# 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 [1], for both the OpenMM [2] and AMBER [3] simulation engines.

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 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 all electrostatic calculations use the SDR method.

All the functionalities from the ABFE calculations are also included in the RBFE option, such as a fully automated protocol, choice of various restraint schemes, use of the 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 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 three different thermodynamic paths to obtain the difference in affinity between any pair of ligands. In Figure 1 we show our proposed cycles, 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. In the next sections we will explain in detail the cycles and each free energy component.

### 2.1 Restraints

The springs on Fig. 1 denote the three types of restraints ( $\Delta G_{rest}$ 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:

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

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

The restraints are always calculated separately for each ligand, target or reference, and therefore are identical to the ones from the ABFE calculations. More details on the restraints free energy terms can be found in our previous work on the BAT software [1] and in the software User Guide.

### 2.2 The 'x' component

The first path we will use to calculate the RBFEs is shown in the left side of Fig. 1, and will include the new **x** free energy component. This component describes the exchange of two neutral and fully restrained ligands (colored white) between binding site and bulk, as shown in the cycle. It is calculated with a simulation box similar to the one used for the SDR calculations, but with two ligands coexisting in the same box, both in the binding site and in the bulk solvent. The free energy associated with this transformation is given by:

$$-\Delta G_{LJ1\to 2} = (\Delta G_{LJ,2,site} - \Delta G_{LJ,1,site}) - (\Delta G_{LJ,2,bulk} - \Delta G_{LJ,1,bulk}) \quad (3)$$

, with the *LJ* index denoting the decoupling of the ligand Lennard-Jones interactions. The electrostatic components of the calculation for ligands 1 and 2 (components **e**), will be calculated separately for each ligand, and thus are the same as the ones from the ABFE calculations [1]. Thus the full transformation shown in the cycle is equivalent to the equation below:

$$\begin{aligned} -(\Delta G^{o}_{bind,2} - \Delta G^{o}_{bind,1}) = {} & (\Delta G_{rest,2,site} - \Delta G_{rest,2,bulk}) + (\Delta G_{elec,2,site} - \Delta G_{elec,2,bulk}) + (\Delta G_{LJ,2,site} - \Delta G_{LJ,1,site}) \\ & - (\Delta G_{LJ,2,bulk} - \Delta G_{LJ,1,bulk}) + (\Delta G_{elec,1,bulk} - \Delta G_{elec,1,site}) + (\Delta G_{rest,1,bulk} - \Delta G_{rest,1,site}) \end{aligned}$$

(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:

$$\begin{aligned} -(\Delta G^{o}_{bind,2} - \Delta G^{o}_{bind,1}) = {} & (\Delta G_{rest,2,site} + \Delta G_{elec,2,site} + \Delta G_{LJ,2,site} - \Delta G_{LJ,2,bulk} - \Delta G_{elec,2,bulk} - \Delta G_{rest,2,bulk}) \\ & - (\Delta G_{rest,1,site} + \Delta G_{elec,1,site} + \Delta G_{LJ,1,site} - \Delta G_{LJ,1,bulk} - \Delta G_{elec,1,bulk} - \Delta G_{rest,1,bulk}) \end{aligned}$$

(5)

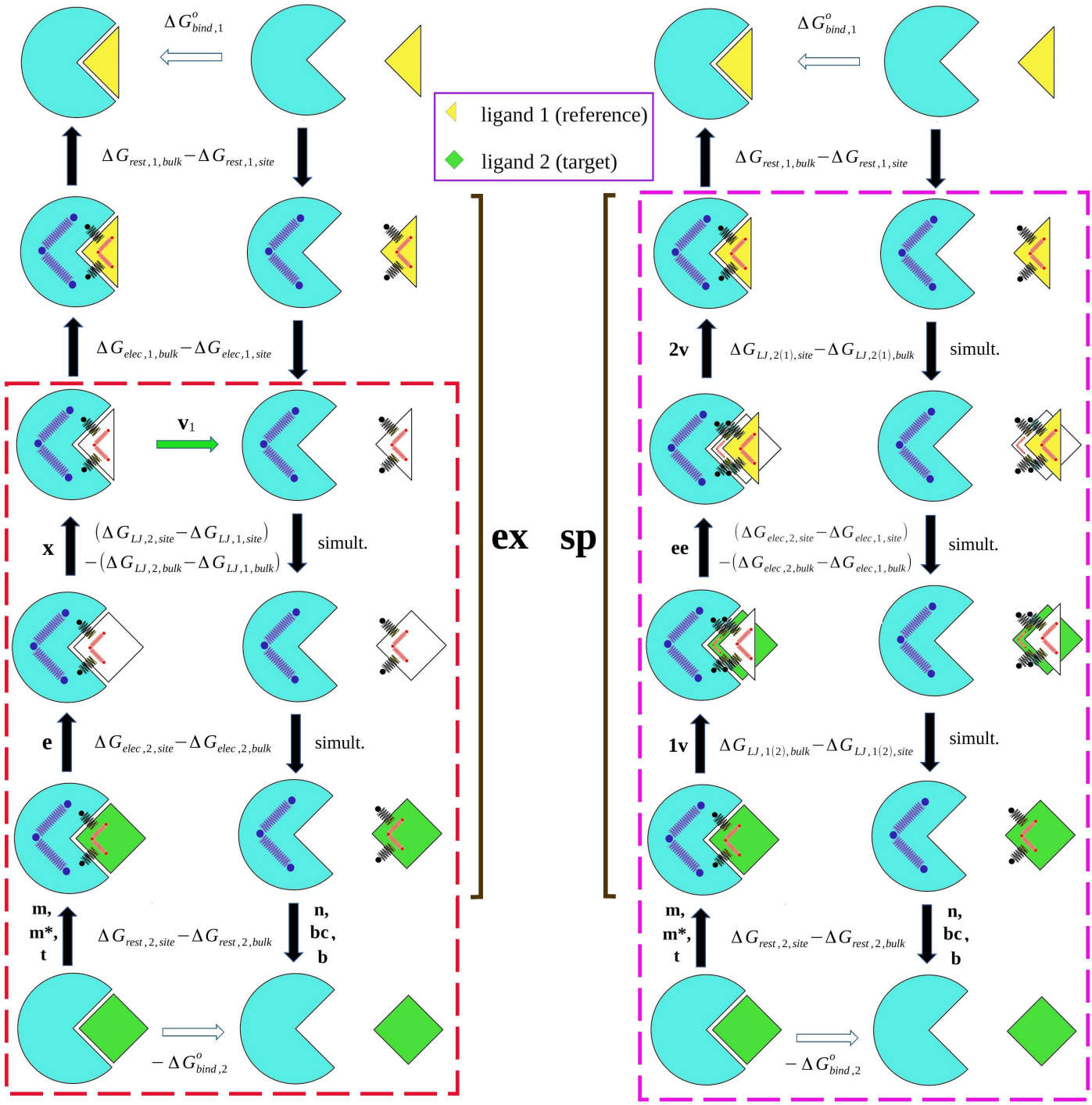


**Figure 1:** Thermodynamic cycles used by BAT.py for RBFE calculations using SepTop. The calculated free energy terms are denoted by the solid black arrows.

When performing RBFE calculations using the **x** component, the free energy components that will be computed for a given target molecule are identified by the solid black arrows inside the dashed red rectangle in 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:

$$-\Delta G_{BAT-x1\to2}=\left(\Delta G_{rest,2,site}-\Delta G_{rest,2,bulk}\right)+\left(\Delta G_{elec,2,site}-\Delta G_{elec,2,bulk}\right)+\left(\Delta G_{LJ,2,site}-\Delta G_{LJ,1,site}\right)-\left(\Delta G_{LJ,2,bulk}-\Delta G_{LJ,1,bulk}\right) \quad (6)$$

This result differs from the relative binding free energy between the two ligands (Eq. 4) by an offset term $\Delta G_0$ that only involves the reference molecule, and consequently is the same for all targets:

$$\Delta G_{BAT-x1\to2}=\left(\Delta G^{o}_{bind,2}-\Delta G^{o}_{bind,1}\right)+\Delta G_0 \quad (7)$$

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

That way, even though the $\Delta G_{BAT-x1\to2}$ 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:

$$\Delta G^{o}_{bind,3}-\Delta G^{o}_{bind,2}=\Delta G_{BAT1\to3}-\Delta G_{BAT1\to2} \quad , \quad (9)$$

and so the 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 ( $\Delta G_{BAT1\to2}$ ) can be subtracted of the result obtained from the first ( $\Delta G_{BAT1\to1}$ ) , which is the reference molecule being transformed into itself. Note that the latter should *not* be zero, due to the presence of the $\Delta G_0$ term.

### 2.3 The 'ex' component

Looking at Eq. 4 and Fig. 1, if we remove the restraint free energies at the end points, we are left with the free energy of exchanging the reference and target ligands, both fully restrained, between the bulk solvent and the protein binding site:

$$-\Delta G^{r}_{1\to2}=\left(\Delta G_{LJ,2,site}-\Delta G_{LJ,1,site}\right)-\left(\Delta G_{LJ,2,bulk}-\Delta G_{LJ,1,bulk}\right)+\left(\Delta G_{elec,2,site}-\Delta G_{elec,1,site}\right)-\left(\Delta G_{elc,2,bulk}-\Delta G_{elc,1,bulk}\right) \quad (10)$$

The **ex** component will compute the value of $\Delta G^{r}_{1\to2}$ from Eq.10 in a single step, decoupling and recoupling the two ligands' LJ and electrostatic interactions simultaneously. The result for a chosen target ligand, when using the **ex** component with restraints, will be the one from Eq. 6 with the addition of the $(\Delta G_{elec,1,bulk} - \Delta G_{elec,1,site})$ term, which is now absent from the offset term $\Delta G_0$ (Eq. 8).

$$-\Delta G_{BAT-\mathbf{ex}\,1\to 2} = \left(\Delta G_{rest,2,site} - \Delta G_{rest,2,bulk}\right) + \left(\Delta G_{elec,2,site} - \Delta G_{elec,2,bulk}\right) + \left(\Delta G_{LJ,2,site} - \Delta G_{LJ,1,site}\right) - \left(\Delta G_{LJ,2,bulk} - \Delta G_{LJ,1,bulk}\right) + \left(\Delta G_{elec,1,bulk} - \Delta G_{elec,1,site}\right) \quad (11)$$

The relations from Eq. 9 remain valid when comparing multiple targets, as long as the calculation method and the reference molecule is the same for all of them. The free energies associated with the application and release of restraints, $\Delta G_{rest,site}$ and $\Delta G_{rest,bulk}$, as with the **x** component above, are only obtained for the target molecule and using the same procedure from the ABFE calculations.

In order to compute the **ex** component, we apply soft-core potentials to both the LJ and the electrostatic energy expressions from the appearing and disappearing atoms. The α and β constants, which respectively scale the degree of "softness" from the LJ and electrostatic potentials, have optimized default values. These values can be changed by the user, by editing the templates for the simulation input files from AMBER (./BAT/amber_files folder) or OpenMM (./BAT/lib/ folder), which might require some degree of knowledge of these MD engines.

### 2.4 The 'sp' component

This component calculates the same quantity as the **ex** one above (Eq. 10), but using a different path, as shown in the right side of Fig. 1. It is essentially the same cycle from Ref. [5], but using the SDR method and with the restraint free energies calculated separately from the ligands Lennard-Jones interactions. When using the **sp** component, the associated equation for the exchange free energy of the fully restrained ligands, between binding site and bulk, is given by:

$$-\Delta G^{r}_{1\to 2} = \left(\Delta G_{LJ,1(2),bulk} - \Delta G_{LJ,1(2),site}\right) + \left(\Delta G_{elec',2,site} - \Delta G_{elec',1,site} - \Delta G_{elec',2,bulk} + \Delta G_{elec',1,bulk}\right) + \left(\Delta G_{LJ,2(1),site} - \Delta G_{LJ,2(1),bulk}\right) \quad (12)$$

keeping in mind that the *elec'* and *LJ* indices denote the *decoupling* of the electrostatic and Lennard-Jones interactions, respectively. The number in parenthesis indicate the presence of ligand 1 or 2 fully coupled to the rest of the system, during the LJ decoupling and recoupling of the other molecule. The electrostatic terms are not the same ones from Eq. 11, due to the presence of the reference and target ligands at all times during these transformations.

The three expressions on the right side of Eq. 12 will be the three sub-components of the **sp** component, respectively from left to right: **1v**, **ee** and **2v**, also shown in the cycle of Fig. 1. The free energy result for a chosen target ligand, when using the **sp** component with restraints, will be the same as when using the **ex** component (Eq. 11):

$$-\Delta G_{BAT-\mathbf{ex}\,1\to 2} = -\Delta G_{BAT-\mathbf{sp}\,1\to 2} \quad , \quad (13)$$

This is shown in Fig. 1 by the solid black arrows inside the purple dashed rectangle, with the relations from Eq. 9 also applicable here. In section 3.2 we give more details on how to interpret the results for RBFE calculations.

# 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`) or `ligand_list` ("rank" option for `calc_type`). 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**, **ex** and **sp** components, which are the ones 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**, **ex** and **sp** transformations 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 is usually the case with the "dock" and "rank" options for `calc_type`. In addition, the `bb_equil` option should always be set to "yes", in order to retain the protein restrained state during equilibration of all target systems (and their reference).

## 3.2 Interpreting the results

Below we show a set of RBFE results using the **x** component with restraints, 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                 Merged components SDR method                 Merged components SDR method
-------------------------------------------- -------------------------------------------- --------------------------------------------

Component            Free Energy; Sigma      Component            Free Energy; Sigma      Component            Free Energy; Sigma

Attach all;             28.16;    0.40       Attach all;             27.27;    0.39       Attach all;             27.34;    0.61
Electrostatic (TI);     -0.87;    0.33       Electrostatic (TI);      3.40;    0.33       Electrostatic (TI);     -1.69;    0.22
LJ exchange (TI);        4.90;    0.13       LJ exchange (TI);        5.93;    0.57       LJ exchange (TI);        5.80;    1.03
Release all;          -37.20;    0.71       Release all;          -35.82;    0.64       Release all;          -35.74;    0.47

Relative free energy;    5.01;    0.88       Relative free energy;   -0.78;    0.99       Relative free energy;    4.29;    1.30

-------------------------------------------- -------------------------------------------- --------------------------------------------
```

pose 1

```
-----------------------------------------------
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 2

```
-----------------------------------------------
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 3

pose 4

pose 5

The relative free energies listed correspond to the $-\Delta G_{BAT-x1\to2}$ values from Eq. 6, 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 using the **x** component with restraints 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):

$$\Delta G_{BAT-x1\to2}-\left(\Delta G_{LJ,1,site}-\Delta G_{LJ,1,bulk}\right)=\Delta G^{o}_{bind,2} \tag{14}$$

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.

The same definitions described above also apply to the RBFE calculations using the **sp** ou **ex** components with restraints, with the only difference being that the result will have the value shown in Eqs. 11 and 13. The comparison between multiple target ligands using Eq. 9 remains valid, but the relation to their ABFE values now also includes the reference molecule electrostatic free energy terms:

$$\Delta G_{BAT-x1\to2}-\left(\Delta G_{LJ,1,site}-\Delta G_{LJ,1,bulk}\right)-\left(\Delta G_{LJ,elec,site}-\Delta G_{elec,1,bulk}\right)=\Delta G^{o}_{bind,2} \tag{15}$$

, which can be verified in the thermodynamic cycle from Fig. 1.

### 3.3 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 RBFE calculations using the **x** component, similar to the ones from Ref. [1], with the restraint scheme and the free energy components needed for each.

- (**m,e,x,n**): 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*,e,x,b,c**): 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.

- (**t,e,x,b**): 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.

These same definitions also apply the the **ex** and **sp** components, but in the latter case three transformations will be performed at each **sp** window (**1v**, **ee** and **2v**), and in the former only a single one is performed (**ex**). The simulation boxes for the application and release of restraints would remain the same for all cases, but the number of SDR transformations would show an increase by one for the **sp** component and a decrease by one for **ex**.

Therefore, even though the cycle from Figure 1 has many steps and transformations, in practice the RBFE calculations from BAT need as little as two free energy components to obtain the relative free energy of a given binding pose or ligand. That would be the case when using the **ex** component with only ligand TR restraints, which corresponds to a (**t,ex,b**) calculation using the notation above.

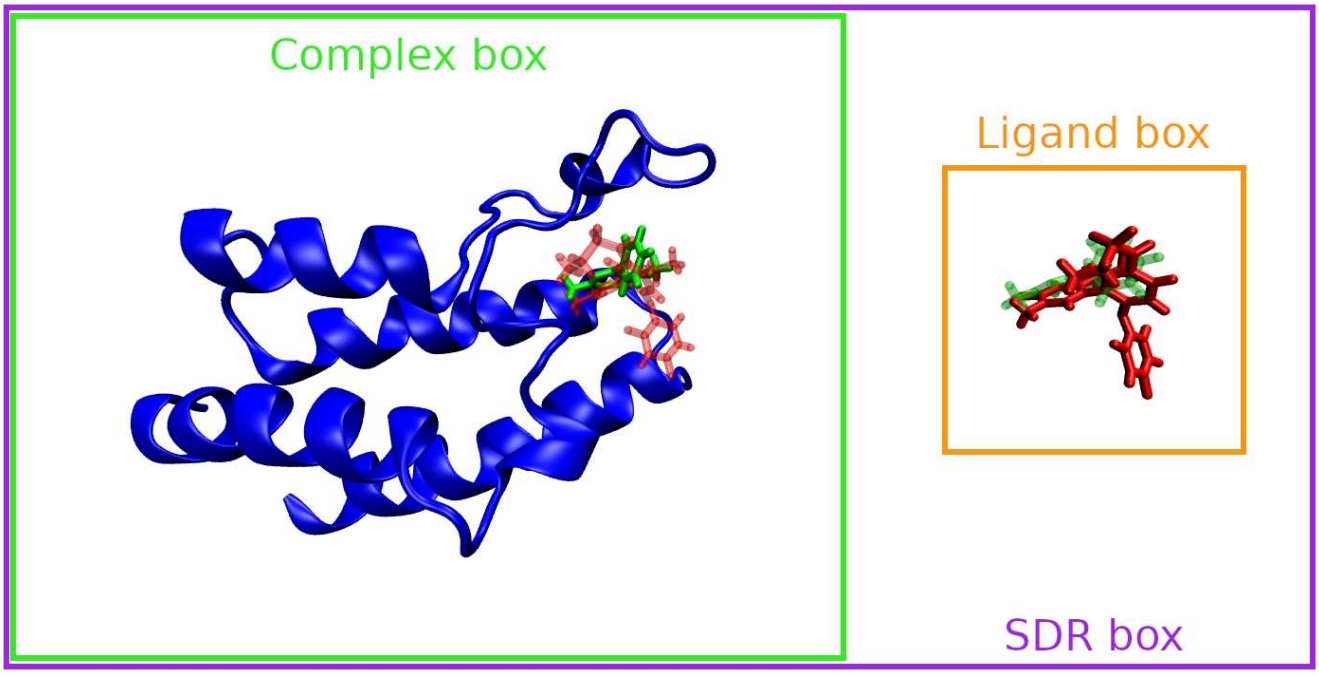


**Figure 2:** The three types of simulation boxes used for the BAT relative calculations, with the reference ligand in green and the target in red. The complex box (green) and the ligand box (orange) only have the target ligand in them, and so does the SDR box (purple) for the e component.

## 4. How to run the RBFE calculations

In order to run the RBFE calculations with BAT, one can select either the “relative”, “relative-ex” or “relative-sp” options for the `fe_type` BAT input variable. The first uses the **e** and **x** components in addition to all conformational and TR restraints, which we identify as an (**m,e,x,n**) calculation. The second and third are also calculations with all restraints, but using the **ex** and **sp** components, respectively, which we can also identify as (**m,ex,n**) and (**m,sp,n**). One can also choose the “custom” option for `fe_type` and select the components individually as shown in Section 3.3, always using the option “exchange” for the `dec_method` variable. The `calc_type` options “dock” and “rank” remain available for relative calculations, as well all the other variables used for ABFE. The software choice between OpenMM and AMBER is also possible with RBFE, except for the **sp** component, which is currently available only for OpenMM.

When selecting the free energy components used for the calculations, it is also necessary to set the number of steps (AMBER) or iterations (OpenMM) for each free energy component simulation window, which for AMBER is done using the `[component]_steps1` and `[component]_steps2` variables, and for OpenMM using the `[component]_itera1` and `[component]_itera2` variables, in addition to `itera_steps` (see User Guide). Note that for the **sp** component three simulations are performed for each **sp** window (**1v**, **ee** and **2v** sub-components), with the values of `sp_itera1` and `sp_itera2` thus being multiplied by three when computing the total simulation time for the calculation. This is already accounted for when BAT includes this time in the Results file.

The folder ./BAT/example-input-files contains a few examples of BAT input files used for relative calculations, including the (**m*,e,x,b,c**) and (**t,e,x,b**) options described above, and also for the **sp** and **ex** components. 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 3.2.

## 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.