

## Supporting Information

# Self-Learning Adaptive Umbrella Sampling Method for the Determination of Free Energy Landscapes in Multiple Dimensions

Wojciech Wojtas-Niziurski<sup>1†</sup>, Yilin Meng<sup>2†</sup>, Benoît Roux<sup>2\*</sup>, Simon Bernèche<sup>1\*</sup>

<sup>1</sup>Swiss Institute of Bioinformatics and Biozentrum, University of Basel,  
Klingelbergstrasse 50/70, CH-4056 Basel, Switzerland

<sup>2</sup>Department of Biochemistry and Molecular Biology, University of Chicago,  
929 E 57<sup>th</sup> Street, Chicago, IL, 60637

<sup>†</sup>Those two authors contributed equally to this work

<sup>\*</sup>Corresponding authors: [simon.berneche@unibas.ch](mailto:simon.berneche@unibas.ch), [roux@uchicago.edu](mailto:roux@uchicago.edu)

## Model System of Fermat Spiral

We focused here on the ability of the described window creation procedure to follow complex pathways. We performed Monte Carlo simulations on an analytical energy function defined as a Fermat spiral:  $r = \pm\theta^{1/2}$  with  $\theta \in \left[0, \frac{9}{4}\pi\right]$ , which was chosen for its non-trivial shape (see Fig. S1). Sampling was performed in Cartesian coordinate with an initial position set to  $[-2, -0.75]$  and the window separation distance to 0.15. Simulation windows were characterized by a biasing harmonic potential with  $k = 6$  kcal/mol•u along both dimensions. Sampling in each window was obtained from a Boltzmann weighted random walk. The WHAM algorithm was applied to calculate the energy landscape from the sampling data. From the initial 9 windows, the automated procedure created new windows in both directions along the spiral function. A total of 562 windows were used to reconstruct the potential function, while the square space in which the Fermat Spiral is inscribed contains 1,156 grid points.

## Model System of Lennard-Jones Particles

### Computational Details

The self-learning umbrella sampling simulations and WHAM with 3 and 6 reaction coordinates were performed to probe a system consisting of Lennard-Jones (LJ) particles. The system consisted of 4 face-centered cubic unit cells of homogeneous LJ particles. The LJ parameters used in our calculations were modified from those of Argon atoms in CHARMM 27 force field<sup>29</sup>, such that  $\epsilon = 2 * \epsilon_{Ar}$  and  $R_{min} = R_{min,Ar}$ . All MD simulations were performed using CHARMM molecular simulation program<sup>33</sup>. Canonical ensemble was applied to this model system. The temperature were controlled by Langevin dynamics and kept at 300 K. No cutoff was applied to non-bonded interactions.

In the first case, we wish to describe the displacement of a given particle, number 166, around which three other particles were removed. The Cartesian coordinates of particle 166 were selected as reaction coordinates to calculate the PMF  $W[x_1, y_1, z_1]$ . A snapshot of the system is illustrated in Figure S2A. All particles except number 166 were restrained to their “lattice” points by a harmonic potential ( $k = 40 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ ). Umbrella windows were generated every  $0.5 \text{ \AA}$  in each direction, within a box  $[-4, 4] \times [-1, 4] \times [-1, 4] \text{ \AA}^3$ . The force constant of the umbrella potential was  $10.0 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . Each window was simulated for 100 ps, with a time step of 1 fs. Free energy landscape expansion procedure was initialized with the  $E_1 = E_2 = 12 \text{ kcal/mol}$ .

In the second case, we wish to characterize the movement of two particles exchanging place in the lattice. The Cartesian coordinates of particle 166 and 170 were chosen as reaction coordinates to calculate the 6D PMF  $W[x_1, y_1, z_1, x_2, y_2, z_2]$ . A string pathway consisting of 22 images was first constructed between two endpoints made by swapping the positions of particles 166 and 170. In order to allow particle 166 and 170 moving inside the system, two particles were removed from the initial configuration. A snapshot of the system used by umbrella sampling was illustrated in Figure S3A. Umbrella windows were generated every  $1.0 \text{ \AA}$  in each dimension, within a box  $[-4, 4] \times [-1, 4] \times [-1, 4] \times [-4, 4] \times [-1, 4] \times [-1, 4] \text{ \AA}^6$ . The force constant of the harmonic umbrella potential was  $5.0 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ . Each window was simulated for 1 ns, with a time step of 1 fs. Free energy landscape expansion procedure was initialized with the  $E_1 = E_2 = 15 \text{ kcal/mol}$ , and  $\Delta_1 = 1.5 \text{ \AA}$ .

## Results and Discussion

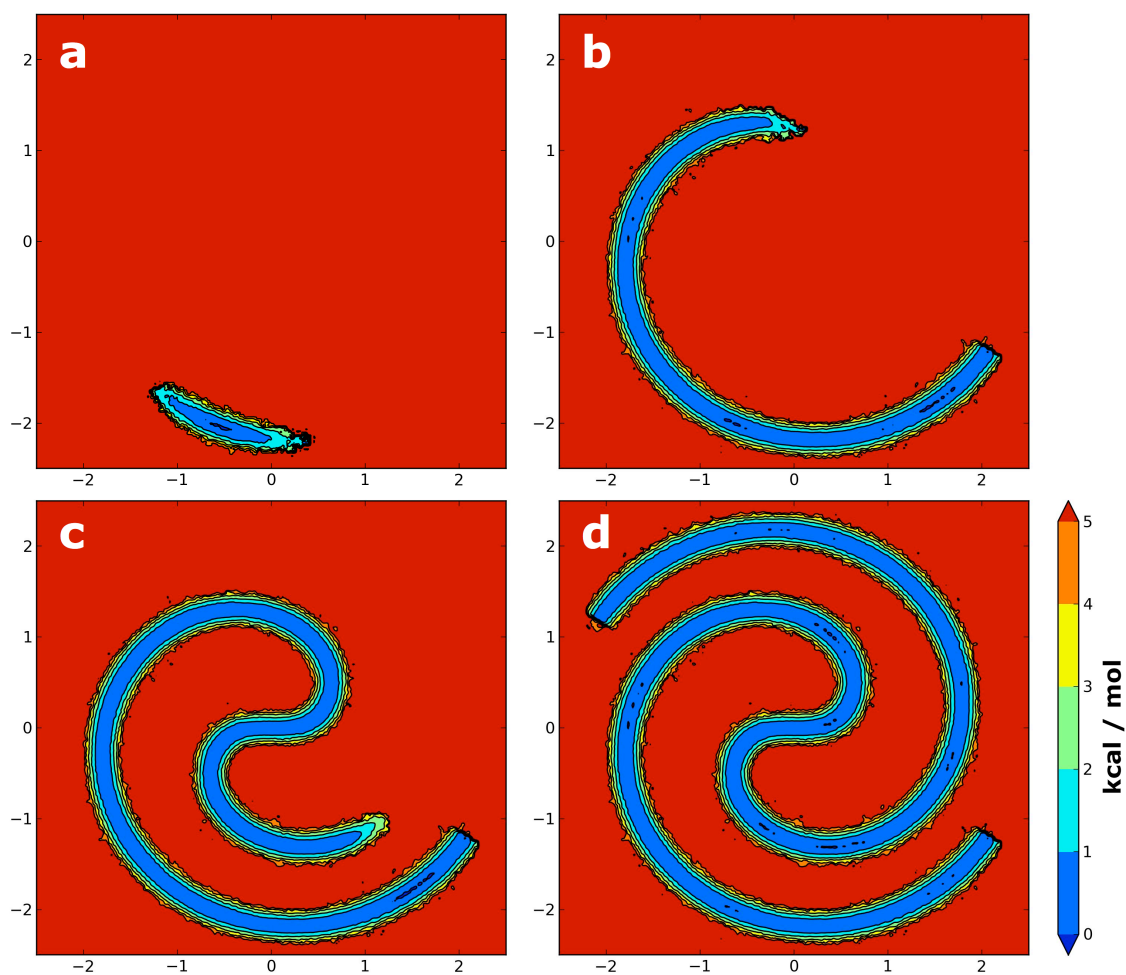
Results generated from the 3D self-learning umbrella sampling are shown in Figure S2. In this case, the free energy landscape was explored starting from only one configuration. Figure S2B shows the cumulative number of windows with respect to cycle index of the self-

learning procedure. The final 3D PMF calculated with WHAM is shown in Figure S2C. Four free energy minima were identified by the self-learning umbrella sampling calculation. Those stable conformations have roughly the same free energy and correspond to particle 166 being at its initial position or at the vacant “lattice” points.

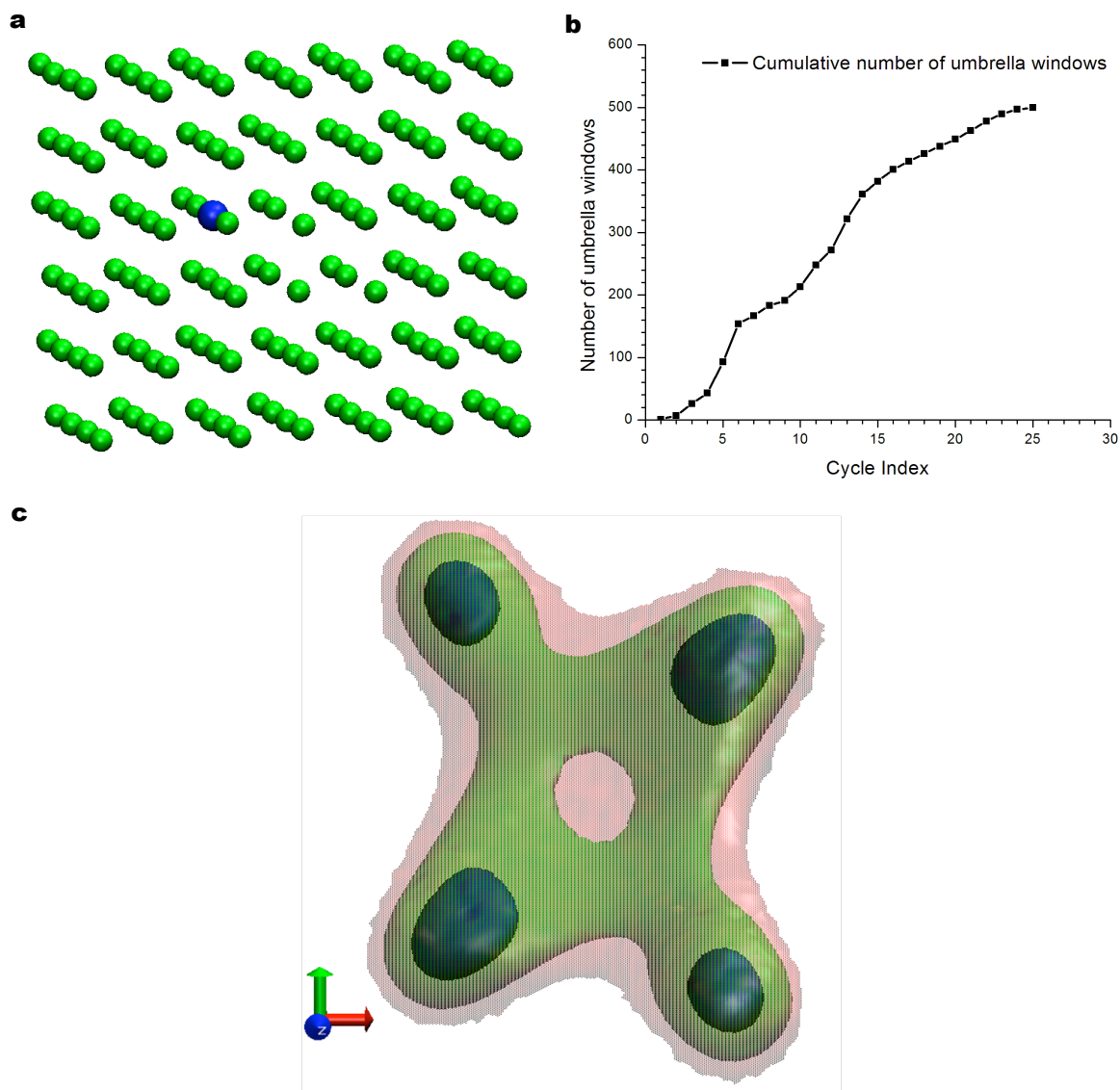
Results generated from the 6D self-learning umbrella sampling are shown in Figure S3. Figure S3B shows the cumulative number of windows with respect to iteration index of the self-learning procedure. By employing a string pathway in combination with the self-learning umbrella sampling approach, 237 windows were eventually needed to characterize the free energy landscape instead of the theoretical 104,976 windows required to occupy the whole 6D space. To facilitate visualization, the 6D PMF was projected onto a 2D landscape (Figure S3C). The two reaction coordinates used here are the distance between particles 166 and 170, and the average of the distances to their initial positions. Three other stable states were found besides the initial and final ones. They represent either one or two particles in the initially vacant “lattice” points. By employing self-learning approach, all essential regions in the free energy landscape were indeed visited. The number of windows that would be required to cover the whole 3D and 6D space, and the number actually used are given in Table S1.

	3D PMF	6D PMF
Full configuration space	2057	104976
Essential configuration space, as delimited by the automated procedure	500	237 (using a predefined string pathway)

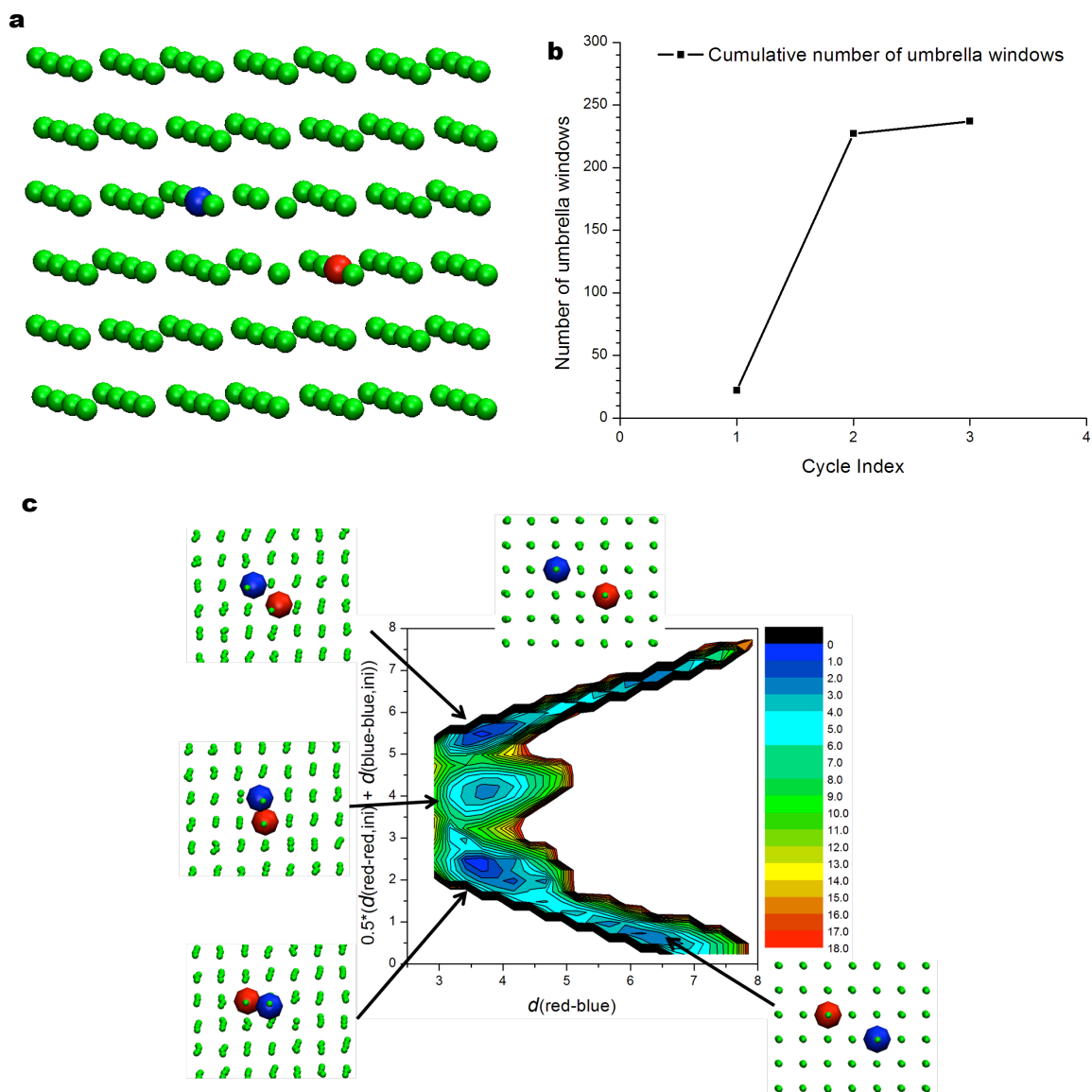
**Table S1.** Number of windows required to calculate a PMF describing the process of moving one (3D) or two (6D) particles in a LJ-particle system.



**Figure S1.** Illustration of the self-learning umbrella sampling approach using Monte Carlo simulations and an energy function defined as a Fermat Spiral. The different panels show the energy landscape at different stages of the calculation: (a) Initial energy landscape generated with 9 windows, (b-c) Intermediate stages with 212 and 352 windows respectively (d) The final landscape obtained from 562 windows. The whole space contains 1,156 grid points.



**Figure S2.** Three-dimension sampling in a system of Lennard-Jones particles. (a) The starting configuration used for the 3D umbrella sampling is shown, with particle 166 colored in blue. (b) The cumulative number of umbrella windows is plotted with respect to cycle index from the self-learning umbrella sampling. (c) The isosurfaces describe the 3D-PMF with the blue, green, and red color representing  $W[x_1, y_1, z_1] = 2, 6,$  and  $10$  kcal/mol, respectively. The four minima in the PMF correspond to particle 166 being at its initial position and at the vacant “lattice” points.



**Figure S3.** Six-dimension sampling initiated with a string pathway. (a) The initial configuration of the system used for the 6D umbrella sampling is shown with particles 166 and 170 colored in blue and red, respectively. (b) The cumulative number of umbrella windows is plotted with respect to cycle index from the self-learning umbrella sampling. (c) A 2D projection of the final 6D-PMF is shown accompanied by representative structures corresponding to free energy wells.