BAT2: An open-source tool for flexible, automated and low cost absolute binding free energy calculations

Germano Heinzelmann,*,† David J. Huggins,‡,¶ and Michael K. Gilson§

†*Departamento de Fisica, Universidade Federal de Santa Catarina, Florianopolis, Brasil*

‡*Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, New York 10065, United States*

¶*Sanders Tri-Institutional Therapeutics Discovery Institute, 1230 York Avenue, Box 122, New York, New York 10065, United States*

§*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, USA*

E-mail: germano.heinzelmann@ufsc.br

Abstract

Absolute binding free energy (ABFE) calculations with all-atom molecular dynamics (MD) have the potential to greatly reduce costs in the first stages of drug discovery. Here we introduce BAT2, the new version of the Binding Affinity Tool (BAT.py), designed to combine full automation of ABFE calculations with high-performance MD simulations, making it a potential tool for virtual screening. We describe and test several changes and new features that were incorporated into the code, such as relative restraints between the protein and the ligand instead of using fixed dummy atoms, support for the OpenMM simulation engine, a merged approach to the application/release of restraints, support for cobinders and proteins with multiple chains, and many others.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

We also reduced the simulation times for each ABFE calculation, assessing the effect on the expected robustness and accuracy of the calculations.

1 Introduction

Identifying molecules that bind to a therapeutic target is one of the most important steps in the early stages of drug discovery. Experimental high-throughput screening (HTS) using robots can be carried out on hundreds of thousands of ligands in a single day,[1,2] but the need for large compound libraries and sophisticated machinery makes this approach costly and resource intensive. In the past years, virtual screening (VS) using computational scoring functions have emerged as a viable alternative, with many studies identifying potent molecules that were subsequently validated by experiments.[3,4] Ligand-based VS methods[5,6] rely on available experimental data, and so do advanced structure-based docking[7,8] and ranking methods that use artificial intelligence (AI).[9,10] Consequently, VS on novel targets without existing chemical matter is expected to produce less accurate results.

Another important class of computational tools that can estimate protein-ligand affinities are the physics-based methods, which perform binding free energy calculations using an ensemble of states generated by all-atom molecular dynamics (MD) simulations.[3,11–15] The interactions between the atoms in the MD simulation are described by an atom-based forcefield, so in principle there is a high level of transferability between different biological systems. These free energy methods are commonly divided in two classes, relative (RBFE)[16–20] and absolute (ABFE)[11,13,21–29] binding free energy calculations. RBFE methods, as the name suggests, compute the relative difference in binding free energy between two similar compounds, by alchemically transforming one into the other in the receptor binding site, and doing the opposite transformation in bulk solvent. Even though widely used in the lead optimization stages, RBFE calculations can become challenging when comparing molecules that have little similarity,[30] making it unsuitable for virtual screening on a diverse set of

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

ligands.

Conversely, ABFE calculations estimate the standard binding free energy of a single ligand by calculating the free energy difference of transferring it from the protein binding site to bulk solvent at 1 M concentration. Since each ligand is treated individually and their binding free energies can be directly compared, ABFE can rank ligands regardless of similarity and thus can be applied to virtual screening.[3,13,15] A recent study by Feng et al. has shown that, despite not yet having the same accuracy as RBFE methods, ABFE calculations can effectively distinguish between binders and non-binders to a given receptor and further enrich a set of top-scoring compounds obtained by docking.[15]

For many years after the implementation of ABFE calculations, there were a few practical challenges that prevented its widespread adoption. In recent years, these challenges have been partially or completely addressed:

- Firstly, ABFE's high computational cost when compared to scoring methods or even RBFE. In recent times, the widespread use of Graphics Processing Units (GPUs)[31–34] has greatly increased the speed and scalability of MD simulations, making ABFE calculations considerably quicker and cheaper.

- Secondly, the human factor of building the systems and setting up the calculations manually. Recent tools such as BAT.py,[13] the binding free energy estimator (BFEE),[35] the CHARMM-GUI[36] server and Schrodinger's ABFEP[11] have automated the steps of preparing, running and analyzing the necessary simulations, thus reducing or even eliminating the need of human intervention.

- Thirdly, the need for the correct conformation of the ligand in the binding site in order to produce meaningful results. This challenge can be addressed by considering multiple different poses independently, which is incorporated in the BAT.py ABFE workflow (Figure 1). The pose with the lowest binding free energy would usually dominate the sum from Eq. 1, and thus be the one observed experimentally.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

- Lastly, possible inaccuracies arising from conformational changes in the protein between its holo and apo states. The latter problem is discussed in detail in our previous work with APR calculations on bromodomains, in which such a transition is identified and its free energy contribution is rigorously computed.[21] The BAT.py software also provides ways to address the issue of protein regions that have increased flexibility.

Since the release of the BAT.py 1.0 software (or BAT1) in 2020, and its associated article in 2021,[13] there were important changes made to the code in order to make the calculations faster, easier, more rigorous and applicable to a wider variety of systems. Thus, in the present manuscript we introduce the BAT 2.x software, or BAT2, currently in its 2.3 version and available at https://github.com/GHeinzelmann/BAT.py. BAT2 provides several improvements over its predecessor, with the most important ones being:

- Support for the open-source OpenMM simulation software[37–39] with OpenMMtools.[40]
- Relative restraints between the protein and the ligand, with a simpler procedure to add new receptors to the BAT2 workflow. Previously, both the protein and the ligand anchor atoms were restrained relative to fixed dummy atoms, which was harder to set up for larger proteins.
- The option of merging all attachments and releasing of restraints into two sets of simulation windows, which requires fewer simulations and makes the calculations cheaper.
- Applicability to proteins with multiple chains and in the presence of cobinders.
- Support for the TIP3PF[41] and OPC[42] water models, in addition to the ones already supported in BAT1.
- Automatic determination of the number of ions in all boxes for a chosen ion concentration.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

- Use of the lovoalign software[43] for protein structure alignment, replacing MUSTANG.[44] The change was made because lovoalign can superimpose protein structures that contain multiple polypeptide chains.

- Choice of fixed solvation buffers in the three axes, or a fixed number of water molecules as with BAT1.

- Freedom to include ligands with hydrogens already added, as well as using pre-generated ligand parameters.

- Only two stages, equilibration and free energy calculations. This makes the calculations cheaper, since the preparation step is not needed in the BAT2 workflow (Section 3.1).

In the next sections we explain the theory and methods behind the modifications made, test them in terms of consistency with previous results, and use them on two sample systems. To help determine the feasibility of ABFE calculations on large libraries of ligands, we also explore the possibility of performing them with a small fraction of the simulation time used previously, assessing the effects of this reduction on their expected accuracy.

2 Theory

If we take into account all possible stable and non-overlaping bound states of a given ligand in a given receptor's binding site, N_{poses}, the standard (or absolute) binding free energy of this molecule to a given receptor, ΔG°_{bind}, can be determined using the equation:[13]

$$\Delta G^{\circ}_{bind} = -RT \ln \sum_{i}^{N_{pose}} e^{-\beta \Delta G^{\circ}_{i}}, \tag{1}$$

where i indexes ligand poses, ΔG°_{bind} is the binding free energy computed for pose i, R is the gas constant, T is absolute temperature, and $\beta^{-1} = RT$.[45] This expression assumes that the poses do not interconvert during their individual binding free energy calculations. Due to the

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

exponential character of the term inside the sum, the lowest value of ΔG_i° will dominate the value of ΔG_{bind}°, so we can consider these two quantities to be equivalent in most cases. The dissociation constant (K_d) between the ligand (L) and the protein (P) is related to ΔG_{bind}° by the following expression:[26]

$$K_d = \frac{[L][P]}{[LP]C^\circ} = e^{\Delta G_{bind}^\circ / RT}, \tag{2}$$

where C° is the standard concentration of 1 M, [L], [P] and [LP] are the equilibrium concentrations of the respective species.

As commonly done for ABFE calculations, we will calculate the free energy of transferring the ligand from the binding site to bulk solvent (ΔG_{trans}) in the presence of artificial restraints, in order to accelerate the convergence of the calculations. The final value of ΔG_{bind}° will then include, in addition to ΔG_{trans}, the free energies of attaching and releasing the chosen restraints:

$$-\Delta G_{bind}^\circ = \Delta G_{p,att} + \Delta G_{l,att} + \Delta G_{trans} + \Delta G_{l,rel} + \Delta G_{p,rel} \tag{3}$$

The values of $\Delta G_{p,att}$ and $\Delta G_{p,rel}$ represent the free energies of attaching and releasing restraints to the protein, respectively, and $\Delta G_{l,att}$ and $\Delta G_{l,rel}$ the same for the ligand. In Table 1 we list all the free energy components calculated by the BAT2 program, each identified by a letter. The way they are obtained, and how they correspond to each term from Equation 3, will be explained in the next sections.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Table 1: Letter codes for the contributions to the binding free energy ΔG°_{bind}. The second column shows the merged **m** and **n** components for the attachment and release of restraints, as well as the electrostatic ΔG_{elec} (**e**) and Lennard-Jones (LJ) ΔG_{LJ} (**v**) components of the SDR procedure.

Description	Letter		System	Method	Term
Attachment of protein conformational restraints	**m**	**a**	Complex	MBAR	$\Delta G_{p,att}$
Attachment of ligand conformational restraints		**l**	Complex	MBAR	$\Delta G_{l,conf,att}$
Attachment of ligand TR restraints		**t**	Complex	MBAR	$\Delta G_{l,TR,att}$
Ligand charge decoupling in site	**e**	**e**	Complex[†]	MBAR/TI-GQ	$\Delta G_{elec,bound}$
Ligand charge recoupling in bulk		**f**	Bulk ligand[†]	MBAR/TI-GQ	$-\Delta G_{elec,unbound}$
Ligand LJ decoupling in site	**v**	**v**	Complex[†]	MBAR/TI-GQ	$\Delta G_{LJ,bound}$
Ligand LJ recoupling in bulk		**w**	Bulk ligand[†]	MBAR/TI-GQ	$-\Delta G_{LJ,unbound}$
Release of ligand TR restraints	**n**	**b**	Bulk ligand	Analytical	$\Delta G_{l,TR,rel}$
Release of ligand conformational restraints		**c**	Bulk ligand[†]	MBAR	$\Delta G_{l,conf,rel}$
Release of protein conformational restraints		**r**	Apo protein[†]	MBAR	$\Delta G_{p,rel}$

[†] For the SDR **e** and **v** components, and the merged **n** component, the complex (or apo protein) and the bulk ligand are placed far from each other in the same box.

3 Methods

3.1 BAT2 Workflow

The BAT2 automated workflow (Figure 1) encompasses all the steps needed to perform a full protein-ligand ABFE calculation, using either the double decoupling method (DDM) [26] or the simultaneous decoupling and recoupling (SDR) method. [13,28] BAT2 requires a few third-party

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

programs such as OpenBabel,[46] Visual Molecular Dynamics (VMD)[47] and lovoalign, which are listed in the software's main page and will be referred to throughout the manuscript. The simulations can be performed using either the the *pmemd.cuda* software from AMBER,[48] version 20 or later, or the OpenMM simulation engine with OpenMMtools, versions 7.7.0 or later for the former and 0.21.3 or later for the latter.

The main inputs to BAT2 are a protein-ligand pair, and an input file that has all the parameters needed for the calculation (Table 2). These include simulation times for each step, number of simulation windows for the free energy calculations, solvation options, as well as specific variables that have to be set up for a new receptor system. Also needed are a protein reference structure file and, depending on the system, molecule force-field parameters that were not generated automatically, as in the case of cobinders. The BAT2 User Guide provides further instructions on how to set up all the needed files, and can be found in the BAT2 distribution.

Table 2: Input files needed for the BAT.py software.

File description	File format	Function	Comments
Docked receptor	PDB file	Provides structure of protein/cobinders	Can be replaced by a protein-ligand complex
Docked poses	PDB files	Docked ligand structures	Not needed if using a protein-ligand complex
Reference file	PDB file	Contains the protein in the reference rotation	Protein structure alignment performed with lovolign
BAT2 input file	Text file	Provides the BAT2 calculation parameters	Includes the variables needed for a given protein system
Additional parameters	.mol2 and .frcmod	Simulation parameters for cobinders and ligand	The ligand parameters can also be generated automatically

As outlined in Figure 1, the workflow starts from either a receptor accompanied by a set of docked poses, or the structure of a docked complex such as a protein-ligand cocrystal structure. First, hydrogens are added to the ligand, and its force-field parameters to be used

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

in the MD simulations are determined. The user can also maintain the ligand protonation and/or its parameters, in case they were already available before running BAT2. Next comes the alignment of the protein-ligand complex to a reference structure provided by the user, and determining the ligand anchor atoms that will be used in restraining the ligand relative to the protein. The complex is then solvated in water with a desired ion concentration, restraints are applied to the ligand relative to the protein, and equilibration simulations are performed in which these restraints are gradually released.

After the equilibration simulations are finished, the next step is the free energy step, which starts by checking if the ligand is still in the proposed binding site after equilibration. If not, the initial docked pose is considered unstable and no free energy calculations are performed. If it is, the system is rebuilt in the equilibrated configuration but with a new set of ligand anchors and reference values for the restraints, and all the free energy simulations are carried out. Once the simulations are concluded, the analysis step will calculate the binding free energy of the ligand for that particular binding pose. The main BAT2 output is a file that contains the calculated free energy value for each component (Table 1), the sum of the components that make up binding free energy (Eq. 3), and the total simulation time needed for the calculation. BAT2 also outputs the equilibrated complex used as the starting point for the free energy step, which provides the reference coordinates used for the applied restraints. The binding free energies across different binding modes can then be evaluated to determine the correct one, and also compared to the values obtained for other molecules.

The main difference of the present workflow, when compared to BAT1, is the elimination of the preparation stage. This stage would carry out a procedure similar to steered molecular dynamics (SMD), generating initial states along the pulling coordinate used in the Attach-Pull-Release (APR)[21,24] binding free energy method. This is not needed for BAT2, since the latter only applies the alchemical double decoupling (DDM)[26] and the simultaneous decoupling and recoupling (SDR) methods[13,28] for the ABFE calculations, the latter being suitable for ligands with net charge.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

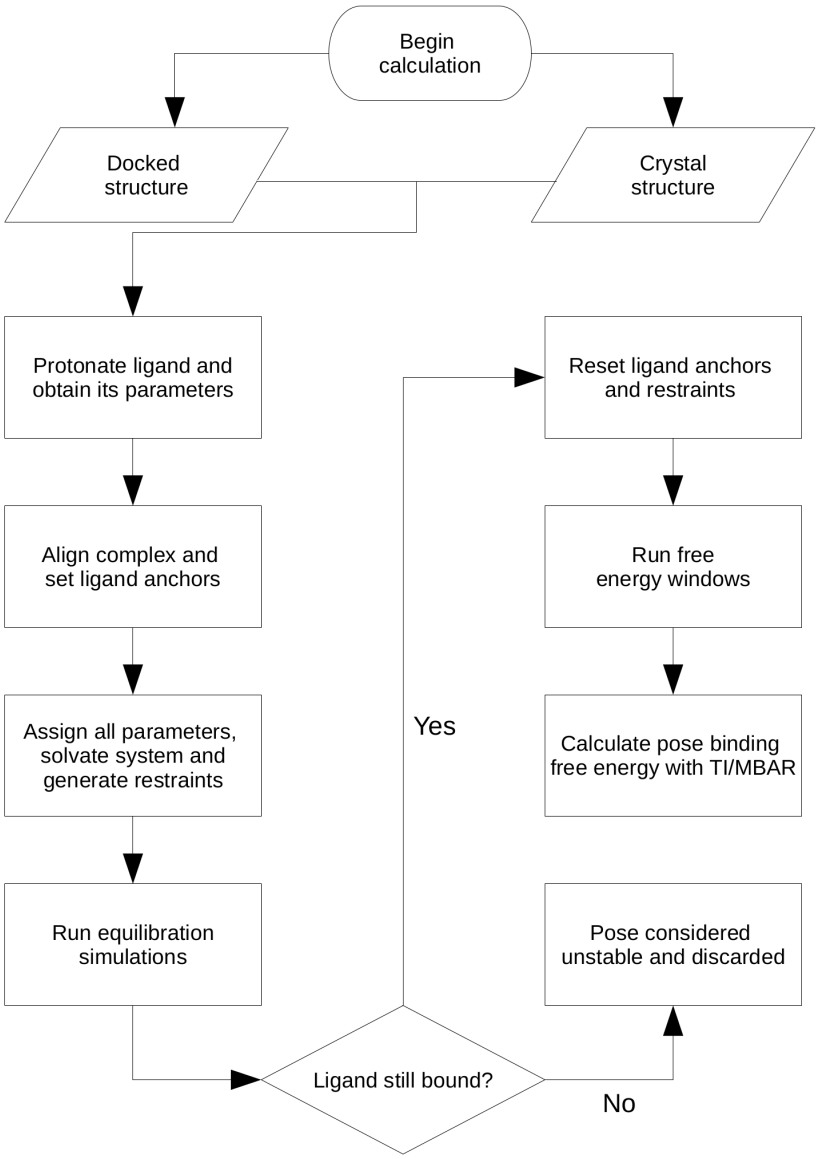

Figure 1: New workflow of the BAT.py software. See text for details.

3.2 System setup

Here there are several differences relative to the BAT1, so in the sections below we explain how BAT2 will use its input to set up the necessary systems and parameters for the calculations.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

3.2.1 Anchor atom selection

When applying BAT2 to a new protein system, three protein anchor atoms (P1, P2 and P3) have to be selected by the user, following a few rules to avoid gimbal-locking and effects on the internal degrees of freedom of the receptor. These rules include choosing backbone anchor atoms in regions that are typically rigid, such as alpha-helices, as well as avoiding short distances and angles between anchors that approximate 0 or 180 degrees.

In order to check if the ligand is inside a predefined protein binding site, which happens in the beginning of the equilibration and free energy stages, BAT2 uses a spherical search zone relative to the position of P1 using extrinsic cartesian coordinates (Fig. 2). For this reason, the protein needs to be in a predefined orientation, which is done by aligning the former to a reference structure using the program lovoalign. The reference structure, the coordinates of the center of the search zone and the search radius are all provided by the user. The BAT2 User Guide has detailed instructions on how to add a new protein system to the BAT2 workflow, with the help of visualization tools such as VMD or Chimera.[49] Once that is concluded, any ligand that binds to the same binding site can be evaluated in a fully automated way without any human intervention.

The ligand anchors, called L1, L2 and L3, will be chosen automatically with the protein already in the reference orientation. The L1 anchor will be the ligand atom closest to the center of the selected search zone; if no ligand atoms are encountered inside the search zone at the start of the free energy stage, this means that the ligand has left the binding site during equilibration. In this case, the starting docked pose is considered to be unstable and no free energy calculation is performed (Figure 1). If L1 is found, the L2 anchor will be selected as the atom in which the P1-L1-L2 angle is closest to 90 degrees, and the L1-L2 distance is within a specified range defined in the BAT2 input file. The choice of L3 follows the same procedure, but now using the L2-L3 distance and the L1-L2-L3 angle.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

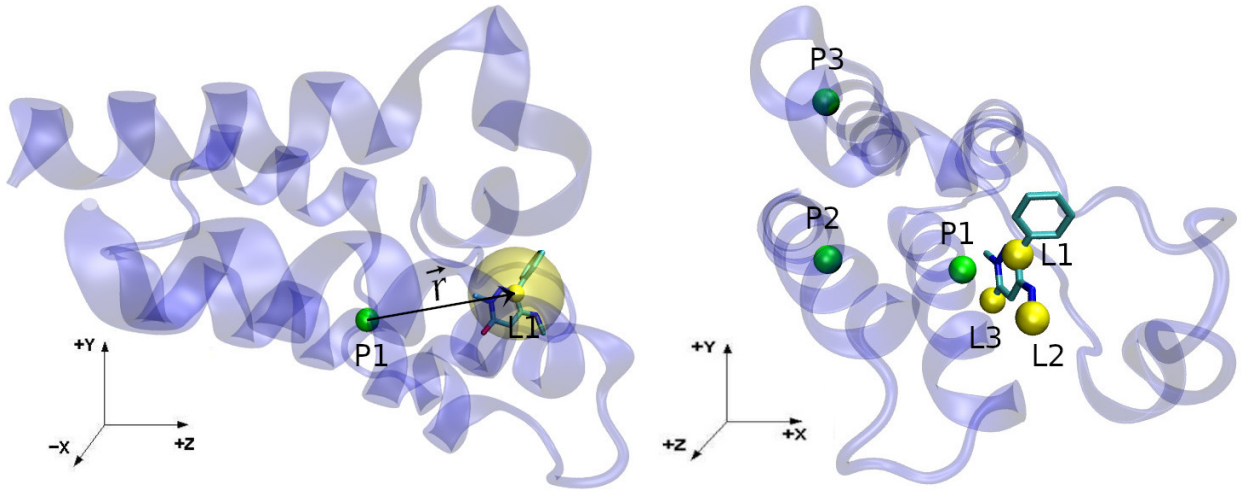

Figure 2: (*left*) The P1 and L1 anchors, as the green and yellow solid spheres, and the $\vec{r}$ vector connecting the two. The transparent yellow region shows the L1 anchor spherical search region. (*right*) All the protein and ligand anchors, as in the right of Figure 3, but from a different perspective.

3.2.2 Force field parameters, solvation and ionization

If no information is provided by the user regarding the ligand protonation state and/or its simulation parameters, BAT2 will use the program Openbabel to add all hydrogens to the ligand molecule and estimate its net charge. The AM1-BCC charge model[50] is then used on the protonated ligand to determine its partial charges, and versions 1 or 2 of the General AMBER Force Field (GAFF)[51,52] for the LJ and bonded interactions. The user can also start with an already protonated ligand, choosing its net charge accordingly in the BAT2 input file, or generate the necessary ligand parameters separately and include them in the BAT2 workflow.

Regarding the protein, the protonation states of its titratable groups are predetermined from the associated residue templates. Several AMBER protein force-fields such as *ff14SB*[53] are available for use with BAT2, and can be selected by the user in the input file. The same goes for water models, with the ones currently supported by BAT2 being TIP3P,[54] TIP3PF,[41] TIP4PEw,[55] SPC/E[56] and OPC,[42] with the associated Li/Merz cation and anion parameters for each.[57–60] For the TIPEP, TIP4PEw and SPC/E water models, monovalent

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

ions use the Joung and Cheatham ion parameters[61] designed for that specific model. If there are other molecules or ion types in the system in addition to the protein, ligand, water and solvated ions, such as cobinders, the user needs to obtain the parameters for them separately before adding them to BAT2. In particular, the user should provide .mol2 and .frcmod files for each molecule, which can be generated using Antechamber by following the tutorial at https://ambermd.org/tutorials/basic/tutorial4b/index.php.

The Ambertools tleap software[48] is used to solvate the system and add the ions to a concentration chosen by the user, or instead just add the counterions needed for neutralization. The user also specifies either the solvation buffers on the three cartesian axes, or the buffers in two of the axes and the total number of water molecules that will be added to the system. These definitions will be used for all simulation boxes built for equilibration and free energy calculations, except the ones that have only the ligand in them. In the latter case, there will be specific solvation buffers for the small ligand box, with the ion concentration the same as the one chosen for the others.

3.3 Application and release of restraints

We can separate the restraints applied to the protein and the ligand in three types: center of mass (COM) restraints, conformational restraints, and rigid-body translational/rotational (TR) restraints. The latter restrains both the position and orientation of the ligand relative to the protein using three anchor atoms in each molecule, with a total of six restrained degrees of freedom (Fig. 3). These are commonly called the Boresch[23] restraints, which are designed to keep the ligand in the binding site during the decoupling of the latter's interactions with its environment.

The COM restraints are used to maintain the location of the complex (or apo protein) inside the box, and also the position of the ligand in bulk solvent when the latter shares the simulation box with the first (left of Fig 3). These restraints are applied to either all backbone atoms of the protein or all non-hydrogen atoms of the bulk ligand, keeping the

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

center of mass of the chosen atoms fixed throughout the simulation, but leaving each molecule free to rotate around its center of mass. No free energy calculations are performed for the COM restraints, since they only maintain the chosen reference frame and do not interfere with the internal degrees of freedom of a given species.

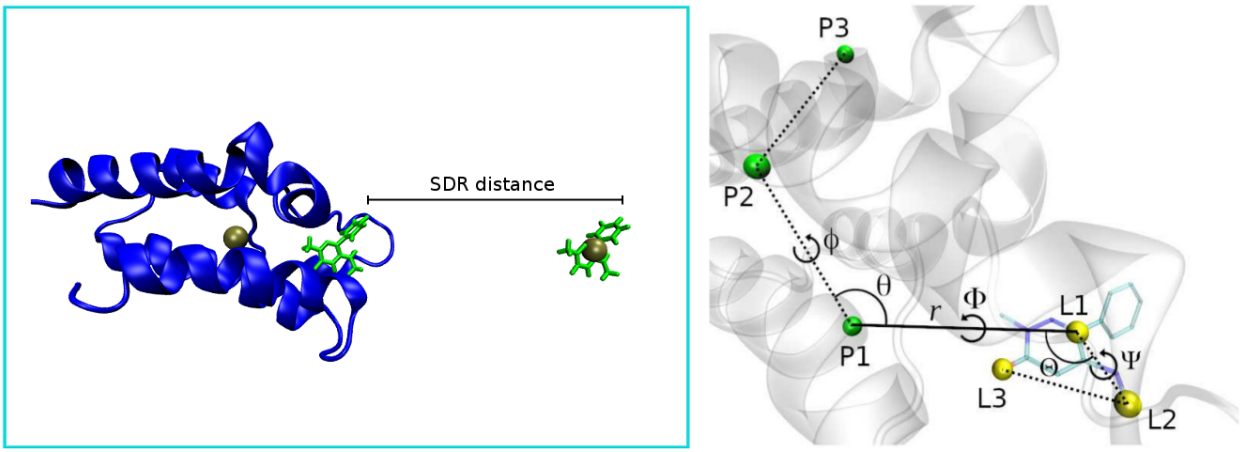

Figure 3: (*left*) Depiction of the SDR box, with the protein in blue, the ligand in green, and the centers of mass of the protein backbone and bulk ligand as the golden spheres. The SDR distance is defined between the two ligands along the z axis, and is chosen in the BAT2 input file. (*right*) The TR restraints on the ligand relative to the protein, showing the protein anchors in green, the ligand anchors in yellow, and the six restrained degrees of freedom: the distance r, the θ and Θ angles, and the ϕ, Φ and Ψ dihedrals.

The conformational restraints on the protein are the same as previously described,[13] so here we will go over them briefly. They are applied to one or more sections of the protein backbone, more specifically to the ϕ and ψ backbone dihedrals in this range, which may aid in convergence depending on the system. The free energies of attaching and releasing the protein backbone restraints correspond to the $\Delta G_{p,att}$ and $\Delta G_{p,rel}$ terms, respectively, and are calculated using a set of simulation windows with intermediate values of the applied spring constants. The first term is computed from simulations that have the ligand in the binding site, and the second has the protein in the apo state. The final attachment/releasing free energy differences are obtained from the respective simulations using the Multistate Bennett Acceptance Ratio (MBAR) method[62] (Table 1). In the case of $\Delta G_{p,rel}$, this contribution is the same across poses if the protein restraint reference state is the same for all of them, and

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

thus only has to be computed once.

The restraints applied to the ligand include both conformational and TR, so the $\Delta G_{l,att}$ and $\Delta G_{l,rel}$ terms from Eq. 3 will have two components each:

$$\Delta G_{l,att} = \Delta G_{l,conf,att} + \Delta G_{l,TR,att} \tag{4}$$

$$\Delta G_{l,rel} = \Delta G_{l,conf,rel} + \Delta G_{l,TR,rel} \tag{5}$$

The $\Delta G_{l,conf,att}$ and $\Delta G_{l,conf,rel}$ terms correspond to the free energy of attaching and releasing the ligand conformational restraints. The same way as in our previous work,[13] all dihedrals of the ligand that do not involve hydrogens are restrained with a spring constant specified by the user. The attachment of these restraints happens with the ligand in the binding site, and their release can happen either with the ligand in a separate box (DDM), or in the same box as the protein but at a distance in which they do not interact (SDR method). As with the protein conformational restraints, we use a number of simulation windows and obtain the final free energy difference using MBAR.

The ligand TR restraints on the BAT2 program are not relative to three fixed dummy atoms, as with BAT1, but instead they are defined relative to the protein. They involve one distance, two angles and three dihedrals between three protein anchors and the three ligand anchors as shown in Figure 3. The attachment of these restraints takes place when the ligand is in the binding site, using a series of simulation windows and the MBAR estimator. The release of the TR restraints to the standard concentration of 1 M is done analytically using Equation 6:

$$\begin{aligned}\Delta G_{l,TR,rel} = k_B T ln\left(\frac{C^\circ}{8\pi^2}\right) + k_B T ln \int_0^\infty \int_0^\pi \int_0^{2\pi} exp[-\beta(u_r + u_\theta + u_\phi)] r^2 sin\theta d\theta d\phi dr + \\ k_B T ln \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} exp[-\beta(u_\Theta + u_\Phi + u_\Psi)] sin\Theta d\Theta d\Phi d\Psi\end{aligned} \tag{6}$$

, where u is the restraining potential applied to a given coordinate from Fig. 3 (right) during

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

the decoupling calculations.

3.3.1 Merged restraints

In BAT1, the attachment of the restraints in the bound state happens in sequence, so the ligand conformational free energy calculation has the protein conformational restraints already in place, and the ligand TR restraints free energy calculation has both the protein and ligand conformational restraints fully attached.[13] New to BAT2 is the possibility of attaching and/or releasing all restraints using a single set of simulation windows with MBAR, which could reduce the computational cost of the calculations. We call these new components the merged restraint components, and they are identified by the **m** and **n** letters (Table 1).

The **m** component is the free energy of attaching all restraints to both the protein and the ligand in the bound state simultaneously, corresponding to the free energy term:

$$\Delta G_{all,att} = \Delta G_{p,att} + \Delta G_{l,att} \tag{7}$$

, with the $\Delta G_{l,att}$ free energy contributions shown in Eq. 4. The **m** simulation boxes are identical to the **a**, **l**, and **t** ones, with the protein-ligand complex fully interacting in solution.

The **n** component computes the restraint releasing free energies for both the protein and the ligand, with the corresponding free energy difference:

$$\Delta G_{all,rel} = \Delta G_{p,rel} + \Delta G_{l,rel} \tag{8}$$

, with the $\Delta G_{l,rel}$ term defined in Eq. 5. The two conformational contributions to $\Delta G_{all,rel}$, $\Delta G_{p,rel}$ for the protein and $\Delta G_{l,conf,rel}$ for the ligand, are calculated simultaneously using simulations that have the apo protein and the bulk solvent ligand in the same box but separated by COM restraints, as shown in Fig. 3 but without the bound ligand. As done with the separated restraints, the remaining $\Delta G_{l,TR,rel}$ term is calculated analytically using Eq. 6.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

3.4 Transfer of the ligand from binding site to bulk solvent

The transfer free energy in Eq. 3, ΔG_{trans}, is defined as the free energy of decoupling the ligand interactions with its environment when it is in the bound state, and recoupling them back when the ligand is unbound in bulk solvent:

$$\Delta G_{trans} = \Delta G_{dcpl,bound} + \Delta G_{rcpl,unbound}, \tag{9}$$

In BAT2 there are two ways of performing the transformation of a protein-bound ligand to a bulk ligand: the double decoupling (DDM) and the simultaneous decoupling and recoupling (SDR) methods.[13]

In the case of DDM, four free energy calculations are performed, one for each of the following terms on the right hand side of the expressions below:

$$\Delta G_{dcpl,bound} = \Delta G_{elec,bound} + \Delta G_{LJ,bound}, \tag{10}$$

$$\Delta G_{rcpl,unbound} = -\Delta G_{elec,unbound} - \Delta G_{LJ,unbound}, \tag{11}$$

$\Delta G_{elec,bound}$ is the free energy of decoupling all electrostatic interactions of the ligand with its environment in the bound state, and $\Delta G_{elec,unbound}$ the same for the ligand in bulk solvent. The $\Delta G_{LJ,bound}$ and $\Delta G_{LJ,unbound}$ terms follow the same definitions, but now for the Lennard-Jones interactions of a molecule that already has its charged interactions with the environment fully decoupled (also called a "neutral" ligand). Following the letter identification from Table 1, the **e** and **v** DDM components will be performed on the solvated protein-ligand complex, and the **f** and **w** components in a smaller simulation box containing only the solvated ligand.

For the SDR method, only two free energy calculations are performed, each of them

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

combining two components of the double decoupling method:

$$\Delta G_{elec} = \Delta G_{elec,bound} - \Delta G_{elec,unbound}, \tag{12}$$

$$\Delta G_{LJ} = \Delta G_{LJ,bound} - \Delta G_{LJ,unbound}, \tag{13}$$

The SDR box has the protein-ligand complex and the bulk ligand in the same system (Fig. 3), kept separated at a non-interacting distance by COM restraints applied separately to the receptor and the free ligand. The decoupling of the bound ligand and the recoupling of the bulk ligand take place simultaneously in the same system, both for the electrostatic (Eq. 12) and the Lennard Jones component (Eq. 13). This approach keeps the net charge of the simulation box constant throughout the electrostatic leg of the calculation, avoiding numerical artifacts associated with charged ligands.[63–65] The letter identifiers for the two SDR calculations are also **e** and **v**, the former for electrostatic and the latter for the LJ windows (Table 1). In contrast to the DD method, none of the SDR windows sample the true end-points of the calculations, which have the ligand fully coupled to the binding site and fully decoupled from bulk, and vice-versa. Nonetheless, the same terms are being computed using the two approaches, which can be verified by comparing Equations 9 to 13.

BAT2 can use either MBAR or Thermodynamic Integration with Gaussian Quadrature (TI-GQ)[21,48] for the decoupling/recoupling calculations. OpenMM uses Hamiltonian Replica Exchange (HREX)[40] when performing free energy calculations with MBAR, both for the attachment/release of restraints and for the transfer free energies, a feature that is not included in the AMBER calculations. The TI-GQ method requires the computation of the ensemble average of the derivative of the system potential energy U relative to the decoupling reaction coordinate λ, on a number of pre-determined λ_i values, with the final value of the free energy difference given by:

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

$$\Delta G_{TI-GQ} = \int_0^1 \left\langle \frac{\partial U(\vec{X},\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda = \sum_{i=1}^{n} w_i \left\langle \frac{\partial U(\vec{X},\lambda)}{\partial \lambda} \right\rangle_{\lambda_i} , \tag{14}$$

, with $\vec{X}$ the generalized coordinates for the position of the system particles. The expression on the right shows the integration using Gaussian quadrature for n windows, with its associated λ_i values and w_i Gaussian weights. There is an unique set of values for λ_i and w_i for each n, so the user only has to choose the value of the latter in the BAT2 input file.

In contrast with AMBER, it is not straightforward to obtain $\partial U/\partial\lambda$ when using the OpenMM/OpenMMtools software. For this reason, we have developed a finite difference method to obtain this quantity when using the latter software, so that the TI-GQ method is available for both simulation packages. We make use of the following approximation, by considering $\langle \partial U/\partial\lambda \rangle$ to be constant along a small interval $\delta\lambda$ (default value is $\delta\lambda = 0.001$):

$$\int_{\lambda-\delta\lambda/2}^{\lambda+\delta\lambda/2} \left\langle \frac{\partial U(\vec{X},\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \sim \left\langle \frac{\partial U(\vec{X},\lambda)}{\partial \lambda} \right\rangle_\lambda \delta\lambda = \delta G_{BAR} \tag{15}$$

The value of δG_{BAR} above is obtained by using the OpenMM MBAR/HREX procedure on two decoupling windows located at $\lambda_i - \delta\lambda/2$ and $\lambda_i + \delta\lambda/2$, which is done for each point in which the derivative of the potential is to be calculated. The obtained derivatives are then plugged into the Gaussian Quadrature expression from Eq. 14 to obtain the desired value of ΔG_{TI-GQ}.

3.5 Equilibration stage

The equilibration simulations using AMBER start the same way as with the BAT1 program, with an initial minimization followed by 100 ps of heating from 10 K to the desired temperature using the Langevin thermostat.[66] Then, a series of 15 ps simulations are performed to bring the system to 1 atm pressure using the Monte Carlo barostat.[67] This procedure avoids possible AMBER crashes caused by excessive shrinking of the initial box. The OpenMM

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

simulations added to BAT2 also start by performing an initial energy minimization on the solvated complex, after which the atom velocities are set to a random distribution that reflects the chosen temperature of the system. The simulation box is then coupled to a Langevin thermostat and a Monte Carlo barostat, in order to maintain the chosen temperature and a pressure of 1 atm.

The subsequent runs using both programs are carried out with constant temperature and pressure, with the parameters for the thermostat and barostat chosen in the BAT2 input file. A series of simulations slowly release TR and conformational restraints applied to the ligand, so that the surrounding protein has time to relax around the docked molecule. The protein conformational restraints, if chosen, can also be present at this stage. After the ligand restraints are removed, a (usually) longer simulation is performed, in which the ligand might find a nearby free energy minimum or leave the initial binding site. The magnitude of the restraints, and the simulation times for their release and for the unrestrained simulations, are all chosen by the user.

3.6 Free energy calculations

The free energy calculations are performed after equilibration, if the ligand has not left the binding site during the unrestrained simulations. The complex will be realigned to the reference structure, the ligand anchors and restraints will be redefined in order to reflect the equilibrated conformation, and for each chosen component of Table 1 (except $\Delta G_{l,TR,rel}$) a series of simulation windows will be created. They can be related to the attachment/release of restraints, or the transfer of the ligand from the binding site to bulk solvent. For each window, an initial equilibration is performed, followed by a production simulation in which data is collected.

Whereas for AMBER the user specifies a number of equilibration and production simulation steps for each window, in the case of OpenMM the user will select a number of equilibration/production HREX iterations and the number of MD steps for each. Other

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

options for both programs can be selected in the BAT2 input file, such as the number of windows and the intermediate lambda values.

Once all simulations from the free energy step are concluded, BAT2 will compute the free energy contribution from each component (Table 1) used in the calculation, and the final binding free energy between the protein and the ligand for that particular binding mode. The uncertainties will be computed from block data analysis as done with the BAT1 version,[13] with the number of data blocks chosen by the user.

3.7 Protein-ligand test systems

In order to exemplify the use of the BAT2 program, we will apply it to two protein systems. The first is the second bromodomain of the Bromodomain-contaning protein 4, or BRD4(2), bound to a unique fragment that was used as a starting point for a series of new binders,[68] the same system used in our previous study on the BAT1 program.[13] As in the latter, we will perform calculations with the ligand from the 5uf0 crystal structure docked to the 5uez receptor, as well as the on the 5uf0 crystal structure itself.

The second one is the human HIV-1 protease protein, also bound to a small molecule, with PDB ID 5ivq.[69] This system is suitable to demonstrate some of the new features of BAT2, such as support for proteins with multiple chains, inclusion of protonated residues and the presence of co-binders. Here we will also perform ABFE calculations on 5 docked poses and the original 5ivq cocrystal structure.

3.8 Computational details

For all simulations we use rectangular periodic boxes, with a cutoff value of 9.0 Å and long-range electrostatics calculated using Particle Mesh Ewald (PME).[70] AMBER uses the SHAKE[71] and SETTLE[72] algorithms to keep rigid all bonds involving hydrogens, and OpenMM additionally uses the CCMA algorithm[73] for the same purpose. All simulations in this study use hydrogen mass repartitioning (HMR),[74] and a 4.0 fs time step. In the HMR

https://doi.org/10.26434/chemrxiv-2024-nzt5j ORCID: https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. License: CC BY-NC 4.0

procedure, performed using the Ambertools *parmed* program, the mass of each hydrogen is multiplied by a factor of 3 and this enhanced hydrogen mass is subtracted from the atom to which the hydrogen is bonded. Soft-core potentials are applied to the Lennard Jones interactions during the ligand LJ decoupling calculations, using the soft-core parameters default values for both AMBER and OpenMM. All other parameters used for the calculations, such as lambda values and simulation times, can be found in the BAT2 input files included in the Supporting Information (SI).

4 Results and discussion

In this section we test and validate the new methods included in BAT2 software, such as new restraint schemes and support for OpenMM, by comparing its results to the ones from the original BAT release on BRD4(2).[13] We also use use this same system to test calculations with very short time scales, aiming to reduce their computational cost. Finally, we apply the BAT2 workflow to a HIV-1 protease system, and evaluate the calculations in terms of accuracy, robustness, and computational cost.

4.1 BRD4(2)

For this protein, we perform all the ABFE calculations starting from the five equilibrated poses and cocrystal structure from our previous study,[13] so they have the same starting points and restraint reference states, allowing the results to be directly compared.

4.1.1 SDR and merged restraints

We perform two types of calculations for each system: a double decoupling procedure with each restraint component calculated separately, and the SDR procedure using the merged restraint scheme described in Methods. The first one corresponds to all the letters in the third column of Table 1, and we will call it "split". The second corresponds to the letters

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

in the second column of the same table, and we will call it "merged". We use both the AMBER and the OpenMM simulation engines in each case, and summarize the results in Table 3. Detailed results, with the separate values obtained for each free energy component, can be found in the Supporting Information. Also included in the SI are all the needed files to reproduce the results shown here in an automated way using the BAT2 software.

Table 3 shows the calculated binding free energies, using BAT1 and BAT2, for six equilibrated states obtained from Ref.,[13] five starting from docked poses and one from the 5uf0 crystal structure. For all starting structures, there is good agreement between the AMBER and OpenMM softwares, the split and merged procedures, and generally between BAT1 and BAT2, with the variations inside the associated uncertainties. The original BAT 1.0 article uses a total of 1.16 μs of simulations for each calculation using the split scheme, while here we use 148 ns for the same procedure, using either AMBER or OpenMM. The merged method uses a total of 100.8 ns for both softwares, bringing a more than 10-fold reduction in the simulation time needed for a single calculation when compared Ref.,[13] with little (if any) loss of accuracy. We do notice a slight increase in the uncertainties when the shorter times are used, even though the results seem to be robust across the five different calculation types for each pose. The table also shows the ligand structural root mean square deviation (RMSD) relative to the 5uf0 crystal structure, for each of the six starting states. The RMSD value is used to quantify the similatity between a ligand binding mode, generated by simulations or docking, to the one determined experimentally.

It is also important to compare the SDR and merged components to their split counterparts, in order to demonstrate that the merged free energy terms are being computed correctly. If the initial and reference states are the same, the free energy value of the merged **m** component should be the sum of the **a**, **l** and **t** free energies from the split method, and the merged **n** component should be the sum of the split **b**, **c** and **r** components. Also, the **e** component of the SDR method should be the sum of the **e** and **f** components of the DD method, and the same goes for the SDR **v** component when compared to the DD **v** and

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Table 3: Binding free energy (ΔG_{bind}) results (in kcal/mol) using different methods for the five equilibrated docked poses and the equilibrated 5uf0 cocrystal structure, with the associated uncertainties in parenthesis. Also shown are the ligand RMSDs of the initial equilibrated states relative to the initial 5uf0 structure. The BAT1 calculations used the "split" procedure with the AMBER software.

	Crystal	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5
RMSD (Å)	1.30	5.26	0.45	5.33	4.23	0.74
BAT 1.0	-6.1 (0.6)	-2.5 (0.9)	-6.7 (0.6)	-2.6 (0.8)	-1.5 (0.6)	-6.5 (0.8)
Split AMBER	-7.0 (1.5)	-2.7 (1.2)	-5.9 (1.5)	-1.7 (1.3)	-1.5 (1.9)	-7.5 (1.2)
Merged AMBER	-6.8 (1.5)	-3.5 (1.0)	-5.7 (0.8)	-3.3 (1.3)	-2.8 (1.5)	-7.3 (1.2)
Split OpMM	-6.6 (0.8)	-2.9 (0.8)	-5.9 (0.9)	-1.3 (1.1)	-1.3 (1.0)	-6.8 (0.8)
Merged OpMM	-6.2 (1.1)	-2.4 (1.4)	-6.1 (0.7)	-3.7 (0.8)	-1.5 (1.2)	-7.2 (0.8)

w free energies. This comparison is shown in Table 4 for the OpenMM software, with the same comparison for AMBER included in the SI. We observe good agreement for all merged components, with no discrepancies over 1.1 kcal/mol and most of them inside the uncertainty values. The merged method uses less simulation time and is suitable for ligands with net charge, so it might be preferable over the split one in most cases.

Table 4: Comparison between the free energies (in kcal/mol) using the merged and split schemes, for each of the merged restraints and SDR components.

	Attach restraints		Electrostatic		Lennard-Jones		Release restraints	
System	merged **m**	split **a**,**l**,**t**	SDR **e**	DD **e**,**f**	SDR **v**	DD **v**,**w**	merged **n**	split **b**,**c**,**r**
Pose 1	28.2 (0.6)	29.3 (0.4)	-0.7 (0.3)	-0.5 (0.2)	12.2 (1.0)	12.5 (0.6)	-37.3 (0.6)	-38.4 (0.5)
Pose 2	26.5 (0.3)	26.5 (0.3)	3.2 (0.2)	3.5 (0.3)	12.4 (0.6)	12.1 (0.6)	-36.1 (0.2)	-36.2 (0.5)
Pose 3	28.2 (0.4)	27.8 (0.4)	-1.2 (0.3)	-1.4 (0.3)	12.1 (0.4)	11.5 (0.9)	-35.4 (0.3)	-36.5 (0.4)
Pose 4	31.2 (0.6)	31.6 (0.4)	0.3 (0.3)	0.5 (0.3)	7.7 (0.9)	6.6 (0.8)	-37.6 (0.5)	-37.4 (0.3)
Pose 5	26.3 (0.3)	26.5 (0.3)	3.6 (0.1)	2.8 (0.4)	12.7 (0.4)	13.5 (0.5)	-35.3 (0.6)	-35.9 (0.3)

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

4.2 Reducing Simulation Time

Computational cost has always been a barrier to apply ABFE calculations on a high-throughput scale, even with the significant gains in performance obtained with the use of GPUs. Thus, one of our main goals with BAT2 is to significantly reduce the time necessary for a single ABFE calculation.

With that in mind, here we will perform two types of short calculations on the BRD4(2) system: one of them using a total of 17.4 ns of simulations per pose, and another using a total of 20.4 ns. We will call the first one the short **tevb** calculation, which as the name suggests uses only these four components, with the **e** and **v** components using the SDR method. No conformational restraints are applied to the ligand or the protein, so their corresponding free energy contributions are not present. The second type we will call the short **m*****evbc** calculation, which also uses SDR and applies conformational restraints to the ligand only. The **m*** component applies the ligand TR and conformational restraints using a single set of windows, with the ligand conformational restraints being released in a small box (**c** component) and the ligand TR release analytically (**b** component).

We have performed six ABFE calculation replicas, all using OpenMM, for each equilibrated pose from Tables 3 and 4, and the results are shown in Tables 5 and 6. Even though we have drastically reduced the time necessary for a single ABFE calculation, the results are consistent with the longer calculations, and also similar between independent replicas. Most importantly, the calculations are still able to identify the correct binding modes, with poses that have an RMSD under 2.0 Å (2 and 5, Table 3) always showing a higher affinity when compared to the other ones. There is also reasonable agreement between the binding free energy of these two poses with the experimental value of -5.2 kcal/mol,[68] even though there is a more pronounced overestimation of the affinities when compared to the longer BAT 1.0 calculations.

The short calculations from this section retain surprising accuracy when compared the ones performed using much longer time scales, reducing the computational cost by a factor of

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

more than 50 when compared to our previous BAT1 results. Even though the **tevb** approach uses slightly less simulation time, the ligand does not have its conformation restrained during the application of the TR restraints and the decoupling/recoupling steps. BAT uses three anchor atoms on the protein and three on the ligand to define the relative position and rotation of the ligand as a whole. This means that anchor atoms placed in highly flexible regions of the ligand could potentially overestimate the magnitude of the TR attachment free energy contribution.[23] Since the ligand anchor atoms are chosen automatically based on geometrical criteria only, an approach that keeps the ligand rigid such as **m*****evbc** is usually recommended. The only additional simulations in that case, when compared to a **tevb** calculation, are the release of the ligand conformational restraints in a small box (**c** component), which has a relatively low computational cost.

It is also important to put these short calculations in terms of computational effort. Using the BAT2 workflow with the OpenMM software, a single NVIDIA GTX 1070 GPU can perform a short calculation for a single pose from Tables 5 and 6 in 2.6 and 2.75 hours, respectively. Since the calculations can be run doing trivial parallelization across windows, components, poses and ligands, it is possible for a server with a few hundred GPUs to perform several thousands of calculations in the time frame of a few weeks. Newer graphics cards such as the RTX 30 and 40 series can significantly reduce the time needed for a single calculation, thus also increasing the number of ligands that can be tested in a given time.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Table 5: Calculated binding free energies for the six replicas of each pose, using the short **tevb** procedure. The uncertainties for the replicas are computed through the usual block data analysis, and the uncertainty of the averages are the standard deviation across replicas. For comparison, the last row shows the results for the longer calculations using 100.8 ns (last row of Table 3).

tevb calculation (17.4 ns)					
Replica	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5
1	-2.5 (1.8)	-6.9 (1.2)	0.0 (0.9)	-2.0 (0.7)	-8.6 (1.2)
2	-0.5 (0.8)	-7.2 (0.6)	-1.0 (0.9)	-0.2 (0.7)	-8.7 (0.9)
3	-3.1 (1.4)	-4.5 (0.6)	-4.0 (1.0)	-1.7 (1.8)	-8.4 (1.2)
4	-1.2 (1.0)	-8.6 (1.0)	-2.2 (1.0)	-0.6 (1.2)	-7.4 (1.1)
5	-3.4 (0.8)	-8.1 (1.2)	-2.1 (0.6)	-2.7 (1.5)	-6.9 (0.9)
6	-1.2 (0.9)	-6.0 (0.9)	-3.1 (1.1)	-3.3 (1.4)	-6.5 (0.9)
Average	-2.0 (1.1)	-6.9 (1.4)	-2.1 (1.3)	-1.8 (1.1)	-7.7 (0.9)
Long	-2.4 (1.4)	-6.1 (0.7)	-3.7 (0.8)	-1.5 (1.2)	-7.2 (0.8)

Table 6: Same as Table 5, but using the short **m*evbc** procedure for the six replicas. Here also the last row shows the results for the longer calculations using 100.8 ns.

m*evbc calculation (20.4 ns)					
Replica	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5
1	-3.3 (1.1)	-7.2 (1.6)	-2.2 (0.9)	-1.1 (1.1)	-8.1 (0.8)
2	-3.6 (1.4)	-7.4 (1.1)	-3.2 (1.2)	-3.3 (1.1)	-6.2 (1.3)
3	-1.8 (1.2)	-7.4 (0.8)	-2.7 (1.1)	0.5 (2.0)	-7.2 (0.6)
4	-3.4 (1.3)	-7.6 (1.4)	-3.5 (1.3)	-1.3 (1.4)	-8.9 (1.0)
5	-3.2 (1.7)	-7.0 (1.6)	-3.0 (0.8)	-0.7 (1.0)	-6.9 (1.4)
6	-1.8 (1.2)	-7.4 (1.7)	-2.6 (1.5)	-2.4 (2.1)	-6.8 (1.3)
Average	-2.9 (0.8)	-7.3 (0.2)	-2.9 (0.4)	-1.4 (1.2)	-7.4 (0.9)
Long	-2.4 (1.4)	-6.1 (0.7)	-3.7 (0.8)	-1.5 (1.2)	-7.2 (0.8)

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

4.3 HIV-1 Protease

To illustrate some of the new features of BAT.py, we also apply our workflow to the HIV-1 protease system with PDB ID 5ivq (Figure 4). Binding free energy calculations are performed after equilibration on five self-docked poses using the Autodock Vina[75] software, and also on the equilibrated 5ivq crystal structure. We choose OpenMM for all calculations performed in this section, choosing two different sets of free energy components.

The first set has the protein backbone restraints applied to four non-contiguous sections of the protein backbone (Fig. 4), which we call **mevn**, or simply "merged" as in section 4.1.1. The second does not include protein conformational restraints, and we name **m*evbc** as in section 4.2. Both use a total of 100.8 ns of simulations per ABFE calculation, but **m*evbc** is slightly cheaper, due to the smaller size of the **c** component simulation box when compared to **n**. Complete results and all input files, including the initial structures and the BAT2 parameters used for both approaches, are included in the SI.

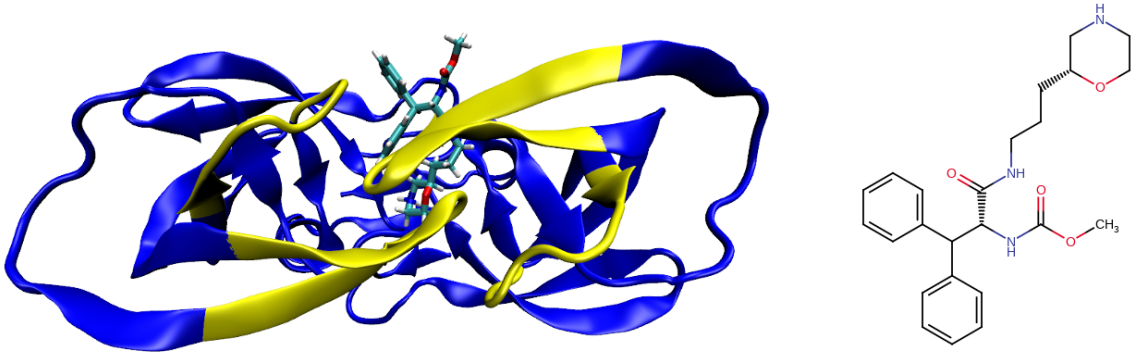

Figure 4: (*left*) The 5ivq structure of the HIV-1 protease, with the restrained backbone regions in yellow. (*right*) The ligand from this structure, where we include a +1 net charge on the morpholine ring.

The results are shown in Table 7. For both methods, the docked pose that is closest to the crystal structure, which is pose 3, has the highest affinity. The binding free energy values obtained for this pose are also near the values obtained for the known crystal structure,

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

even though they are slightly lower, but still within the uncertainties (added in quadrature). The experimental data for this ligand reports 11 % inhibition of HIV-1 protease, at a ligand concentration of 1 μM and a protein concentration of 20 pM.[69] This puts the affinity roughly between -7 and -8 kcal/mol, and thus near the values calculated using BAT2. Thus, for this example we were able to correctly identify the correct docked pose, even though it was not the one with the best Vina score, and also produce binding free energy values comparable to experiments.

Table 7: Binding free energies (in kcal/mol) for the five docked and equilibrated poses, as well as for the equilibrated 5ivq cocrystal structure. Also shown are the ligand RMSDs of the equilibrated states relative to the initial 5ivq structure.

	Crystal	Pose 1	Pose 2	Pose 3	Pose 4	Pose 5
RMSD (Å)	0.35	5.48	9.01	1.65	5.64	8.91
mevn	-7.3 (1.5)	2.5 (2.4)	-1.5 (1.9)	-9.6 (2.8)	-0.4 (2.4)	0.1 (2.0)
m*evbc	-7.2 (1.4)	4.2 (1.7)	0.1 (1.1)	-9.7 (1.9)	-2.7 (1.6)	-3.0 (1.3)

5 Conclusions

We have presented here the BAT2 software, currently in its 2.3 version, explaining the theory behind it and testing it on two sample systems. On the first, good agreement is observed between calculations performed using AMBER or OpenMM, and also between different choices for the free energy components that make up the total binding free energy. We also show a significant reduction in the computational cost of ABFE calculations, making it now possible to test hundreds of thousands of ligands on a reasonable time. The second system illustrates some of the new BAT2 features, that now make it applicable to virtually any protein-ligand system. Here the results are also consistent with the available experimental data, with BAT2 being able to find the experimental binding pose and to correctly estimate the binding free energy.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

When compared to its earlier version BAT1, BAT2 is more broadly applicable, easier to set up for new systems, cheaper to run due to the merged windows scheme, and compatible with the free and open-source OpenMM simulation engine. BAT.py itself is an open-source software, freely available for download at the GitHub platform.

Acknowledgement

We thank Andrea Rizzi for helpful discussions. The authors gratefully acknowledge the generous support provided by the Sanders Tri- Institutional Therapeutics Discovery Institute (TDI), a 501(c)(3) organization. TDI receives financial support from Takeda Pharmaceutical Company, TDI's parent institutes (Memorial Sloan Kettering Cancer Center, The Rockefeller University and Weill Cornell Medicine), and from a generous contribution from Lewis Sanders and other philanthropic sources. GH thanks FAPESC and CNPq for the research grants. MKG has an equity interest in and is a cofounder and scientific advisor of VeraChem LLC, Denovicon and InCerebro.

Supporting Information Available

SI includes additional tables and all input files needed to reproduce the results from this work.

References

(1) Michael, S.; Auld, D.; Klumpp, C.; Jadhav, A.; Zheng, W.; Thorne, N.; Austin, C. P.; Inglese, J.; Simeonov, A. A robotic platform for quantitative high-throughput screening. *Assay and Drug Development Technologies* **2008**, *6*.

(2) Dörr, M.; Fibinger, M. P.; Last, D.; Schmidt, S.; Santos-Aberturas, J.; Böttcher, D.; Hummel, A.; Vickers, C.; Voss, M.; Bornscheuer, U. T. Fully automatized high-

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

throughput enzyme library screening using a robotic platform. *Biotechnology and Bioengineering* **2016**, *113*.

(3) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous free energy simulations in virtual screening. *Journal of Chemical Information and Modeling* **2020**, *60*.

(4) Walters, W. P.; Wang, R. New trends in virtual screening. *Journal of Chemical Information and Modeling* **2020**, *60*.

(5) Chen, B.; Harrison, R. F.; Papadatos, G.; Willett, P.; Wood, D. J.; Lewell, X. Q.; Greenidge, P.; Stiefl, N. Evaluation of machine-learning methods for ligand-based virtual screening. *Journal of Computer-Aided Molecular Design* **2007**, *21*.

(6) Plewczynski, D.; Spieser, S.; Koch, U. Performance of Machine Learning Methods for Ligand-Based Virtual Screening. *Combinatorial Chemistry & High Throughput Screening* **2009**, *12*.

(7) Torres, P. H.; Sodero, A. C.; Jofily, P.; Silva-Jr, F. P. Key topics in molecular docking for drug design. 2019.

(8) Ferreira, L. G.; Santos, R. N. D.; Oliva, G.; Andricopulo, A. D. Molecular docking and structure-based drug design strategies. 2015.

(9) Murugan, N. A.; Priya, G. R.; Sastry, G. N.; Markidis, S. Artificial intelligence in virtual screening: Models versus experiments. *Drug Discovery Today* **2022**, *27*.

(10) Pandey, M.; Fernandez, M.; Gentile, F.; Isayev, O.; Tropsha, A.; Stern, A. C.; Cherkasov, A. The transformational role of GPU computing and deep learning in drug discovery. *Nature Machine Intelligence* **2022**, *4*.

(11) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beautrait, A.; Vendome, J.; Abel, R.; Friesner, R. A.; Wang, L. Enhancing Hit Discovery in

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Virtual Screening through Absolute Protein-Ligand Binding Free-Energy Calculations. *Journal of Chemical Information and Modeling* **2023**, *63*.

(12) Muegge, I.; Hu, Y. Recent Advances in Alchemical Binding Free Energy Calculations for Drug Discovery. *ACS Medicinal Chemistry Letters* **2023**, *14*.

(13) Heinzelmann, G.; Gilson, M. K. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific Reports* **2021**, *11*.

(14) Fu, H.; Zhou, Y.; Jing, X.; Shao, X.; Cai, W. Meta-Analysis Reveals That Absolute Binding Free-Energy Calculations Approach Chemical Accuracy. *Journal of Medicinal Chemistry* **2022**, *65*.

(15) Feng, M.; Heinzelmann, G.; Gilson, M. K. Absolute binding free energy calculations improve enrichment of actives in virtual compound screening. *Scientific Reports* **2022**, *12*.

(16) Tembe, B. L.; Mc Cammon, J. A. Ligand-receptor interactions. *Computers & Chemistry* **1984**, *8*, 281–283.

(17) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of Chemical Information and Modeling* **2017**, *57*.

(18) Yang, Q.; Burchett, W.; Steeno, G. S.; Liu, S.; Yang, M.; Mobley, D. L.; Hou, X. Optimal designs for pairwise calculation: An application to free energy perturbation in minimizing prediction variability. *Journal of Computational Chemistry* **2020**, *41*.

(19) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.;

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society* **2015**, *137*.

(20) Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *Journal of Chemical Information and Modeling* **2015**, *55*.

(21) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *Journal of Chemical Theory and Computation* **2017**, *13*.

(22) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science* **2016**, *7*.

(23) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *Journal of Physical Chemistry B* **2003**, *107*.

(24) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational Calorimetry: High-Precision Calculation of Host-Guest Binding Thermodynamics. *Journal of Chemical Theory and Computation* **2015**, *11*.

(25) Woo, H. J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*.

(26) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophysical Journal* **1997**, *72*.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

(27) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water. *The Journal of Chemical Physics* **1988**, *89*.

(28) Heinzelmann, G.; Chen, P. C.; Kuyucak, S. Computation of standard binding free energies of polar and charged ligands to the glutamate receptor glua2. *Journal of Physical Chemistry B* **2014**, *118*.

(29) Huggins, D. J. Comparing the Performance of Different AMBER Protein Forcefields, Partial Charge Assignments, and Water Models for Absolute Binding Free Energy Calculations. *Journal of Chemical Theory and Computation* **2022**, *18*.

(30) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead optimization mapper: Automating free energy calculations for lead optimization. *Journal of Computer-Aided Molecular Design* **2013**, *27*.

(31) Lee, T.-S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13*, 3077–3084.

(32) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **2013**, *9*, 3878–3888.

(33) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(34) 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.;

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13*, e1005659.

(35) Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 556–560.

(36) Kim, S.; Oshima, H.; Zhang, H.; Kern, N. R.; Re, S.; Lee, J.; Roux, B.; Sugita, Y.; Jiang, W.; Im, W. CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations. *Journal of Chemical Theory and Computation* **2020**, *16*, 7207–7218, Publisher: American Chemical Society.

(37) Eastman, P.; Pande, V. S. Constant constraint matrix approximation: A robust, parallelizable constraint method for molecular simulations. *Journal of Chemical Theory and Computation* **2010**, *6*.

(38) Eastman, P.; Pande, V. S. OpenMM: A Hardware-Independent Framework for Molecular Simulations. *Computing in Science and Engineering* **2010**, *12*.

(39) Peter, E.; Pande, V. S. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *Journal of Computational Chemistry* **2010**, *31*.

(40) Chodera, J. D.; Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *Journal of Chemical Physics* **2011**, *135*.

(41) Wang, L. P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *Journal of Physical Chemistry Letters* **2014**, *5*.

(42) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: A different approach. *Journal of Physical Chemistry Letters* **2014**, *5*.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

(43) Martínez, L.; Andreani, R.; Martínez, J. M. Convergent algorithms for protein structural alignment. *BMC Bioinformatics* **2007**, *8*.

(44) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. MUSTANG: a multiple structural alignment algorithm. *Proteins* **2006**, *64*, 559–574.

(45) Setiadi, J.; Boothroyd, S.; Slochower, D. R.; Dotson, D. L.; Thompson, M. W.; Wagner, J. R.; Wang, L.-P.; Gilson, M. K. Tuning Potential Functions to Host–Guest Binding Data. *Journal of Chemical Theory and Computation* **2024**, *20*, 239–252, PMID: 38147689.

(46) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J Cheminform* **2011**, *3*, 33.

(47) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(48) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; III,; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Forouzesh, N.; Giambaşu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, J.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Zhu, Q.; ; Kollman, P. A. *AMBER23*; University of California, San Francisco: San Francisco, CA, 2023.

(49) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.;

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

Meng, E. C.; Ferrin, T. E. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25*, 1605–1612.

(50) 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.

(51) 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.

(52) 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.

(53) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*.

(54) 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.

(55) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(56) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(57) Li, P.; Merz, K. M. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *Journal of Chemical Theory and Computation* **2014**, *10*.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

(58) Li, P.; Song, L. F.; Merz, K. M. Systematic parameterization of monovalent ions employing the nonbonded model. *Journal of Chemical Theory and Computation* **2015**, *11*.

(59) Li, Z.; Song, L. F.; Li, P.; Merz, K. M. Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *Journal of Chemical Theory and Computation* **2020**, *16*.

(60) Sengupta, A.; Li, Z.; Song, L. F.; Li, P.; Merz, K. M. Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *Journal of Chemical Information and Modeling* **2021**, *61*.

(61) Joung, I. S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B* **2008**, *112*.

(62) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* **2008**, *129*.

(63) Lin, Y.-L.; Aleksandrov, A.; Simonson, T.; Roux, B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2690–2709.

(64) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139*, 184103.

(65) Öhlknecht, C.; Perthold, J. W.; Lier, B.; Oostenbrink, C. Charge-Changing Perturbations and Path Sampling via Classical Molecular Dynamic Simulations of Simple Guest–Host Systems. *Journal of Chemical Theory and Computation* **2020**, Publisher: American Chemical Society.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

(66) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32*, 523–535.

(67) Aqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **2004**, *384*, 288–294.

(68) Wang, L.; Pratt, J. K.; Soltwedel, T.; Sheppard, G. S.; Fidanze, S. D.; Liu, D.; Hasvold, L. A.; Mantei, R. A.; Holms, J. H.; McClellan, W. J.; Wendt, M. D.; Wada, C.; Frey, R.; Hansen, T. M.; Hubbard, R.; Park, C. H.; Li, L.; Magoc, T. J.; Albert, D. H.; Lin, X.; Warder, S. E.; Kovar, P.; Huang, X.; Wilcox, D.; Wang, R.; Rajaraman, G.; Petros, A. M.; Hutchins, C. W.; Panchal, S. C.; Sun, C.; Elmore, S. W.; Shen, Y.; Kati, W. M.; McDaniel, K. F. Fragment-Based, Structure-Enabled Discovery of Novel Pyridones and Pyridone Macrocycles as Potent Bromodomain and Extra-Terminal Domain (BET) Family Bromodomain Inhibitors. *J. Med. Chem.* **2017**, *60*, 3828–3850.

(69) Bungard, C. J.; Williams, P. D.; Ballard, J. E.; Bennett, D. J.; Beaulieu, C.; Bahnck-Teets, C.; Carroll, S. S.; Chang, R. K.; Dubost, D. C.; Fay, J. F.; Diamond, T. L.; Greshock, T. J.; Hao, L.; Holloway, M. K.; Felock, P. J.; Gesell, J. J.; Su, H. P.; Manikowski, J. J.; McKay, D. J.; Miller, M.; Min, X.; Molinaro, C.; Moradei, O. M.; Nantermet, P. G.; Nadeau, C.; Sanchez, R. I.; Satyanarayana, T.; Shipe, W. D.; Singh, S. K.; Truong, V. L.; Vijayasaradhi, S.; Wiscount, C. M.; Vacca, J. P.; Crane, S. N.; McCauley, J. A. Discovery of MK-8718, an HIV Protease Inhibitor Containing a Novel Morpholine Aspartate Binding Group. *ACS Medicinal Chemistry Letters* **2016**, *7*.

(70) 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.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

(71) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*.

(72) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **1992**, *13*, 952–962.

(73) Eastman, P.; Pande, V. S. CCMA: A Robust, Parallelizable Constraint Method for Molecular Simulations. *Journal of chemical theory and computation* **2010**, *6*.

(74) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

(75) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **2010**, *31*.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0

TOC Graphic

Some journals require a graphical entry for the Table of Contents. This should be laid out "print ready" so that the sizing of the text is correct.
Inside the `tocentry` environment, the font used is Helvetica 8 pt, as required by *Journal of the American Chemical Society*.
The surrounding frame is 9 cm by 3.5 cm, which is the maximum permitted for *Journal of the American Chemical Society* graphical table of content entries. The box will not resize if the content is too big: instead it will overflow the edge of the box.
This box and the associated title will always be printed on a separate page at the end of the document.

https://doi.org/10.26434/chemrxiv-2024-nzt5j **ORCID:** https://orcid.org/0000-0001-7168-6941 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC 4.0