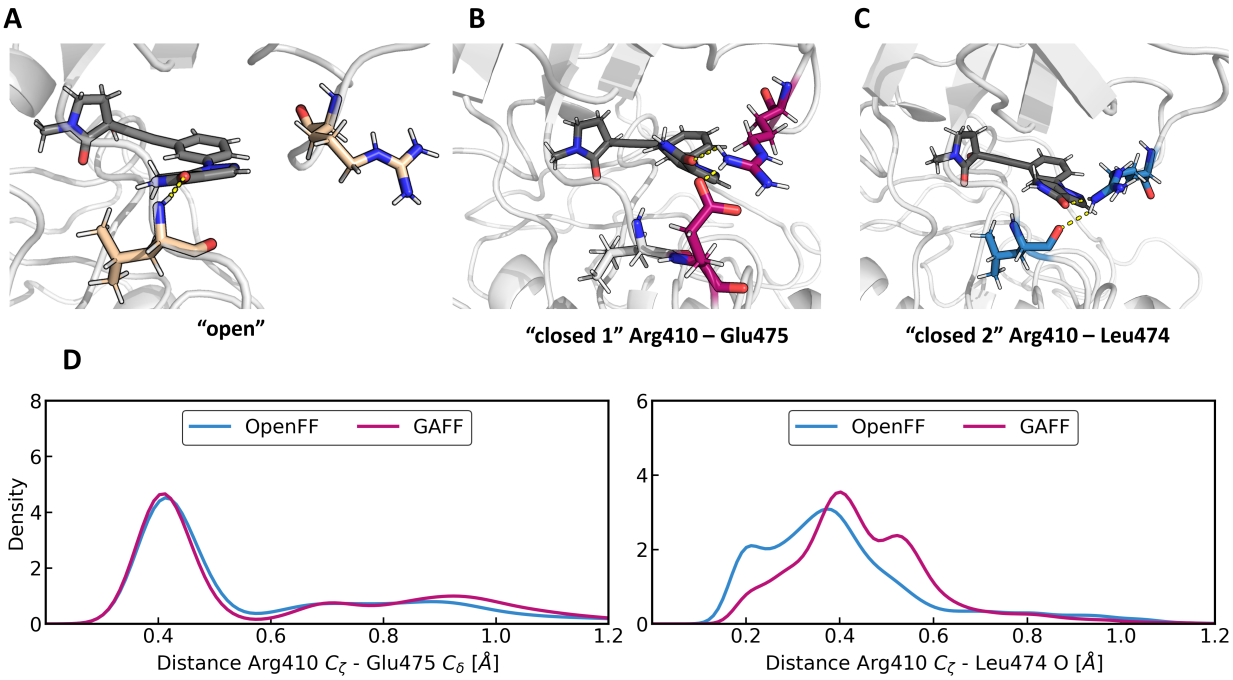

Figure 8: (**A-C**): Three key conformations of Arg410 seen in the simulations of the protein binding to ligand **N.1**. (**D**): Distributions of the distance between Arg410 and Glu475 (left) and between Arg410 and Leu474 (right) from the combined trajectories (five repeats concatenated). Blue and purple lines correspond to simulations with OpenFF and GAFF, respectively.

Table 2: Statistical metrics used to evaluate the accuracy of binding free energies calculated for the NIK dataset using RE-EDS and H-RE TI calculations with either hybrid or dual topology representation. Numbers reported corresponds to the mean, with upper and lower bounds of the 95% confidence interval in square brackets.

Method	Simulation	MUE [kJ mol^{-1}]	Kendall τ	Spearman ρ
RE-EDS	GAFF (dual)	3.8 [3.64 – 4.04]	0.4 [0.36 – 0.47]	0.6 [0.56 – 0.69]
RE-EDS	OpenFF (dual)	7.7 [7.59 – 7.93]	0.0 [-0.03 – 0.03]	0.1 [0.10 – 0.19]
RE-EDS	OpenFF (hybrid)	7.0 [6.90 – 7.19]	0.1 [0.06 – 0.10]	0.3 [0.26 – 0.28]
H-RE TI	OpenFF (dual)	5.3 [5.14 – 5.48]	0.1 [0.06 – 0.15]	0.2 [0.13 – 0.25]

To further investigate whether force-field deficiencies or sampling are the main cause of deviations, we performed H-RE TI simulations with the same force field (OpenFF). We found overall good agreement with the RE-EDS results (Figure 7A), but a lower MUE of 5.3 kJ mol^{-1} compared to experiment. We note that the binding affinity of the largest outlier in the RE-EDS simulations (**N.1**) is also predicted poorly by H-RE TI, and the increased accuracy mainly stems from slightly better predictions for other ligands (**N.2**, **N.3**, **N.4** and **N.6**). This suggests that incomplete sampling in the RE-EDS simulations with OpenFF may at least be partially responsible for the deviations observed. In agreement with our previous analysis, we find that the Arg410 side chain position in the H-RE TI calculations varied for the different ligands (Figure S7 in the Supporting Information), suggesting that this degree of freedom may indeed be the source of part of the observed error in both RE-EDS and H-RE TI simulations.

The sampling issues may also be related to motion slower than the trajectory, which is supported by observations from Shih et al. [80], who reported that FEP+ results for inhibitors of the NIK kinase may differ by up to 8.4 kJ mol^{-1} when using a starting structure prepared with a different workflow to model the missing loop residues. Finally, we note that the simulations with a hybrid topology perform slightly better than those with a dual topology representation (Table 2 and Figure 7B). This result may be explained by the larger number of round trips observed in the simulations with a hybrid topology

(~ 5.6 round-trips per ns) than with dual topology (~ 2.3 round trips per ns), thus favouring transitions between states. Consequently, we recommend the use of the hybrid topology representation for binding free energy calculations of ligands with a common core due to the advantages they confer (e.g., no hindering of motion with distance restraints) [55, 69], as shown below for other datasets.

4.2.3 Checkpoint Kinase 1 (CHK1)

The CHK1 dataset consists of two subsets of five and 13 ligands, respectively. The first subset of five ligands is the same as studied by Wang *et al.* [81] with FEP+, Jespers *et al.* [19] with QligFEP, and Ries *et al.* [32] with RE-EDS in combination with the GROMOS force field. The five ligands include challenging ring growing and ring opening modifications. The second subset was prepared from the same set of published experimental values [62], choosing molecules for which the experimental IC_{50} values were spread over a broader range. This second subset focuses on the alchemical modifications of aromatic rings (five-membered, six-membered, and fused rings).

The results for the smaller subset are very good with MUE of 1.7 and 1.5 kJ mol^{-1} for GAFF and OpenFF, respectively (Figure 9), which is lower than previous results obtained with GROMOS force-field parameters (MUE of 2.8 kJ mol^{-1}). This confirms the hypothesis that deviations seen in Ries *et al.* can be attributed mostly to force-field parametrization (done manually based on similar molecules). These results are also very similar to those obtained for the same system with other methods (FEP+ [81] and QligFEP [19] (Figure S8 in the Supporting Information). Note that the standard deviations in the RE-EDS simulations with GAFF and OpenFF are generally smaller than those observed with FEP+ and QligFEP as not all possible pairwise transformations were simulated with the latter methods.

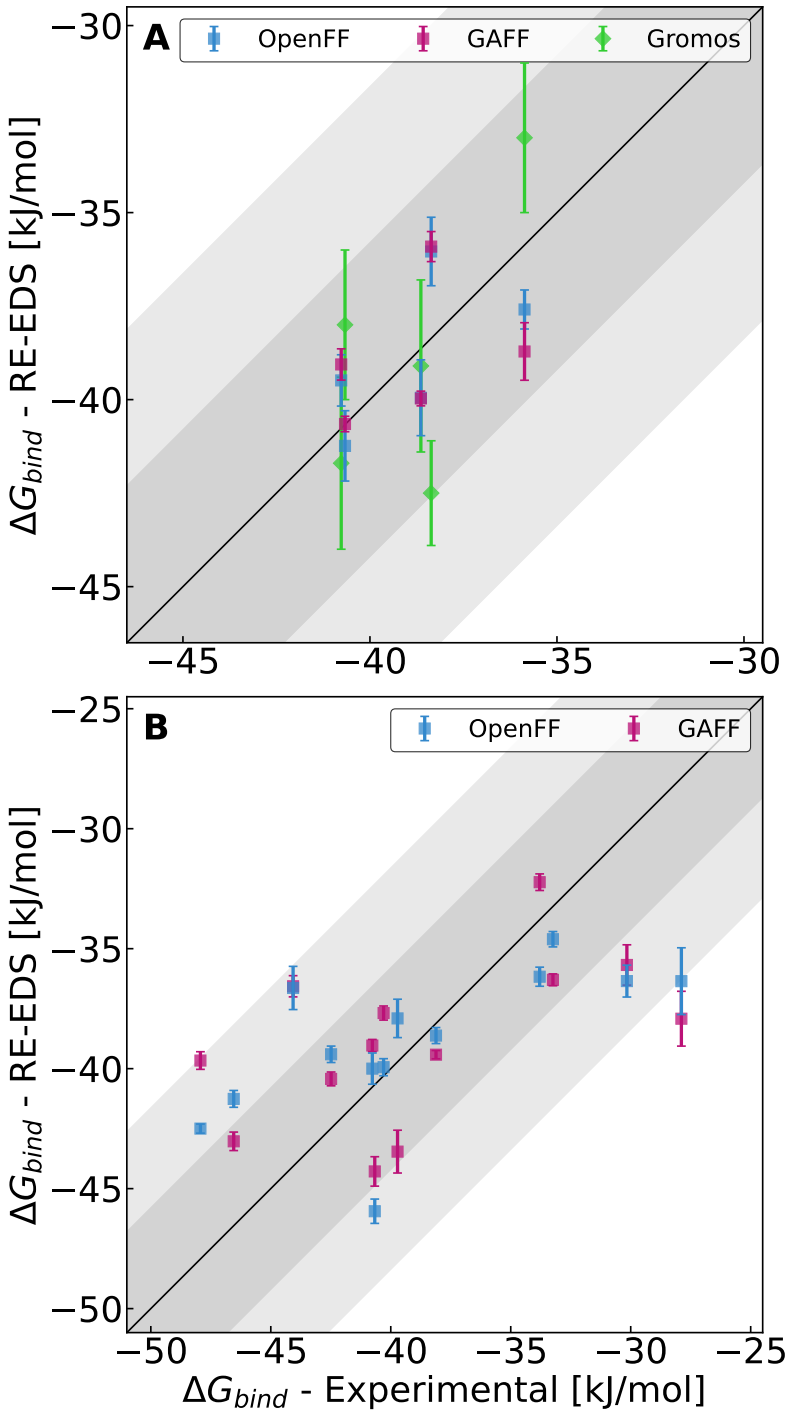

Figure 9: Results obtained for the CHK1 dataset. (**A**): Comparison of calculated and experimental binding free energies for the smaller subset of five inhibitors. Dark and light grey regions correspond to margins of error of 1 and 2 kcal mol^{-1}, respectively. Results are shown with GAFF (purple) and OpenFF (blue), as well as results taken from Ries *et al.*[32] performed with the GROMOS force field (green) for comparison. Additional comparisons to other state-of-the-art methods can be found in Figure S9 in the Supporting Information. (**B**): Results obtained for the larger set of 13 inhibitors.

Results for the subset of 13 ligands show a good performance and demonstrate that more ligands can be included in the same RE-EDS simulation. For both force fields, we find average errors below the margin of error (1 kcal mol^{-1}) and the correlation coefficients show a reasonable ranking of the ligands. Most of the remaining deviations can likely be attributed to the force field as the sampling of all end states is almost uniform (Figure S9 in the Supporting Information). Sampling is facilitated in this system by the modified substituent being solvent exposed. As re-arrangement of water molecules occurs on a

faster timescale than re-arrangement of a protein side chains or loops, the transition between end states should be easier.

Table 3: Statistical metrics used to evaluate the accuracy of binding free energies calculated for the CHK1 datasets. Numbers reported corresponds to the mean, with upper and lower bounds of the 95% confidence interval in square brackets.

Simulation	MUE [kJ mol^{-1}]	Kendall τ	Spearman ρ
GAFF (subset 1)	1.7 [1.61 – 1.78]	0.3 [0.27 – 0.38]	0.5 [0.36 – 0.55]
OpenFF (subset 1)	1.5 [1.39 – 1.67]	0.5 [0.39 – 0.52]	0.7 [0.60 – 0.71]
Gromos FF (subset 1)	2.8 [2.47 – 3.05]	0.3 [0.20 – 0.39]	0.4 [0.30 – 0.49]
GAFF (subset 2)	4.2 [4.16 – 4.24]	0.3 [0.31 – 0.34]	0.5 [0.47 – 0.51]
OpenFF (subset 2)	3.7 [3.69 – 3.78]	0.5 [0.52 – 0.56]	0.7 [0.67 – 0.71]

4.2.4 Proviral Insertion in Murine Lymphoma Kinase 1 (PIM)

The PIM dataset was constructed to further validate the ability of RE-EDS to simulate more than 10 alchemical modifications of five- and six-membered rings with potentially bulky substituents (e.g., -CN in **M.3**, and -CF_3 in **M.4**). In summary, the 13 ligands were sampled well in all simulations (except one, see discussion below), while the resulting binding free energies show overall a slightly larger deviation from experimental values compared to the CHK1 dataset (Table 4). The five repeats of the production runs showed larger deviations for the PIM system than for CHK1 (see Figures S8 and S10). These differences result directly from less frequent transitions between the end states, which depend strongly on the nature of the binding pocket. Upon inspection of the binding pose (see Figure 10), we hypothesize that accommodation of the bulky -CF_3 of ligand **M.4** requires a rearrangement of the neighbouring Arg122 side chain, which could explain the larger standard deviations in ΔG_{bind} observed for this ligand (rearrangement of Arg122 is unlikely to occur in short ns simulations typically performed in free-energy calculations). When comparing the rotation around the bond connecting the core and perturbed substituent of ligand **M.4** (see Figures S11 and S12), we found that this dihedral angle remained close to the initial conformation (-CF_3 pointing towards Arg122, dihedral angle of 360°) over the whole simulation with OpenFF, while it rotates with GAFF in the "dummy state" (i.e., when the system samples other end states) such that the bulky -CF_3 substituent points in the other direction, which is an unfavourable conformation hindering transitions to that end state. We note that this does not impact the other ortho-substituted ligands (**M.4**, **M.10**, **M.12**, and **M.13**) as those are symmetric. For the PIM dataset, ligand **M.4** was sampled in one of the five random repeats with GAFF >0.1% of the time due to this issue. Thus, the average values for $\Delta G^{bound}_{4\rightarrow R}$, which were subsequently converted to ΔG_{bind}, were obtained based only on four repeats (Figure 10 and Table 4). A more detailed discussion of the implications of lack of sampling of end states is provided in the following section.

Table 4: Statistical metrics used to evaluate the accuracy of binding free energies calculated for the PIM datasets. Numbers reported corresponds to the mean, with upper and lower bounds of the 95% confidence interval in square brackets.

Simulation	MUE [kJ mol^{-1}]	Kendall τ	Spearman ρ
GAFF	4.6 [4.56 – 4.71]	0.2 [0.15 – 0.17]	0.3 [0.24 – 0.27]
OpenFF	4.9 [4.83 – 4.94]	0.3 [0.28 – 0.30]	0.5 [0.43 – 0.46]

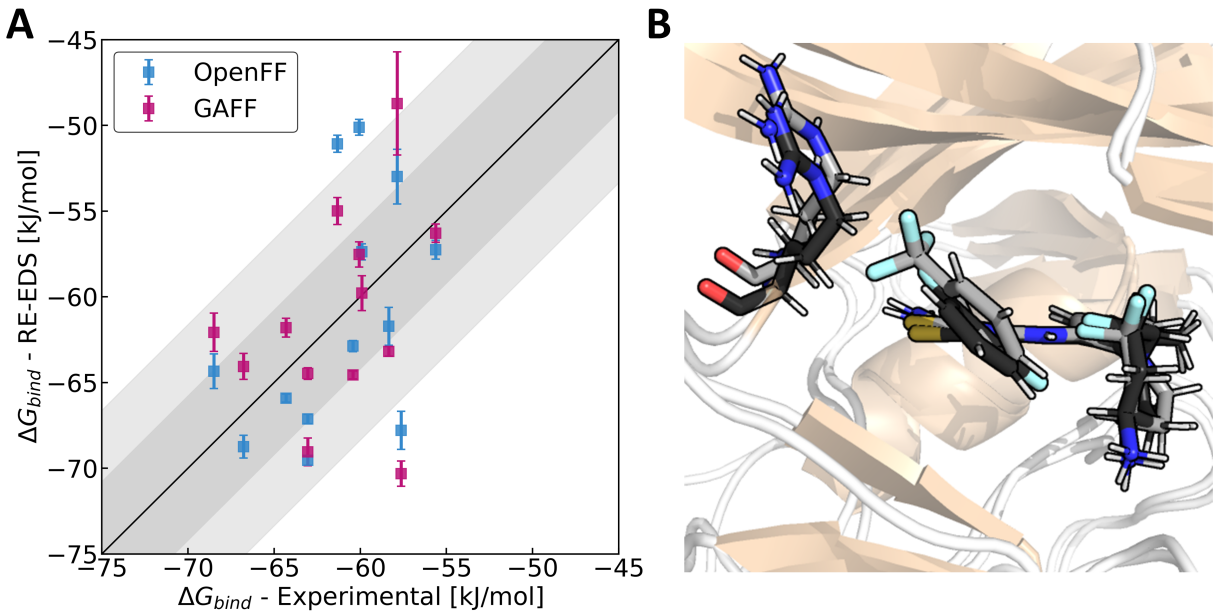

Figure 10: (**A**): Comparison of calculated and experimental binding free energies for the PIM dataset. Dark and light grey regions correspond to margins of error of 1 and 2 kcal mol^{-1}, respectively. Results are shown with GAFF (purple) and OpenFF (blue). (**B**) SSM conformations (minimum energy conformers from a 1 ns simulation) for ligands **M.1** (black) and **M.4** (light grey). Alignment was performed by using the atoms of the maximum common core to showcase the shift in position of the Arg122 side chain (colored as the corresponding ligand) required to accommodate the bulkier -CF_3 substituent.

4.3 General Discussion

4.3.1 Sources of Errors

The results shown demonstrate that the RE-EDS methodology can be applied to different protein-ligand systems, allowing to calculate RBFEs within the 1 kcal mol^{-1} margin of error, which is the common performance criterion for free-energy methods (see recent large-scale comparisons [11, 16, 25, 82]). As with standard free-energy methods like FEP or TI, deviations may stem from either incomplete sampling, inaccurate force-field parameters, or a combination thereof, and it is often difficult to disentangle them [83].

In the presented simulations, the main source of deviations appeared to be related to dihedral-angle rotations when a ligand is not actively sampled in the RE-EDS simulation (i.e., the protein environment does not "see" it). If an unfavourable configuration is reached in this "dummy state", transitions to this end state become more difficult. Sampling of degrees of freedom with timescales similar to that of the simulation is a known problem for free-energy methods [22, 83]. In its current implementation in GROMOS, the EDS reference state (Eq. 2) includes only the non-bonded potential energy. Inclusion of the dihedral-angle terms may be a way to address this issue for RE-EDS.

4.3.2 Simulation Time Requirements

Calculating binding free energies with RE-EDS provides a significant decrease in computational time compared to standard pairwise methods like FEP or TI. The performance enhancement obtained grows with the number of ligands (n) included simultaneously (Figure 11 and Table S6 in the Supporting Information). Note that the hyper-parameters for FEP/TI (21 λ-values with 5 ns each) were chosen based on common procedures, although there is no strict consensus on the best parameters to use, and it is generally accepted that they are system dependent [11, 12, 56, 84]. Furthermore, we report estimates for FEP/TI with both the minimum number of pairwise calculations ($n-1$) and for the

maximum number ($\binom{n}{2}$) of edges, as the actual number of calculations will lie in between and depends on the user. Recent work from Pitman *et al.* suggests that the minimum number of edges to simulate from a perturbation graph scales with respect to $O(n \log n)$ [85], based on a rigorous statistical analysis assuming normally distributed errors (with $\sigma = 1.0$ kcal mol^{-1}) on a set of synthetic data. Simulating a more challenging system may, however, require inclusion of a larger number of edges to satisfy the same stability in precision. Multi-scale methods like RE-EDS, on the other hand, will likely have a better scaling because the interactions between the unperturbed particles in the system has to be calculated only once and not for each pair of ligands.

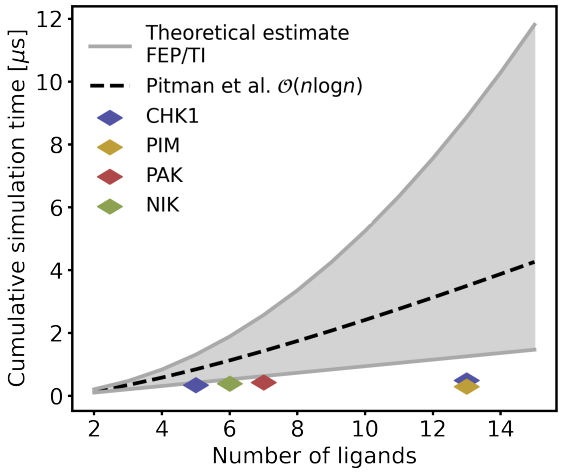

Figure 11: Comparison of the cumulative sampling time in the RE-EDS simulations performed in this work (colored diamonds) to the expected range of corresponding FEP/TI sets of simulations (grey). The black dashed line corresponds to the minimum number of edges to simulate in a perturbation graph to maintain a stable precision in free-energy differences [85].

The results shown in Figure 11 are for a single production run. It is, however, widely acknowledged that performing multiple repeats (as done in this study) provides more accurate results [77, 78]. The speed-up provided by RE-EDS is amplified when repeats of the production run are performed because the preparation phase does not need to be repeated (Table S7 in the Supporting Information). Although not all ligands were sampled in all repeats for the more complex PIM dataset, we would like to highlight that it was still possible to estimate all binding free energies with reasonable accuracy. Importantly, lack of sampling of one specific ligand in a simulation did not impact in any way the results for the other end states.

5 Conclusions

In this study, we have applied the multi-state free-energy method RE-EDS to calculate protein-ligand binding free energies of a set of four kinases and their inhibitors (42 in total). The ligands involved relatively large modifications, changes in ring size, ring opening, buried and solvent-exposed substituents. The results demonstrate that the method is suitable to estimate relative binding affinities for > 10 ligands simultaneously, giving a substantial decrease in cumulative simulation time compared to conventional methods like FEP and TI. We have presented a set of improvements made to the RE-EDS pipeline, in particular a combined optimization of the energy offsets and s-distribution.

Akin to other free-energy methods, results obtained with RE-EDS are also affected by force-field deficiencies and practical limitations in simulation time restricting the sampling of all accessible conformations. For the sets of molecules studied in this work, we did not observe a clear superiority of one of the two force fields studied. OpenFF appeared to perform better for all systems except for the NIK dataset. These observations are in line with the benchmarking study reported by OpenFF developers

where GAFF and OpenFF show similar performance [48]. In addition to deviations arising from larger conformational changes of the protein (e.g., re-organization of a loop), which cannot be expected to be sampled within a few nanoseconds [86], we have identified that dihedral-angle rotations of the ligand in the "dummy state" may hinder transitions to this ligand, affecting sampling and thus the resulting free-energy differences. This issue will be addressed in future developments of the methodology.

In conclusion, RE-EDS has been shown to be an attractive and efficient free-energy method , with a pipeline readily available on GitHub [76] and freely available implementations in GROMOS [87] and OpenMM [34].

Data and Software Availability

The *reeds* module for parameter optimization is freely available on Github at https://github.com/rinikerlab/reeds. The equilibrated protein-ligand coordinates and topology files for the simulations can be obtained at https://github.com/rinikerlab/reeds_applications, as well as example input files for the RE-EDS simulations.

Acknowledgments

The authors gratefully acknowledge financial support by the Swiss National Science Foundation (grant number 200021_212732). The authors thank Genentech for providing access to unpublished data and Nicholas Skelton for helpful discussions. The authors thank Felix Pultar for his review of the manuscript.

References

[1] L. Frye, S. Bhat, K. Akinsanya, R. Abel, *Drug Discov. Today Technol.* **2021**, *39*, 111–117.

[2] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. Zhang, T. Hou, *Chem. Rev.* **2019**, *119*, 9478–9508.

[3] P. B. Cox, R. Gupta, *ACS Med. Chem. Lett.* **2022**, *13*, 1016–1029.

[4] L. Pinzi, G. Rastelli, *Int. J. Mol. Sci.* **2019**, *20*, 4331.

[5] J. D. Chodera, D. L. Mobley, *Annu. Rev. Biophys.* **2013**, *42*, 121–142.

[6] P. W. Snyder, J. Mecinović, D. T. Moustakas, S. W. Thomas III, M. Harder, E. T. Mack, M. R. Lockett, A. Héroux, W. Sherman, G. M. Whitesides, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17889–17894.

[7] C. Barillari, J. Taylor, R. Viner, J. W. Essex, *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.

[8] F. Spyrakis, M. H. Ahmed, A. S. Bayden, P. Cozzini, A. Mozzarelli, G. E. Kellogg, *J. Med. Chem.* **2017**, *60*, 6781–6827.

[9] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420–1426.

[10] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 300–313.

[11] V. Gapsys, L. Pérez-Benito, M. Aldeghi, D. Seeliger, H. Van Vlijmen, G. Tresadern, B. L. de Groot, *Chem. Sci.* **2020**, *11*, 1140–1152.

[12] C. E. M. Schindler, H. Baumann, A. Blum, D. Böse, H.-P. Buchstaller, L. Burgdorf, D. Cappel, E. Chekler, P. Czodrowski, D. Dorsch, M. K. I. Eguida, B. Follows, T. Fuchß, U. Grädler, J. Gunera, T. Johnson, C. Jorand Lebrun, S. Karra, M. Klein, T. Knehans, L. Koetzner, M. Krier, M. Leiendecker, B. Leuthner, L. Li, I. Mochalkin, D. Musil, C. Neagu, F. Rippmann, K. Schiemann, R. Schulz, T. Steinbrecher, E.-M. Tanzer, A. Unzue Lopez, A. Viacava Follis, A. Wegener, D. Kuhn, *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.

[13] H. Kokubo, T. Tanaka, Y. Okamoto, *J. Chem. Theory Comput.* **2013**, *9*, 4660–4671.

[14] V. Limongelli, M. Bonomi, M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6358–6363.

[15] Z. Cournia, B. Allen, W. Sherman, *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.

[16] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel, *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

[17] Q. Yang, W. Burchett, G. S. Steeno, S. Liu, M. Yang, D. L. Mobley, X. Hou, *J. Comput. Chem.* **2020**, *41*, 247–257.

[18] L. Carvalho Martins, E. A. Cino, R. S. Ferreira, *J. Chem. Theory Comput.* **2021**, *17*, 4262–4273.

[19] W. Jespers, M. Esguerra, J. Åqvist, H. Gutiérrez-de Terán, *J. Cheminf.* **2019**, *11*, 26.

[20] S. Liu, Y. Wu, T. Lin, R. Abel, J. P. Redmann, C. M. Summa, V. R. Jaber, N. M. Lim, D. L. Mobley, *J. Comput. Aided Mol. Des.* **2013**, *27*, 755–770.

[21] H. Xu, *J. Chem. Inf. Model.* **2019**, *59*, 4720–4728.

[22] D. F. Hahn, C. I. Bayly, H. E. B. Macdonald, J. D. Chodera, A. S. Mey, D. L. Mobley, L. P. Benito, C. E. Schindler, G. Tresadern, G. L. Warren, *LiveCoMS* **2022**, *4*, 1497.

[23] J. L. Knight, C. L. Brooks III, *J. Chem. Theory Comput.* **2011**, *7*, 2728–2739.

[24] J. Z. Vilseck, K. A. Armacost, R. L. Hayes, G. B. Goh, C. L. Brooks III, *J. Phys. Chem. Lett.* **2018**, *9*, 3328–3332.

[25] E. P. Raman, T. J. Paul, R. L. Hayes, C. L. Brooks III, *J. Chem. Theory Comput.* **2020**, *16*, 7895–7914.

[26] J. Z. Vilseck, X. Ding, R. L. Hayes, C. L. Brooks III, *J. Chem. Theory Comput.* **2021**, *17*, 3895–3907.

[27] C. D. Christ, W. F. van Gunsteren, *J. Chem. Phys.* **2007**, *126*, 184110.

[28] C. D. Christ, W. F. van Gunsteren, *J. Chem. Phys.* **2008**, *128*, 174112.

[29] R. L. Hayes, K. A. Armacost, J. Z. Vilseck, C. L. Brooks III, *J. Phys. Chem. B* **2017**, *121*, 3626–3635.

[30] D. Sidler, A. Schwaninger, S. Riniker, *J. Chem. Phys.* **2016**, *145*, 154114.

[31] D. Sidler, M. Cristòfol-Clough, S. Riniker, *J. Chem. Theory Comput.* **2017**, *13*, 3020–3030.

[32] B. Ries, K. Normak, R. G. Weiss, S. Rieder, E. de Barros, C. Champion, G. König, S. Riniker, *J. Comput. Aided Mol. Des.* **2022**, *36*, 117–130.

[33] S. R. Rieder, B. Ries, K. Schaller, C. Champion, E. P. Barros, P. H. Hünenberger, S. Riniker, *J. Chem. Inf. Model.* **2022**, *62*, 3043–3056.

[34] S. R. Rieder, B. Ries, A. Kubincová, C. Champion, E. P. Barros, P. H. Hünenberger, S. Riniker, *J. Chem. Phys.* **2022**, *157*, 104117.

[35] J. W. Perthold, C. Oostenbrink, *J. Phys. Chem. B* **2018**, *122*, 5030–5037.

[36] J. W. Perthold, D. Petrov, C. Oostenbrink, *J. Chem. Inf. Model.* **2020**, *60*, 5395–5406.

[37] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **2002**, *298*, 1912–1934.

[38] F. M. Ferguson, N. S. Gray, *Nat. Rev. Drug Discov.* **2018**, *17*, 353–377.

[39] P. Cohen, D. Cross, P. A. Jänne, *Nat. Rev. Drug Discov.* **2021**, *20*, 551–569.

[40] R. Roskoski, *Pharmacol. Res.* **2021**, *175*, 106037.

[41] F. Carles, S. Bourg, C. Meyer, P. Bonnet, *Molecules* **2018**, *23*, 908.

[42] M. Patil, N. Pabla, Z. Dong, *Cell. Mol. Life Sci.* **2013**, *70*, 4009–4021.

[43] K. M. Pflug, R. Sitcheran, *Int. J. Mol. Sci.* **2020**, *21*, 8470.

[44] G. M. Bokoch, *Annu. Rev. Biochem.* **2003**, *72*, 743–781.

[45] X. Zhang, M. Song, J. K. Kundu, M.-H. Lee, Z.-Z. Liu, *J. Cancer Prev.* **2018**, *23*, 109.

[46] R. Roskoski Jr, *Pharmacol. Res.* **2016**, *103*, 26–48.

[47] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[48] S. Boothroyd, P. K. Behara, O. C. Madin, D. F. Hahn, H. Jang, V. Gapsys, J. R. Wagner, J. T. Horton, D. L. Dotson, M. W. Thompson, J. Maat, T. Gokey, L.-P. Wang, D. J. Cole, M. K. Gilson, J. D. Chodera, C. I. Bayly, M. R. Shirts, D. L. Mobley, *J. Chem. Theory Comput.* **2023**, *19*, 3251–3275.

[49] E. P. Barros, B. Ries, C. Champion, S. R. Rieder, S. Riniker, *J. Chem. Inf. Model.* **2023**, *63*, 1794–1805.

[50] C. H. Bennett, *J. Comput. Phys.* **1976**, *22*, 245–268.

[51] M. R. Shirts, J. D. Chodera, *J. Chem. Phys.* **2008**, *129*, 124105.

[52] V. Gapsys, D. Seeliger, B. L. de Groot, *J. Chem. Theory Comput.* **2012**, *8*, 2373–2382.

[53] C. D. Christ, W. F. van Gunsteren, *J. Chem. Theory Comput.* **2009**, *5*, 276–286.

[54] S. Riniker, C. D. Christ, N. Hansen, A. E. Mark, P. C. Nair, W. F. van Gunsteren, *J. Chem. Phys.* **2011**, *135*, 07B604.

[55] B. Ries, S. Rieder, C. Rhiner, P. H. Hünenberger, S. Riniker, *J. Comput. Aided Mol. Des.* **2022**, *36*, 175–192.

[56] A. S. J. S. Mey, B. K. Allen, H. E. B. Macdonald, J. D. Chodera, D. F. Hahn, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, A. Rizzi, J. Scheen, M. R. Shirts, G. Tresadern, H. Xu, *LiveCoMS* **2020**, *2*, 18378–18378.

[57] N. Blaquiere, G. M. Castanedo, J. D. Burch, L. M. Berezhkovskiy, H. Brightbill, S. Brown, C. Chan, P.-C. Chiang, J. J. Crawford, T. Dong, P. Fan, J. Feng, N. Ghilardi, R. Godemann, E. Gogol, A. Grabbe, A. J. Hole, B. Hu, S. G. Hymowitz, M. H. Alaoui Ismaili, H. Le, P. Lee, W. Lee, X. Lin, N. Liu, P. A. McEwan, B. McKenzie, H. L. Silvestre, E. Suto, S. Sujatha-Bhaskar, G. Wu, L. C. Wu, Y. Zhang, Z. Zhong, S. T. Staben, *J. Med. Chem.* **2018**, *61*, 6801–6813.

[58] H. D. Brightbill, E. Suto, N. Blaquiere, N. Ramamoorthi, S. Sujatha-Bhaskar, E. B. Gogol, G. M. Castanedo, B. T. Jackson, Y. C. Kwon, S. Haller, J. Lesch, K. Bents, C. Everett, P. B. Kohli, S. Linge, L. Christian, K. Barrett, A. Jaochico, L. M. Berezhkovskiy, P. W. Fan, Z. Modrusan, K. Veliz, M. J. Townsend, J. DeVoss, A. R. Johnson, R. Godemann, W. P. Lee, C. D. Austin, B. S. McKenzie, J. A. Hackney, J. J. Crawford, S. T. Staben, M. H. Alaoui Ismaili, L. C. Wu, N. Ghilardi, *Nat. Commun.* **2018**, *9*, 179.

[59] J. Rudolph, L. J. Murray, C. O. Ndubaku, T. O'Brien, E. Blackwood, W. Wang, I. Aliagas, L. Gazzard, J. J. Crawford, J. Drobnick, W. Lee, X. Zhao, K. P. Hoeflich, D. A. Favor, P. Dong, H. Zhang, C. E. Heise, A. Oh, C. C. Ong, H. La, P. Chakravarty, C. Chan, D. Jakubiak, J. Epler, S. Ramaswamy, R. Vega, G. Cain, D. Diaz, Y. Zhong, *J. Med. Chem.* **2016**, *59*, 5520–5541.

[60] C. C. Ong, S. Gierke, C. Pitt, M. Sagolla, C. K. Cheng, W. Zhou, A. M. Jubb, L. Strickland, M. Schmidt, S. G. Duron, D. A. Campbell, W. Zheng, S. Dehdashti, M. Shen, N. Yang, M. L. Behnke, W. Huang, J. C. McKew, J. Chernoff, W. F. Forrest, P. M. Haverty, S.-F. Chin, E. A. Rakha, A. R. Green, I. O. Ellis, C. Caldas, T. O'Brien, L. S. Friedman, H. Koeppen, J. Rudolph, K. P. Hoeflich, *Breast Cancer Res.* **2015**, *17*, 1–12.

[61] X. Wang, W. Blackaby, V. Allen, G. K. Y. Chan, J. H. Chang, P.-C. Chiang, C. Diène, J. Drummond, S. Do, E. Fan, E. B. Harstad, A. Hodges, H. Hu, W. Jia, W. Kofie, A. Kolesnikov, J. P. Lyssikatos, J. Ly, M. Matteucci, J. G. Moffat, V. Munugalavadla, J. Murray, D. Nash, C. L. Noland, G. Del Rosario, L. Ross, C. Rouse, A. Sharpe, D. Slaga, M. Sun, V. Tsui, H. Wallweber, S.-F. Yu, A. J. Ebens, *J. Med. Chem.* **2019**, *62*, 2140–2153.

[62] 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*, 123–128.

[63] D. A. Case, R. C. Walker, C. Cheatham, Thomas E. Simmerling, A. Roitberg, K. M. Merz, R. Luo, T. Darden, J. Wang, R. E. Duke, D. R. Roe, S. LeGrand, J. Swails, D. Cerutti, G. Monard, C. Sagui, J. Kaus, R. Betz, B. Madej, C. Lin, D. Mermelstein, P. Li, A. Onufriev, S. Izadi, R. M. Wolf, X. Wu, A. W. Götz, H. Gohlke, N. Homeyer, W. M. Botello-Smith, L. Xiao, T. Luchko, T. Giese, T. Lee, H. T. Nguyen, H. Nguyen, P. Janowski, I. Omelyan, A. Kovelnko, P. A. Kollman, AMBER Reference Manual; University of California: San Francisco, **2016**.

[64] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, *Nucleic Acids Res.* **2003**, *31*, 3381–3385.

[65] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

[66] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[67] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2000**, *21*, 132–146.

[68] J. Wagner, M. Thompson, D. Dotson, hyejang, S. Boothroyd, J. Rodríguez-Guerra, openforcefield/openff-forcefields: Version 2.0.0 "Sage", version 2.0.0, **2021**.

[69] W. Jiang, C. Chipot, B. Roux, *J. Chem. Inf. Model.* **2019**, *59*, 3794–3802.

[70] J. Z. Vilseck, L. F. Cervantes, R. L. Hayes, C. L. Brooks III, *J. Chem. Inf. Model.* **2022**, *62*, 1479–1488.

[71] G. Landrum, P. Tosco, B. Kelley, S. Riniker, Ric, gedeck, R. Vianello, N. Schneider, A. Dalke, D. N, K. Eisuke, B. Cole, S. Turk, M. Swain, S. Alexander, D. Cosgrove, A. Vaucher, M. Wójcikowski, G. Jones, D. Probst, G. Godin, V. F. Scalfani, A. Pahl, B. Francois, JLVarjo, strets123, JP, DoliathGavin, G. Sforna, J. H. Jensen, rdkit/rdkit: 2021_03_2 (Q1 2021) Release, version Release_2021_03_2, Accessed: 2022/05/05, **2021**.

[72] M. T. Lehner, B. Ries, S. R. Rieder, S. Riniker, **2021**, Accessed: 2022/05/05, DOI 10.5281/zenodo.4621710.

[73] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, W. F. van Gunsteren, *Comp. Phys. Comm.* **2012**, *183*, 890–903.

[74] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.

[75] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[76] B. Ries, S. R. Rieder, C. Champion, E. P. Barros, S. Riniker, rinikerlab/reeds: An Automatized RE-EDS Sampling Procedure (v1.0), Accessed: 2022/05/05, **2021**.

[77] A. P. Bhati, S. Wan, Y. Hu, B. Sherborne, P. V. Coveney, *J. Chem. Theory Comput.* **2018**, *14*, 2867–2880.

[78] B. Knapp, L. Ospina, C. M. Deane, *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.

[79] I. V. Khavrutskii, A. Wallqvist, *J. Chem. Theory Comput.* **2011**, *7*, 3001–3011.

[80] A. Y. Shih, M. Hack, T. Mirzadegan, *J. Chem. Inf. Model.* **2020**, *60*, 5287–5289.

[81] 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*, 42–54.

[82] V. Gapsys, D. F. Hahn, G. Tresadern, D. L. Mobley, M. Rampp, B. L. de Groot, *J. Chem. Inf. Model.* **2022**, *62*, 1172–1177.

[83] H. M. Baumann, V. Gapsys, B. L. de Groot, D. L. Mobley, *J. Phys. Chem. B* **2021**, *125*, 4241–4261.

[84] M. Kuhn, S. Firth-Clark, P. Tosco, A. S. J. S. Mey, M. Mackey, J. Michel, *J. Chem. Inf. Model.* **2020**, *60*, 3120–3130.

[85] M. Pitman, D. F. Hahn, G. Tresadern, D. L. Mobley, *J. Chem. Inf. Model.* **2023**, *63*, 1776–1793.

[86] F. Fratev, S. Sirimulla, *Sci. Rep.* **2019**, *9*, 16829.

[87] W. F. van Gunsteren, gromos.net, http://www.gromos.net/, Accessed: 2022/01/06, **2021**.