User guide for BAT.py relative calculations (v2.4)

Table of Contents

1. Introduction

In this guide we explain the theory, methodology and practical aspects of the relative binding free energy (RBFE) calculations implemented in the 2.4 version of the BAT.py software. All the functionalities from the BAT.py ABFE calculations [1] are also included in the RBFE option, such as a fully automated protocol, choice of various restraint schemes, use of the OpenMM [2] and AMBER [3] software on Graphics Processing Units (GPUs), and application to any protein system of choice. For that reason, here we will only focus on the specificities of the RBFE calculations, and for all other BAT aspects we invite the reader to read the BAT.py User Guide available in the GitHub distribution.

2. Theory

The RBFE calculations from BAT.py will use the separate topologies (SepTop) approach, which was originally proposed by Rocklin et al. [4], and more recently reintroduced by Baumann et al. [5]. Here we add a few modifications when compared to the latter, and we propose a new thermodynamic cycle to obtain the difference in affinity between any pair of ligands.

In Figure 1 we show our proposed cycle, with a full transformation that computes the binding free energy difference between a reference (1) and a target (2) ligands, with the associated BAT.py free energy components and free energy terms. The springs denote the three types of restraints (ΔG_{rest} terms) for the systems that contain ligands 1 and/or 2: black for translational/rotational (TR) restraints on the ligand ($\Delta G_{l,TR,att/rel}$), blue for protein backbone conformational restraints ($\Delta G_{p,att/rel}$), and red for ligand conformational restraints ($\Delta G_{l,conf,att/rel}$), following the equations:

$$\Delta G_{rest,site} = \Delta G_{p,att} + \Delta G_{l,conf,att} + \Delta G_{l,TR,att} \quad (1)$$

$$\Delta G_{rest,bulk} = \Delta G_{p,rel} + \Delta G_{l,conf,rel} + \Delta G_{l,TR,rel} \quad , \quad (2)$$

with more details on these free energy terms and their calculation methods in the BAT.py User Guide. The electrostatic components of the calculation using simultaneous decoupling and recoupling (SDR), for either ligand 1 or 2, are also the same as the ones used for the original ABFE calculations [1], and calculated separately for each ligand as done with the restraint free energies.

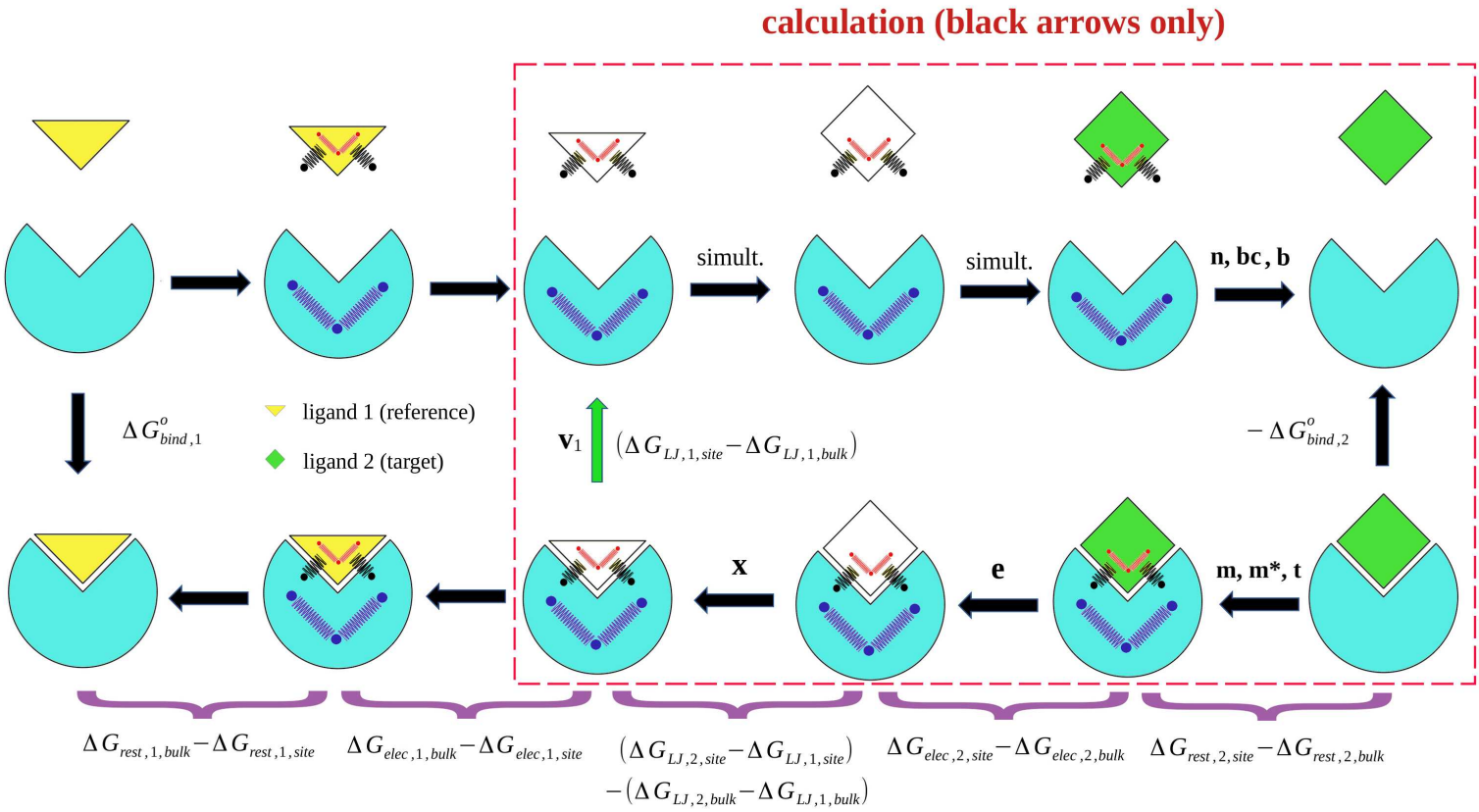

Figure 1: Thermodynamic cycle used by BAT.py for RBFE calculations using SepTop.

The main difference here from the original BAT ABFE calculations is the inclusion of the **x** component, which describes the exchange between two neutral ligands (colored white) between binding site and bulk, as shown in the cycle. This component is calculated with a simulation box similar to the one used for the SDR calculations, but here one ligand will be transformed into the other in the binding site, and the opposite transformation will be carried out in bulk. The **x** component is the only calculation in which the two ligands will coexist in the same simulation box, both of them fully restrained. The free energy associated with this transformation is given by:

$$\Delta G_{exchange} = (\Delta G_{LJ,2,site} - \Delta G_{LJ,1,site}) - (\Delta G_{LJ,2,bulk} - \Delta G_{LJ,1,bulk}) \quad (3)$$

, with the *LJ* index denoting the decoupling of the ligand Lennard-Jones interactions. Thus the full transformation shown in the cycle is equivalent to the equation below:

$$-(\Delta G^o_{bind,2} - \Delta G^o_{bind,1}) = (\Delta G_{rest,2,site} - \Delta G_{rest,2,bulk}) + (\Delta G_{elec,2,site} - \Delta G_{elec,2,bulk}) + (\Delta G_{LJ,2,site} - \Delta G_{LJ,1,site}) - (\Delta G_{LJ,2,bulk} - \Delta G_{LJ,1,bulk}) + (\Delta G_{elec,1,bulk} - \Delta G_{elec,1,site}) + (\Delta G_{rest,1,bulk} - \Delta G_{rest,1,site})$$

(4)

Rearranging the terms of this equation, we can easily check that is also corresponds to the difference in the ABFE of the two ligands, and thus our cycle is valid:

$$-(\Delta G^o_{bind,2} - \Delta G^o_{bind,1}) = (\Delta G_{rest,2,site} + \Delta G_{elec,2,site} + \Delta G_{LJ,2,site} - \Delta G_{LJ,2,bulk} - \Delta G_{elec,2,bulk} - \Delta G_{rest,2,bulk}) - (\Delta G_{rest,1,site} + \Delta G_{elec,1,site} + \Delta G_{LJ,1,site} - \Delta G_{LJ,1,bulk} - \Delta G_{elec,1,bulk} - \Delta G_{rest,1,bulk})$$

(5)

As noted elsewhere [4,5], there are a few reasons why RBFE with SepTop can be advantageous over regular RBFE or ABFE in certain cases. Compared to the former, it allows for the transformation

between ligands that have no similarities, broadening the scope of this method. Also, as opposed to ABFE, in SepTop there is no need for sampling of the protein apo state, which can bring convergence issues in the cases of protein conformational changes or for occluded binding sites. Our particular protocol has the further advantage of being able to compare ligands that have different net charges, since the electrostatic calculations are done separately for each ligand using the SDR method.

3. Methods

3.1 Preparation and equilibration

The workflow for the relative calculations is the exact same one from the original ABFE procedure from BAT.py [1], in which a series of systems containing different binding poses or ligands to the same protein is prepared and equilibrated before the free energy calculations. This equilibration procedure will always be done for the target ligand from the cycle of Fig. 1, with the reference ligand being the first molecule listed in the `poses_list` ("dock" option for `calc_type`), `ligand_list` ("rank" option for `calc_type`) or `celpp_receptor` ("crystal" option for `calc_type`) variable. The reference molecule will also be equilibrated since it has itself as the target molecule for the first ligand/pose in the list.

Once the equilibration is concluded, the resulting state will be used to build the different systems needed for the binding free energy calculations. For the **x** component, which is the one in which both molecules coexist, the equilibrated receptor and target ligand will adopt the coordinates from the equilibrated system that has the target molecule bound, and the coordinates for the reference ligand will come from its own equilibrated state. The various restraints for each molecule will have their reference values from the starting coordinates after equilibration, and thus will follow the same definitions.

For the cycle of Fig. 1 to be valid, the restraint coordinates at the initial and final states of the **x** transformation have to be consistent for all poses/ligands, including the reference one. In the case of the ligand conformational restraints, this always holds true because the equilibrated conformation of the reference ligand does not depend on the target. The same goes for the TR restraints, which are defined relative to the equilibrated reference receptor. However, if the user also applies protein conformational restraints, the same receptor should be used for all targets, as with the "dock" and "rank" options for `calc_type`. If using the "crystal" option, make sure that their backbones are at least very similar. In both cases, the `bb_equil` option should always be set to "yes", in order to retain the protein restrained state during equilibration of the target systems.

3.2 Free energy terms

The free energy components that will be calculated for a given target molecule are the ones inside the dashed red square in the cycle of Fig. 1. They include the application/removal of restraints and the electrostatic calculations for the target molecule (2), as well as the exchange free energy between the two neutral ligands. Thus the BAT result for a transformation between the reference and target molecules will contain the following terms:

$$-\Delta G_{BAT\,1\to2} = (\Delta G_{rest,2,site} - \Delta G_{rest,2,bulk}) + (\Delta G_{elec,2,site} - \Delta G_{elec,2,bulk}) + (\Delta G_{LJ,2,site} - \Delta G_{LJ,1,site}) - (\Delta G_{LJ,2,bulk} - \Delta G_{LJ,1,bulk}) \quad (6)$$

This result differs from the relative binding free energy between the two ligands (Eq. 4) by an offset term ΔG_0 that only involves the reference molecule, and consequently is the same for all targets:

$$\Delta G_{BAT\,1\to2} = \left(\Delta G^{o}_{bind,2} - \Delta G^{o}_{bind,1}\right) + \Delta G_0 \tag{7}$$

$$\Delta G_0 = +\left(\Delta G_{elec,1,bulk} - \Delta G_{elec,1,site}\right) + \left(\Delta G_{rest,1,bulk} - \Delta G_{rest,1,site}\right) \tag{8}$$

That way, even though the $\Delta G_{BAT\,1\to2}$ term does not have a physical meaning, from Eq. 7 we see that the following equality always holds for any two targets 2 and 3:

$$\Delta G^{o}_{bind,3} - \Delta G^{o}_{bind,2} = \Delta G_{BAT\,1\to3} - \Delta G_{BAT\,1\to2} \quad , \tag{9}$$

and so the BAT results are sufficient to obtain the RBFE between any two or more ligands, as long as they are relative to the same reference. If the goal is only calculating the relative affinity between ligands 1 and 2, the BAT result from the second one ($\Delta G_{BAT\,1\to2}$) can be subtracted of the result obtained from the first ($\Delta G_{BAT\,1\to1}$) , which is the reference molecule being transformed into itself. Note that the latter should *not* be zero, due to the presence of the ΔG_0 term.

3.3 Example of results

Below we show a set of RBFE results from BAT, for 5 poses from the ligand in the 5uf0 PDB structure docked to the BRD4(2) bromodomain receptor from the 5uez PDB structure. We start the free energy calculations from the same equilibrated states as the ones used in Refs. [1,6] for this system, using pose 4 as the reference molecule for all poses including itself:

```
Merged components SDR method

Component            Free Energy; Sigma

Attach all;              28.16;    0.40
Electrostatic (TI);      -0.87;    0.33
LJ exchange (TI);         4.90;    0.13
Release all;            -37.20;    0.71

Relative free energy;      5.01;    0.88
```

pose 1

```
Merged components SDR method

Component            Free Energy; Sigma

Attach all;              27.27;    0.39
Electrostatic (TI);       3.40;    0.33
LJ exchange (TI);         5.93;    0.57
Release all;            -35.82;    0.64

Relative free energy;     -0.78;    0.99
```

pose 2

```
Merged components SDR method

Component            Free Energy; Sigma

Attach all;              27.34;    0.61
Electrostatic (TI);      -1.69;    0.22
LJ exchange (TI);         5.80;    1.03
Release all;            -35.74;    0.47

Relative free energy;      4.29;    1.30
```

pose 3

```
Merged components SDR method

Component            Free Energy; Sigma

Attach all;              30.36;    0.62
Electrostatic (TI);      -0.05;    0.27
LJ exchange (TI);         0.45;    0.23
Release all;            -37.07;    0.25

Relative free energy;      6.30;    0.76
```

pose 4

```
Merged components SDR method

Component            Free Energy; Sigma

Attach all;              25.88;    0.58
Electrostatic (TI);       2.66;    0.23
LJ exchange (TI);         5.08;    0.34
Release all;            -34.61;    0.35

Relative free energy;      0.99;    0.80
```

pose 5

The relative free energies listed correspond to the $\Delta G_{BAT\,ref\to target}$ values, and so the relative affinities between any two poses can be obtained by the difference in the BAT result for each according to Eq. 9.

They should be consistent with the ones obtained by just computing the difference between the ABFE results for two poses (Eqs. 5 and 9), which are available at Ref. [1]. Note that the RBFE values differ from the ABFE ones by an offset value that is the same for all poses, which is in fact the *LJ* decoupling terms for the reference molecule (pose 4):

$$\Delta G_{BAT\ ref \rightarrow target} - \left(\Delta G_{LJ,ref,site} - \Delta G_{LJ,ref,bulk} \right) = \Delta G^{o}_{bind,target} \quad (10)$$

This is demonstrated in the thermodynamic cycle from Fig. 1 by the green arrow. This *LJ* term does not have to be calculated at all, since any offset value can be used. For example, if comparing different ligands, the same number can be added to all relative affinities so that the reference one displays its experimental binding free energy, and all the others would be ranked accordingly.

3.4 Calculation types and computational cost

Even though the example shown above uses all three types of restraints (Eqs. 1 and 2), this is not always needed or desired. To reduce computational cost, it is interesting to perform calculations that have fewer free energy components, or use components that have smaller simulation systems (or boxes). The three types of boxes BAT uses are shown in Figure 2: the SDR box, with the complex/receptor and the bulk ligand; the complex box, with only the receptor or complex; and the small box containing the only the ligand.

With that in mind, we list below three possible types of calculations, similar to the ones from Ref. [1], with the restraint scheme and the free energy components needed for each.

- **mexn**: all restraints applied, with three components that use the SDR boxes (**e**, **x** and **n**), and one free energy component that uses the complex box (**m**). Expensive and not always needed.

- **m*exbc**: no protein conformational restraints, with two components that use the SDR boxes (**e** and **x**), one free energy component that uses the complex box (**m***), and one that uses the small ligand box (**c**), with component **b** calculated analytically. Preferred in most cases.

- **texb**: only ligand TR restraints, using two components with SDR boxes (**e** and **x**) and one with the complex box (**t**). Cheapest, suitable for more rigid ligands.

Thus, even though the cycle from Figure 1 has many steps and transformations, in practice our approach needs as little as three free energy components to obtain the relative free energy of a given binding pose or ligand.

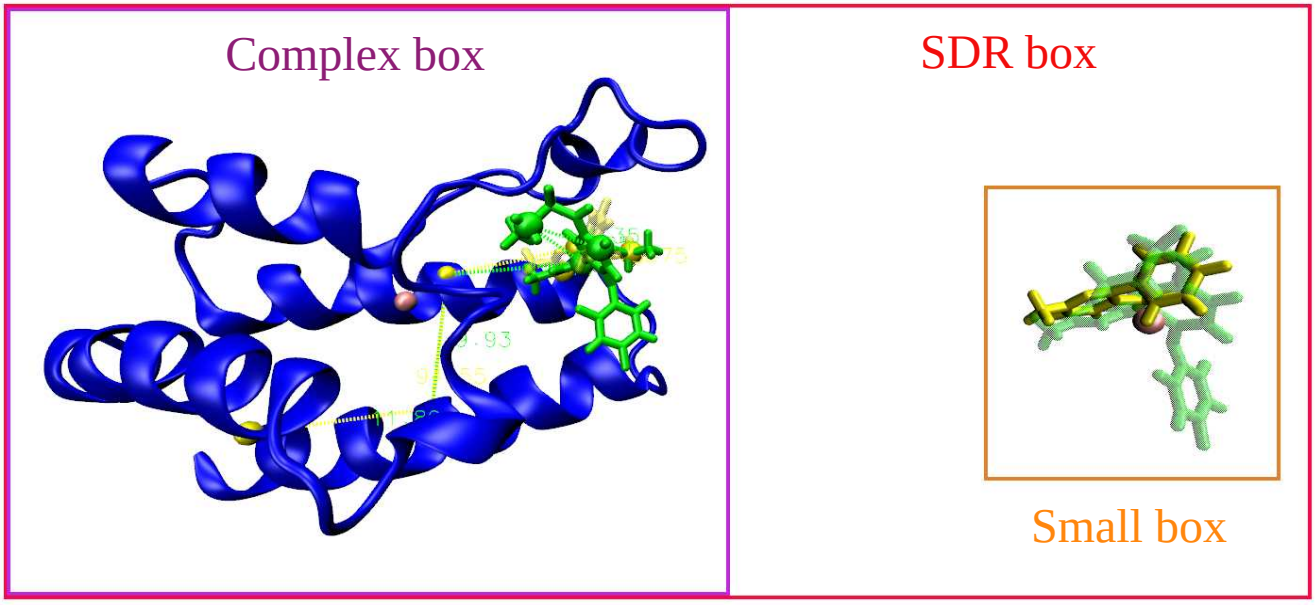

Figure 2: The three box sizes used by BAT.py. Also shown are the two restrained ligands in the binding site and bulk during the **x** transformation that uses the SDR box.

4. How to run the RBFE calculations

In order to run the RBFE calculations with BAT, one can select the "relative" option for the `fe_type` BAT input variable, for which a **mexn** calculation will be performed. One can also choose the "custom" option for `fe_type` and select the components individually (including **x**), using the option "exchange" for the `dec_method` variable. The `calc_type` options "dock", "rank" and "crystal" remain available for relative calculations, as well as the software choice between OpenMM and AMBER, and all the other variables used for ABFE.

The folder ./BAT/example-input-files contains a few examples of BAT input files used for relative calculations, including the **m*exbc** and **texb** options described above. The coordinates for all ligands and receptors are included in the ./BAT/all-poses folder. The procedure to run the RBFE calculations is the same as the one described in the tutorial for ABFE, only changing the BAT input file according to the desired calculation type. The results produced by BAT are then interpreted as explained in section 3.3.

5. References

[1] G. Heinzelmann, D. J. Huggins and M. K. Gilson (2024). "BAT2: an Open-Source Tool for Flexible, Automated, and Low Cost Absolute Binding Free Energy Calculations". *Journal of Chemical Theory and Computation*, **20**, 6518.

[2] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande (2017). "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics." *PLOS Computational Biology,* **13**, e1005659.

[3] D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P.

Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F.Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, N.R.Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman (2020), AMBER 2020, University of California, San Francisco.

[4] G. J. Rocklin, D. L. Mobley and K. A. Dill (2023). “Separated topologies: A method for relative binding free energy calculations using orientational restraints”. *Journal of Chemical Physics*, **138**, 085104.

[5] H. M. Baumann, E. Dybeck, C. L. McClendon, F. C. Pickard IV, V. Gapsys, L. Pérez-Benito, D. F. Hahn, G. Tresadern, A. M. Mathiowetz and D. L. Mobley (2023) “Broadening the Scope of Binding Free Energy Calculations Using a Separated Topologies Approach”. *Journal of Chemical Theory and Computation*, **19**, 5058.

[6] G. Heinzelmann and M. K. Gilson (2021). “Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation”. *Scientific Reports,* **11**, 1116.