Fast and efficient BAT analysis on docked poses

Germano Heinzelmann - 12/2021

1. Overview

In the present report we analyze the possibility of drastically reducing the computational cost of the binding free energy calculations performed by BAT to a set of docked poses. We will apply the calculations to one of the protein systems from [1], which has the ligand from the 5uf0 crystal structure docked to the 5uez receptor in five different poses. This is also the sample system used in the tutorial from the BAT distribution at https://github.com/Gheinzelmann/BAT.py, so the calculations shown here can be reproduced in a fully automated way.

A total of three protocols were carried out to obtain the binding free energies (BFE) of the docked poses, which are summarized in Table I. The first is the double decoupling procedure using the parameters from the BAT.py tutorial, identified simply as “DD”, which uses a total of 1164.8 nanoseconds of simulations for a single pose, the same as Ref [1]. The second protocol also uses the DD procedure, but with a reduced number of windows for the restraints and decoupling procedures, as well as shorter simulation times for each window. It will be called “short DD”, with the total simulation time for a single calculation being 248.4 ns. The third protocol uses the simultaneous decoupling and recoupling method, with merged components for the attachment and release of restraints, which are called **m** and **n**. This last approach will be called “merged SDR”, using a total simulation time of 284 ns.

Table I: Components used on each protocol, the number of windows for the restraint and decoupling components, and the total simulation time for a single calculation on a given pose.

Protocol	Restraint components	Restraint windows (MBAR)	Decoupling components	Decoupling windows (TI-GQ)	Total simulation time
DD	**a**, **l, t, c, r**	16	**e, v, w, f**	12	1164.8 ns
Short DD	**a**, **l, t, c, r**	10	**e, v, w, f**	7	248.4 ns
Merged SDR	**m, n**	10	**e, v**	12	284.0 ns

2. Results

Table II shows the BFE results for the poses in which the protocols were tested. Pose 2 was not assessed because in this system the ligand left the binding site during the equilibration procedure. Pose 0 started with a RMSD of 5.36 Å in the docked system [1], but moved during equilibration, with a final RMSD of 1.21 Å. Thus, from the four poses tested, three of them had RMSDs relative to the 5uf0 structure under 2.00 Å, considered correct, and one (pose 3) had a larger RMSD of 4.86 Å, considered incorrect.

In all three protocols the incorrect pose had the weakest binding from the set, and thus BAT was able to identify the correct poses through the calculated binding free energies. The differences

between the results using the three protocols were inside the error margins, except for pose 1, which showed a difference of 1.92 kcal/mol between the minimum and maximum BFE values.

Table II: RMSDs after equilibration for each pose, the binding free energies (BFE) for each pose using the three protocols, and the maximum range of BFE for each pose using the three protocols.

Docked pose	RMSD after equilibration	BFE DD	BFE Short DD	BFE Merged SDR	BFE Range
Pose 0	1.21 Å	7.13 (1.13)	7.04 (1.06)	7.42 (0.76)	0.38
Pose 1	0.37 Å	5.79 (0.84)	7.71 (0.99)	7.55 (0.79)	1.92
Pose 3	4.86 Å	5.38 (0.70)	4.00 (1.46)	5.56 (1.14)	1.56
Pose 4	0.65 Å	6.07 (0.60)	5.57 (1.16)	6.32 (1.30)	0.75

Due to the larger discrepancy above, we provide a more detailed comparison between the three methods for pose 1. This is shown in Table III below, with the free energy value of each component. Also displayed are the sum of the components of the DD method that are combined in the merged SDR approach, in order to provide a direct comparison between them. The restraint components show good agreement for all three protocols, either separated or combined. The same goes for the decoupling components, except for a discrepancy in the **v** component between the DD and short DD approaches, with a difference of nearly 2 kcal/mol (11.34 vs 13.57 kcal/mol).

Table III: Detailed results for pose 1, showing the calculated free energy for each component using the three protocols. The “Combined DD” column shows the sum of the components from the DD method (ΔG DD column) that correspond to the merged SDR components, for a direct comparison between them.

Component DD	ΔG DD	ΔG short DD	Combined DD	Component Merged SDR	ΔG merged SDR
a	15.46 (0.11)	15.20 (0.34)	26.32 (0.14)	**m**	26.77 (0.35)
l	7.51 (0.06)	7.64 (0.09)			
t	3.35 (0.07)	3.14 (0.13)			
e	-7.47 (0.27)	-7.32 (0.23)	3.84 (0.30)	**e**	4.09 (0.29)
f	11.31 (0.13)	11.14 (0.28)			
v	11.34 (0.67)	13.57 (0.28)	12.05 (0.72)	**v**	12.75 (0.61)
w	0.71 (0.27)	0.56 (0.65)			
b	-11.75	-11.75	-36.42 (0.28)	**n**	-36.07 (0.19)
c	-8.00 (0.08)	-8.14 (0.18)			
r	-16.67 (0.27)	-16.34 (0.43)			
Final BFE	5.79 (0.84)	7.71 (0.99)	5.79 (0.84)	**Final BFE**	7.55 (0.79)

3. Discussion

The results presented here show that it is possible to reduce the simulation time of the absolute binding free energy calculations (ABFE) performed by BAT by a factor of 4- to 5- fold, and still obtain consistent and reliable results. Considering an average performance of a single GTX 1070 card of around 200 ns/day of simulations [1], a full ABFE calculation can be carried out in a single card in a little more than one day. Even though we have not tested it with BAT yet, a newer RTX

3090 card provides a 3-fold performance increase over the GTX 1070, according to the AMBER benchmarks. Thus, considering an RTX 3090 card running at an average of 600 ns/day, a full evaluation of a ligand with five docked poses would take around 2 days using a single card.

We should note that this reduction in the simulation times was just a first attempt without a deep investigation into the fluctuations of each free energy component. Thus, the times can still be optimized, to prioritize the decoupling components over the restraint ones for example, which show lower fluctuations according to Table III. A detailed analysis on a particular system of interest can also shorten the simulation times without a significant loss in precision and accuracy.

4. Reference

[1] 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.