

CL-FEP: An End-State Free Energy Perturbation Approach

Yasser B. Ruiz-Blanco * and Elsa Sanchez-Garcia *

Computational Biochemistry, Center of Medical Biotechnology, University of Duisburg-Essen,
Germany

Supporting Information

Contents

Table S1. Thermodynamic and convergence parameters obtained for all the checkpoints in the CD – PRO system using the second order cumulant estimator (C2).....	S2
Table S2. Thermodynamic and convergence parameters obtained for all the checkpoints in the 18C6 – K ⁺ system using the second order cumulant estimator (C2).	S3
Table S3. Thermodynamic and convergence parameters obtained for all the checkpoints in the SH3 – Ac-APSYSPPPPP system using the second order cumulant estimator (C2).	S4
Table S4. Thermodynamic and convergence parameters obtained for all the checkpoints in the SH2 – Ac-pYEEI system using the second order cumulant estimator (C2).	S5
Finite-size effect correction	S6
Enthalpy estimation	S7

Table S1. Thermodynamic and convergence parameters obtained for all the checkpoints in the CD – PRO system using the second order cumulant estimator (C2).

Point	$\Delta G_{\text{C}2}$	ISR	Convergence
(%)	kcal/mol	%	
1 (30)	-9.5 ± 0.70	13.19	WARNING3
2 (35)	-8.5 ± 0.67	19.18	WARNING3
3 (40)	-8.2 ± 0.67	26.69	WARNING3
4 (45)	-8.2 ± 0.64	19.20	WARNING3
5 (50)	-7.8 ± 0.65	14.91	WARNING3
6 (55)	-6.0 ± 0.63	45.87	WARNING3
7 (60)	-6.2 ± 0.59	31.24	OK
8 (65)	-5.8 ± 0.58	19.22	OK
9 (70)	-5.8 ± 0.54	19.29	OK
10 (75)	-5.4 ± 0.51	28.57	OK
11 (80)	-5.0 ± 0.49	24.62	OK
12 (85)	-4.5 ± 0.52	30.92	OK
13 (90)	-4.9 ± 0.46	43.04*	OK
14 (95)	-5.2 ± 0.44	19.40	OK
15 (100)	-4.6 ± 0.50	21.70	OK
Average	-5.2 ± 0.51		

*Largest value of the importance sampling ratio among the converged checkpoints

Table S2. Thermodynamic and convergence parameters obtained for all the checkpoints in the 18C6 – K⁺ system using the second order cumulant estimator (C2).

Point	ΔG_{c2}	ISR	Convergence
(%)	kcal/mol	%	
1 (30)	-1.6 ± 0.70	20.71	WARNING3
2 (35)	-2.5 ± 0.62	25.70	OK
3 (40)	-2.4 ± 0.56	32.03	OK
4 (45)	-2.2 ± 0.57	13.77	OK
5 (50)	-3.3 ± 0.57	26.68	OK
6 (55)	-3.5 ± 0.51	20.70	OK
7 (60)	-3.8 ± 0.49	34.61	OK
8 (65)	-3.5 ± 0.46	26.89	OK
9 (70)	-3.8 ± 0.42	81.12*	OK
10 (75)	-3.7 ± 0.42	16.59	OK
11 (80)	-3.8 ± 0.42	24.56	OK
12 (85)	-3.6 ± 0.40	18.10	OK
13 (90)	-3.7 ± 0.38	28.94	OK
14 (95)	-3.6 ± 0.38	20.64	OK
15 (100)	-3.7 ± 0.38	26.36	OK
Average	-3.4 ± 0.46		

*Largest value of the importance sampling ratio among the converged checkpoints

Table S3. Thermodynamic and convergence parameters obtained for all the checkpoints in the SH3 – Ac-APSYSPPPPP system using the second order cumulant estimator (C2).

Point	ΔG_{C2}	ISR	Convergence
(%)	kcal/mol	%	
1 (30)	-10.4 ± 0.71	9.85	WARNING3
2 (35)	-9.9 ± 0.76	20.36	WARNING3
3 (40)	-8.2 ± 0.66	18.13	WARNING3
4 (45)	-8.0 ± 0.64	24.00	WARNING3
5 (50)	-7.6 ± 0.62	21.14	WARNING3
6 (55)	-8.1 ± 0.62	13.29	OK
7 (60)	-6.8 ± 0.62	16.70	WARNING3
8 (65)	-7.0 ± 0.57	17.06	OK
9 (70)	-7.1 ± 0.56	39.51	OK
10 (75)	-7.6 ± 0.52	34.77	OK
11 (80)	-7.2 ± 0.57	17.84	OK
12 (85)	-7.2 ± 0.51	26.77	OK
13 (90)	-7.9 ± 0.48	28.93	OK
14 (95)	-7.7 ± 0.54	44.92*	OK
15 (100)	-8.4 ± 0.45	21.28	OK
average	-7.5 ± 0.53		

*Largest value of the importance sampling ratio among the converged checkpoints

Table S4. Thermodynamic and convergence parameters obtained for all the checkpoints in the SH2 – Ac-pYEEI system using the second order cumulant estimator (C2).

Point	ΔGc₂	ISR	Convergence
(%)	kcal/mol	%	
1 (30)	-15.7 ± 0.93	11.49	WARNING3
2 (35)	-17.6 ± 0.85	14.57	WARNING3
3 (40)	-16.3 ± 0.85	27.11	WARNING3
4 (45)	-11.9 ± 0.73	21.01	WARNING3
5 (50)	-9.7 ± 0.65	15.05	OK
6 (55)	-10.1 ± 0.64	18.72	OK
7 (60)	-10.1 ± 0.64	19.88	WARNING3
8 (65)	-8.3 ± 0.61	21.9	OK
9 (70)	-8.4 ± 0.57	27.47	OK
10 (75)	-6.9 ± 0.62	23.16	WARNING3
11 (80)	-6.2 ± 0.58	26.15	OK
12 (85)	-6.1 ± 0.48	36.25*	OK
13 (90)	-5.9 ± 0.49	25.48	OK
14 (95)	-5.5 ± 0.5	26.18	OK
15 (100)	-5.8 ± 0.49	34.9	OK
average	-7.0 ± 0.54		

*Largest value of the importance sampling ratio among the converged checkpoints

Finite-size effect correction

In CL-FEP, the energy samples of the initial and final states are obtained from independent simulations. Hünenberger et al. discussed in detail several finite-size effects that must be corrected particularly in FEP calculations involving charged species.¹ Our CL-FEP approach, unlike the traditional stratified FEP solution, does not involve annihilation or vanishing processes of charged molecules where the net charge of the simulation box may vary. Then, the finite-size effect contribution from the periodicity-induced net charge interaction,¹ which is necessary for the stratified FEP solution, should not affect our approach.

Nonetheless, a correction is needed for the self-interaction contribution of the lattice. This need arises from the fact that, for instance, the ligand molecule is simulated in the initial state within a box with solvent and the necessary counter ions, while in the final state the ligand is bound to the protein in a probably larger simulation box. Given that the self-interactions of the ligand are a function of the size of the periodic box, the balance between initial and final states will carry an excess of self-interaction energy that must be corrected.

$$\Delta U_{corrected}^* = \Delta U^* + \Delta U_{SI}$$

The self-interaction term in a cubic lattice is formulated as:¹⁻³

$$\Delta U_{SI} = -\frac{\xi_{LS}}{8\pi\epsilon_0} \sum_{i=1}^K q_i^2 \left(\frac{1}{L_i^F} - \frac{1}{L_i^I} \right)$$

Where the dimensionless coefficient $\xi_{LS} \approx -2.8373$ is the cubic lattice-sum Wigner integration constant. The quotient $1/8\pi\epsilon_0$ takes the value $166.031 \text{ kcal } \text{\AA mol}^{-1} \text{ e}^{-2}$, K is the number of charged species and L represents the edge of the cubic simulation box in the initial or final state.

Our code CL-FEP.pl computes automatically this lattice correction and applies it to the potential energy differences between initial and final states. We note that the above formulation of the self-interaction correction is strictly valid only for cubic boxes. Therefore, the code implicitly assumes that the simulation boxes of initial and final states have this geometry. This means that if the simulations were performed with rectangular boxes, the code will still compute the edge of the

boxes (L^X) as the cubic root of the box volume. The computed value of this correction is printed in the final output file for the user's information and it is typically below 1 kcal/mol.

Enthalpy estimation

While CL-FEP is intended to compute the free energy change of a system transformation, the resampled energy differences calculated with Eq. 26 (main text) also provide a direct estimation of the enthalpy change (ΔH), which is converged under the CL-FEP algorithm. Together, the converged free energy change and the corresponding enthalpy complete the thermodynamic profile of the system transformation as the entropic contribution can then be obtained by $T\Delta S = \Delta H - \Delta G$.

References

- (1) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hunenberger, 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* (18), 184103.
- (2) Hünenberger, P. H., Lattice-sum methods for computing electrostatic interactions in molecular simulations. *AIP Conference Proceedings* **1999**, *492* (1), 17-83.
- (3) Hummer, G.; Pratt, L. R.; García, A. E., Free energy of ionic hydration. *J. Phys. Chem.* **1996**, *100* (4), 1206-1215.