Ligand Binding Thermodynamic Cycles: Hysteresis, LWHAM, and the Overlapping States Matrix (Supporting Information)

Di Cui*, Bin W. Zhang*, Zhiqiang Tan*, Ronald Levy*∗

E-mail: di.cui@temple.edu,bin.w.zhang@temple.edu,ztan@stat.rutgers.edu,ronlevy@temple.edu

Free Energy Change along Each Edge in a Triangular Cycle

In the supporting material, we show another example of FEP calculations involving three congeneric ligands forming a triangular perturbation map of closed thermodynamic cycle. The three ligands also belong to quinoline-based ALLINIs which bind to HIV-1 integrase. We consider the bound state for the perturbations which is more challenging to sample. The estimated free energy change along each edge in the triangular cycle is summarized in Figure 1. The black and blue values represent the free energy estimations using the data from independent parallel simulations for each λ-states based on BAR and UWHAM, respectively. The hysteresis of cycle closure in the triangular cycle is 1.39 kcal/mol as the free energy change along each edge is evaluated based on BAR; while there is no hysteresis effect when UWHAM is applied. The orange values denote the benchmark results from

Hamiltonian replica exchange simulations. We observe that along the edge of ligand 3 to ligand 1, it shows the largest discrepancy for the estimation based on BAR compared to the benchmark with a value of 1.83 kcal/mol. This would be improved by using UWHAM all the data in the cycle, leading to the drop of discrepancy compared to the benchmark with a value of 1.07 kcal/mol.

Gaussian Statistics Analysis Method

An approach to analyzing hysteresis in the FEP thermodynamic cycles has been proposed by Wang et al.[1] based on Gaussian statistics, which is discussed here for a cycle with four nodes. The basic assumption is that the FEP estimate $\triangle F_i$ along edge i is Gaussian distributed, independently of other edges, with mean $\triangle \bar{F}_i$ and standard deviation σ_i for $i \in \{12, 23, 34, 41\}$. That is, the probability density of $\triangle F_i$ is

$$\rho_i = \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{(\triangle F_i - \triangle \bar{F}_i)^2}{2\sigma_i^2}). \tag{1}$$

There are two distinct methods introduced. In the first method, the standard deviations σ_i are assumed to be known. Maximization of the overall likelihood

$$L = \frac{1}{(2\pi)^2} \exp(-\sum_i \log \sigma_i - \sum_i \frac{(\triangle F_i - \triangle \bar{F}_i)^2}{2\sigma_i^2}), \tag{2}$$

subject to the constraint of cycle closure, $\sum_i \triangle \bar{F}_i = 0$, leads to the estimator

$$\triangle \hat{F}_i = \triangle F_i - \frac{\sigma_i^2 \Delta}{\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{34}^2 + \sigma_{41}^2}, \tag{3}$$

where the hysteresis $\Delta = \triangle F_{12} + \triangle F_{23} + \triangle F_{34} + \triangle F_{41}$. A limitation of this estimator Eq.(3) is that the adjustments from $\triangle F_i$ to $\triangle \hat{F}_i$ for all i are in the same direction as

$-\Delta$. Nevertheless, if the assumption is valid that Δ is Gaussian with mean 0 and standard deviation $s = (\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{34}^2 + \sigma_{41}^2)^{1/2}$, then the cycle closure requirement can be used to assess convergence of the calculations: if $|\Delta| > 2s$, then it is unlikely (P-value < 0.05) that the calculations are converged.

In the second method, the standard deviations σ_i are unknown, but assumed to be equal to a common value σ. Then the estimator Eq.(3) reduces to

$$\triangle \hat{F}_i = \triangle F_i - \frac{\Delta}{4}, \tag{4}$$

Note that the adjustments from $\triangle F_i$ to $\triangle \hat{F}_i$ for all i are all the same, regardless of potentially different degrees of equilibration along different edges. Now the assumption that Δ is Gaussian with mean 0 and standard deviation $s = 2\sigma$, can be used with cycle closure to derive an estimator of σ as $\hat{\sigma} = \Delta/2$. This estimator of standard deviation of each $\triangle F_i$ is the same for all edges, irrespective of their degrees of equilibration. This argument also necessarily implies that $\Delta = 2\hat{\sigma} \approx s = 2\sigma$, corresponding to a P-value= 0.32.

Block Bootstrap

We compared the standard error for the free energy change along each edge calculated based on the fractional replication method as shown in Table 1 and the block bootstrap method as shown in Table 2. For the block bootstrap method, the original data set was first split into non-overlapping n blocks. A new data set was reconstructed by randomly resampling the n blocks of original data set n times with replacement. We call such a new data set which has the same size with original data set as one resample. In this study, we test different number of blocks and different number of resamples in the block bootstrap method.

As seen from Tables 1 and 2, the standard errors estimated using fractional replication

or block bootstrap stabilize quickly as the number of resamples increases. However, the standard errors by block bootstrap decreases considerably as the number of blocks ranges from 4 to 50 or equivalently the block length ranges from 1250 down to 100, with the data size $n = 5000$ for each state. The optimal block size from statistical theory based on stationarity is of order $n^{1/3} \approx 17$, corresponding to the number of blocks equal to 292, which would lead to even smaller standard errors than with the number of blocks equal to 50.

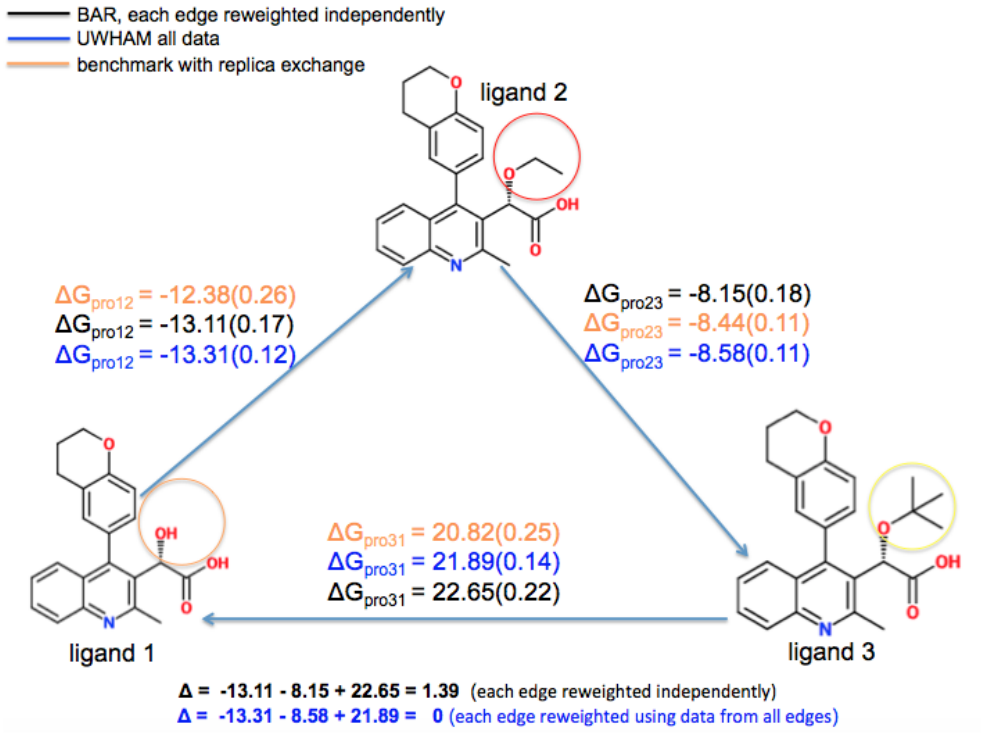

Figure 1: Comparison of free energy change along each edge in the triangular system in the bound state estimated from BAR and UWHAM

Table 1: Standard error of free energy change along each edge in the bound state estimated from UWHAM including all the data. The standard error is calculated based on fractional replication method.

No. of Combinations	SE of ΔF_{pro12}	SE of ΔF_{pro23}	SE of ΔF_{pro34}	SE of ΔF_{pro41}
25	0.047	0.089	0.093	0.059
50	0.046	0.087	0.105	0.062
100	0.048	0.087	0.100	0.063
200	0.044	0.085	0.097	0.058
400	0.044	0.084	0.091	0.058
600	0.043	0.082	0.090	0.056

Table 2: Standard error of free energy change along each edge in the bound state estimated from UWHAM including all the data. The standard error is calculated based on block bootstrap.

No. of Blocks	No. of Resamples	SE of ΔF_{pro12}	SE of ΔF_{pro23}	SE of ΔF_{pro34}	SE of ΔF_{pro41}
4	50	0.043	0.084	0.079	0.050
4	100	0.046	0.076	0.080	0.050
4	200	0.045	0.082	0.093	0.057
4	400	0.046	0.091	0.100	0.059
4	600	0.042	0.087	0.095	0.054
10	50	0.033	0.072	0.078	0.041
10	100	0.034	0.073	0.080	0.042
10	200	0.031	0.068	0.076	0.044
10	400	0.033	0.072	0.083	0.042
10	600	0.035	0.066	0.073	0.044
20	50	0.024	0.049	0.060	0.032
20	100	0.026	0.050	0.059	0.032
20	200	0.029	0.050	0.055	0.033
20	400	0.026	0.050	0.055	0.029
20	600	0.026	0.049	0.054	0.034
40	50	0.023	0.037	0.039	0.024
40	100	0.023	0.040	0.042	0.025
40	200	0.027	0.038	0.042	0.028
40	400	0.021	0.040	0.044	0.029
40	600	0.022	0.040	0.043	0.028
50	50	0.022	0.031	0.035	0.026
50	100	0.021	0.036	0.038	0.025
50	200	0.024	0.038	0.040	0.026
50	400	0.021	0.037	0.040	0.026
50	600	0.022	0.038	0.039	0.025

References

(1) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282–1293.