Supporting Information

Finding an Optimal Pathway on a Multidimensional Free-Energy Landscape

Haohao Fu[†,○], Haochuan Chen[†,○], Xin'ao Wang[†], Hao Chai[†], Xueguang Shao[*,†], Wensheng Cai[*,†], and Christophe Chipot[*,‡,§]

[†]Research Center for Analytical Sciences, College of Chemistry, Nankai University, Tianjin Key Laboratory of Biosensing and Molecular Recognition, State Key Laboratory of Medicinal Chemical Biology, Tianjin 300071, China

[‡]Laboratoire International Associé CNRS and University of Illinois at Urbana−Champaign, F-54506 Vandœuvre-lès-Nancy, France

[§]Department of Physics, University of Illinois at Urbana−Champaign, 1110 West Green Street, Urbana, Illinois 61801, United States

[○]Contribute equally

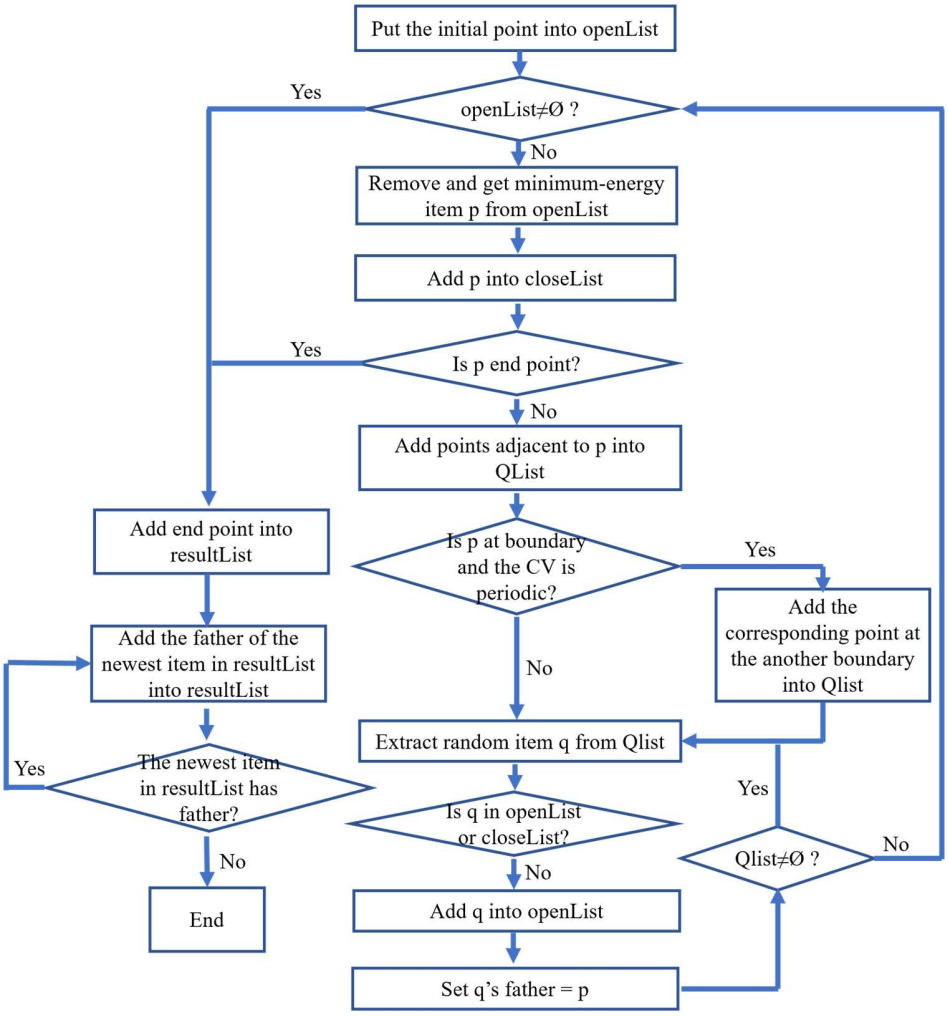

Figure S1. Flow chart of the core algorithm of MULE package.

An abridged proof of the ability of the Dijkstra algorithm to identify the true MFEP

(Continued from the proof in the main text.) Let us note s_t $[t \in (0, n)]$ the transition state, and let us assume that there is only one free-energy barrier along $f(s)$. Hence, from s_0 to s_t, $W[f(s)]$ monotonically increases, and from s_t to, s_n $W[f(s)]$ monotonically decreases. After exploring $f(s_t)$, points adjacent to $f(s_t)$, will be added into $\mathrm{openList}$. Since the Dijkstra algorithm always extracts the point corresponding to the lowest free energy from $\mathrm{openList}$, there is $W[f(s_{t+1})] = min(W[Q]), \forall Q \in$ unexplored points adjacent to $f(s_t)$. Hence, $f(s_i)$ $(i \in [\mathrm{t}, n])$ satisfies the requirement of a steepest-descents path. Let us then consider $f(s_i)$ $\{i \in [0, t)\}$. Using exactly the same proving process as that of the main text, it can be guaranteed that there is no pathway $g(s)$ satisfying $H < I$, where $H = max\{W[f(s_i)]\} = W[f(s_{t-1})]$ and $I = max\{W[g(s_i)]\} = W[g(s_{t-1})]$, $\{i \in [0, t)\}$. Hence, $f(s_i)$ $\{i \in [0, t)\}$ satisfies the requirement of a steepest-descents path. Therefore, $f(s)$ is the union of steepest-descents paths from the saddle points to the minima, that is, the true MFEP.

Comparison between MULE (the Dijkstra algorithm) and the string algorithm

Pros of MULE. The Dijkstra algorithm is parameter-free, and performs consistent, reproducible path searches. In contrast, the string algorithm implemented in MEPSearcher and NAMD has been shown in a number of cases to depend on the initial guess of the path, as well as on tunable parameters, e.g., the number of images along the string. Different settings may lead to distinct results when using this algorithm.[1,2] Moreover, the Dijkstra algorithm can always identify the true MFEP, while the pathway found by the string algorithm is usually an approximation of the MFEP due to the limited number of images.

Cons of MULE. MULE is developed for post treatments, which implies that a complete free-energy landscape must be obtained as a preamble to the identification of the MFEP. If one, however, wishes to perform on-the-fly MFEP searches, MULE, or the Dijkstra algorithm, is nearly serial, while the string algorithm is embarrassingly parallelizable, and, hence, almost surely more computationally efficient.

Other differences. The Dijkstra algorithm was originally designed for finding paths in a discrete graph. Interpolation of the discrete free-energy landscape is, therefore, not required in MULE, though optional if needed. In the string algorithm, however, such an interpolation is almost always required to reparametrize the string.[3] This difference in interpolation between the two algorithms may lead to significantly distinct pathways, as shown in Figure S2.

To highlight the difference between the Dijkstra and the string algorithms, favorable pathways underlying the translocation of a chloride ion through a synthetic membrane channel were identified by the string algorithm implemented in MEPSeacher (https://github.com/chenxin199261/MEPSearcher). Figure S2A showcases a significant difference between the pathways found by MULE and MEPSearcher, arising from the interpolation of the discretized free-energy landscape by the MEPSearcher, as detailed in Figure S3. Figure S2B describes the parameter-dependence of the string algorithm.

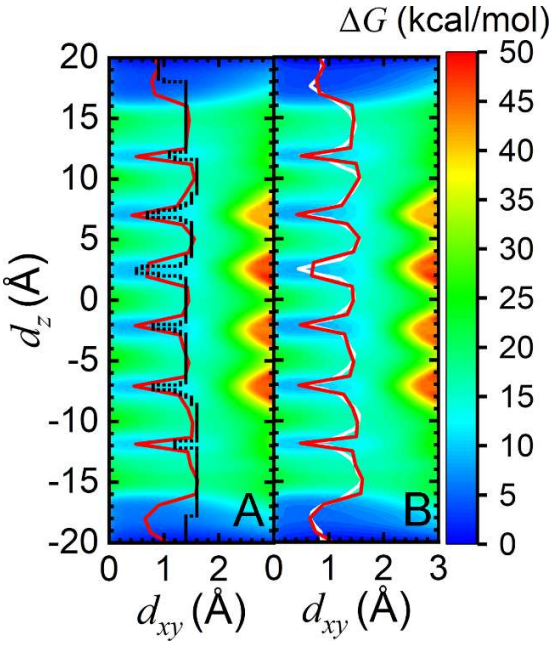

Figure S2. Predicted MFEPs characterizing the translocation of a chloride ion through a transmembrane channel by MULE (black dot), MEPSearcher with 40 beads (red line) and MEPSearcher with 80 beads (white line). The difference between the pathways found by MULE and MEPSearcher is due to the interpolation of the free-energy landscape by the latter, as depicted in Figure S3.

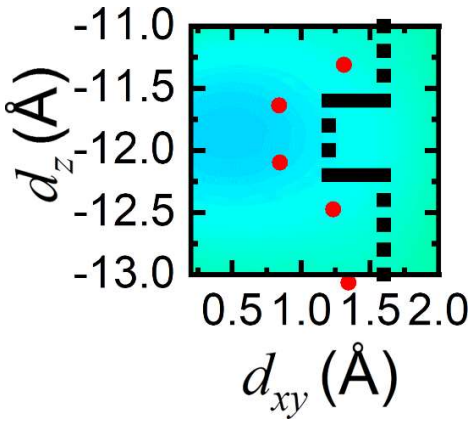

Figure S3. Closeup of Figure S2A highlighting the effect of interpolating the free-energy landscape by MEPSearcher. The contour map describes the interpolated free-energy map

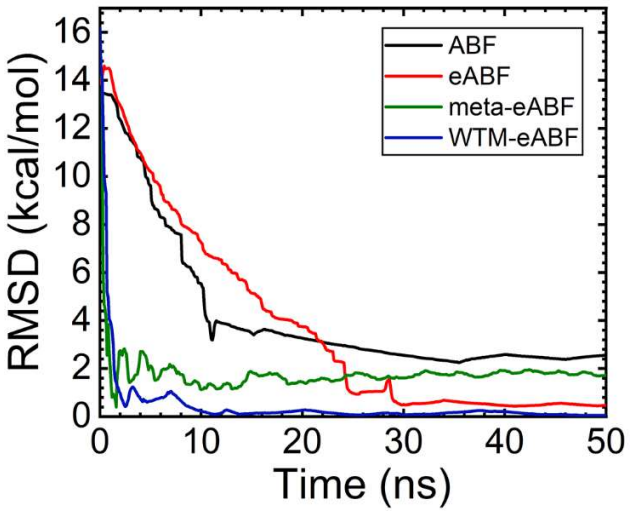

Figure S4. Time evolution of the root-mean-square deviation over the free-energy profiles corresponding to Figure 3. WTM-eABF has the best convergence property among all the methods. The sampling of meta-eABF is almost the same as fast as WTM-eABF, but the fluctuation of biasing forces shown in Figure 3C decrease the quality of the free-energy landscape for a long-time meta-eABF simulation. Plain ABF and eABF have a smooth and relatively slow convergence rate in this case.

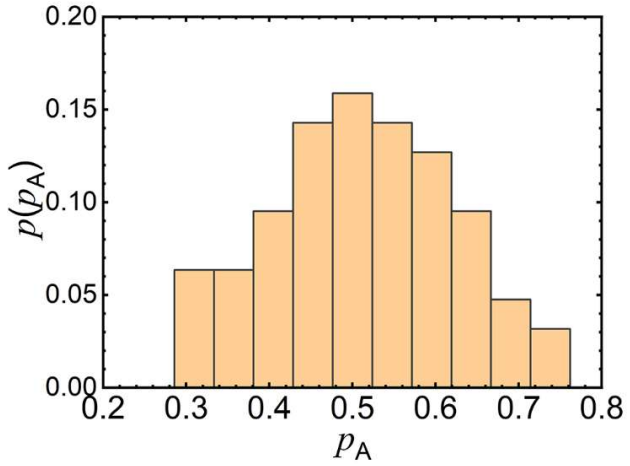

Figure S5. Distributions of the committor, p_A, at the positions near the saddle points around d_z = -14.3 and d_{xy} = 1.7 for the WTM-eABF simulation characterizing the translocation of a chloride ion through a transmembrane channel. The distributions are Gaussian-like with a peak at p_A = 0.5, suggesting the high reliability of the free-energy calculation and the chosen transition coordinate.

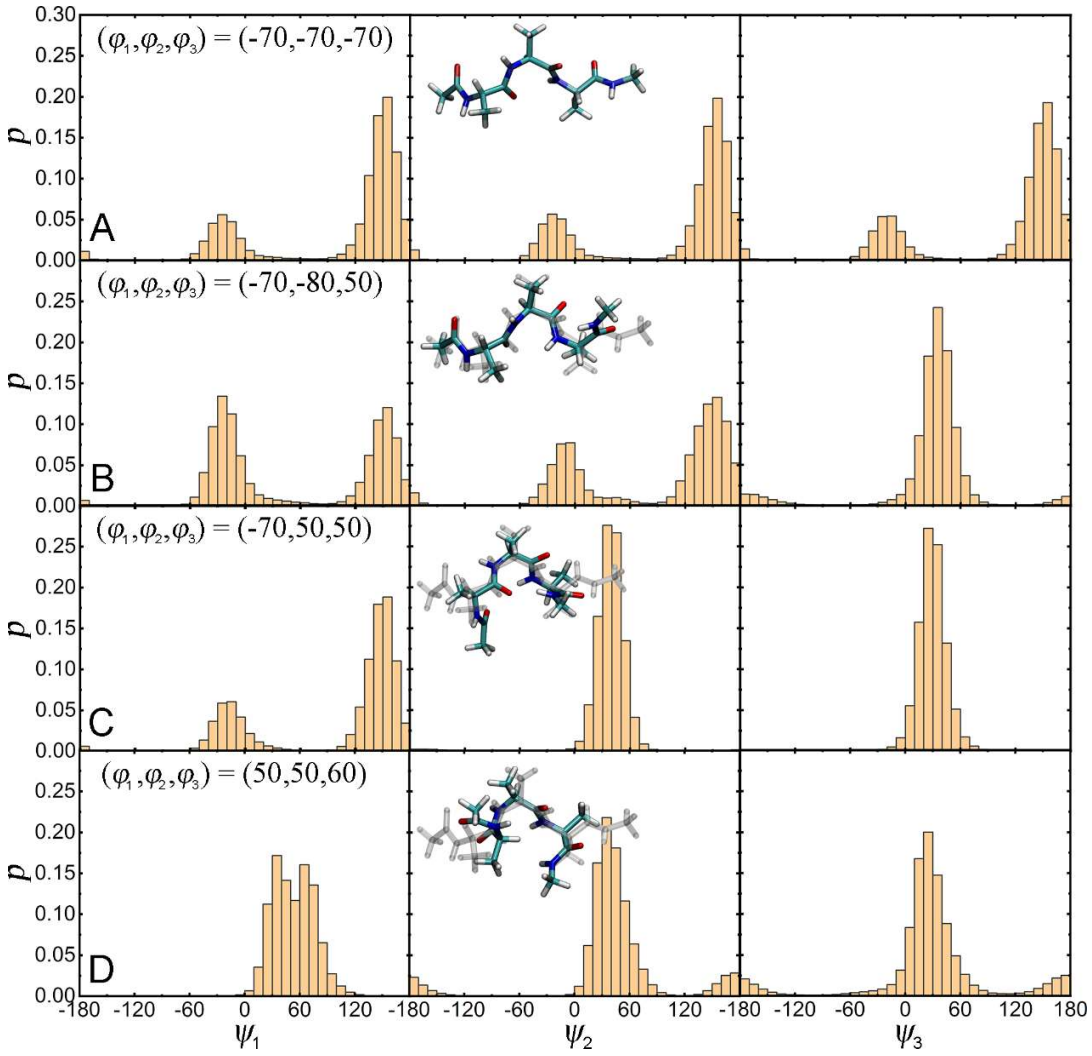

Figure S6. Distribution of ψ_1, ψ_2, and ψ_3 for metastable states in Figure 6. The results show that only one ψ angle rotates accompanied with the change of one φ angle. Insets: milestone structures with superposition of the optimal conformation.

Simulation details

All the simulations were performed using NAMD 2.13[4] with the latest version of Colvars module.[5] Covalent bonds involving hydrogen atoms were constrained to their equilibrium length employing the SHAKE/RATTLE[6] and SETTLE[7] algorithms. For simulations characterizing aqueous solution, the r-RESPA algorithm[8] was utilized to integrate the equations of motion with a timestep of 2 and 4 fs for short- and long-range interactions, respectively. A 12-Å cutoff was introduced to truncate van der Waals and short-range Coulombic interactions. The particle mesh Ewald (PME)[9] method was used to estimate the long-range electrostatic forces. Deca-alanine, N-acetyl-N'methylalaninamide (NANMA), transmembrane synthetic peptide nanotube and lipids were described by the CHARMM force field.[10] Alanine tripeptide was modeled by the Amber force field.[11] Water was characterized by the TIP3P model.[12]

Isomerization of N-acetyl-N'methylalaninamide in aqueous solution

Simulation parameters. To investigate the impact of different parameters of biasing potential on the convergence of WTM-eABF, we followed the control-of-variables strategy and performed a series of simulations using NANMA as a test example. The reaction coordinate model consisted of two torsion angles, φ and ψ of the peptide, as illustrated in Figure S7A, and the CZAR estimator was used to compute the gradients along them. All simulations can be categorized as three groups. In the first group (Figure S7B) different choices of bias temperature were used, while the width and initial height of Gaussians were kept at 5.0° and 0.1 kcal/mol, respectively. In the second group (Figure S7C) different values of initial Gaussian height were evaluated, while the width of Gaussians and the bias temperature were fixed at 5.0° and 4000 K, respectively. The last group

assessed various Gaussian widths using the same settings for bias temperature (4000 K) and initial Gaussian height (0.1 kcal/mol).

Results and discussion. The RMSD of resultant PMF in each group with respect to a reference from an 500-ns ABF simulation is shown in Figure S7.

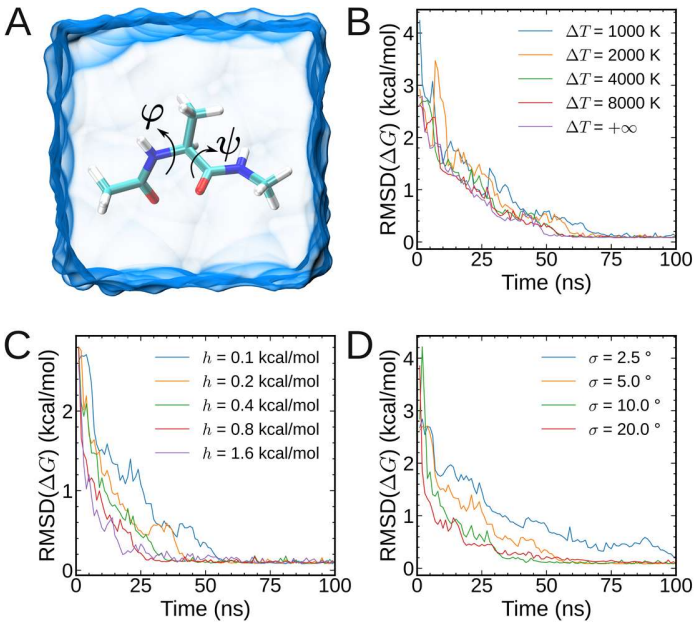

Figure S7. (A) Schematic representation of the transition coordinate of NANMA. Rate of convergence of WTM-eABF simulations with different (B) bias temperatures, (C) initial Gaussian heights and (D) Gaussian widths.

It can be concluded from Figure S7B that increasing the bias temperature slightly speedup the convergence. Nevertheless, as discussed in the manuscript, using the non-tempered parameter ($\Delta T = +\infty$) is not recommended, and the choices of 4000 K and 8000 K share approximately the same convergence rate with the non-tempered case, hence they are suitable for common PMF calculations. Inferred from Figure S7C and Figure S7D, increasing the initial height of Gaussians and the Gaussian width also accelerate the convergence, but extremely high initial height (h = 1.6 kcal/mol) and Gaussian width (σ = 20.0 °) can perturb the measurement of average forces, thus

yielding instable convergence compared with the moderate choices of h = 0.8 kcal/mol and σ = 10.0 °. In summary, the parameter set ΔT = 4000 K, h = 0.8 kcal/mol and σ = 10.0 ° can be used for mapping the multidimensional PMF along Ramachandran angles of peptides.

REFERENCES

(1) Johnson, M. E.; Hummer, G. Characterization of a Dynamic String Method for the Construction of Transition Pathways in Molecular Reactions. *J. Phys. Chem. B* **2012**, *116*, 8573–8583.

(2) Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-Level Characterization of Transport Cycle Thermodynamics in the Glycerol-3-Phosphate:Phosphate Antiporter. *Nat. Commun.* **2015**, *6*, 8393.

(3) E, W.; Ren, W.; Vanden-Eijnden, E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. *J. Chem. Phys.* **2007**, *126*, 164103.

(4) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(5) Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111*, 3345–3362.

(6) Andersen, H. C. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(7) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(8) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

(9) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(10) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2016**, *14*, 71.

(11) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(12) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.