Supporting Information

Two Dimensional Window Exchange Umbrella Sampling for Transmembrane Helix Assembly

*Soohyung Park and Wonpil Im**

Department of Molecular Biosciences and Center for Bioinformatics, The University of Kansas, 2030 Becker Drive, Lawrence, Kansas 66047, USA

S1. Sampling of the TM-TM interfaces near the global PMF minimum in 1D-WEUSMD

Sampling of the GpA-TM interfaces can be represented by the rotation angle distribution of Gly79 in each helix, $P(\theta_1,\theta_2)$. The rotation angles θ_1 and θ_2 are defined as follows. For two helices i and j, θ_1 is defined by the angle between $\mathbf{q}^{(j)} - \mathbf{q}^{(i)}$ and $\mathbf{r}_\alpha^{(i)} - \mathbf{q}^{(i)}$ (Figure S1A), where $\mathbf{q}^{(i)}$ is the projection of $\mathbf{r}_\alpha^{(i)}$, the ^{79}Gly C_α position in helix i, onto its helix axis. θ_2 is defined similarly. For comparison with the distribution obtained from TREXMD, the configurations sampled in the thermally accessible region (i.e., the $2k_BT$ region around the global PMF minimum)[1] for the IS1 and IS2 1D-WEUSMDs (8-11 Å in Figure 1A) were analyzed. As shown in Figures S1B-D, $P(\theta_1,\theta_2)$ from both 1D-WEUSMDs show good agreement with that from TREXMD. For two major peaks shown in $P(\theta_1,\theta_2)$, we calculated the root mean square deviation (RMSD) between the representative structures from 1D-WEUSMD and from TREXMD. All the calculated RMSD values are below 1 Å, demonstrating that the sampling in the thermally accessible region from 1D-WEUSMD is comparable to that from TREXMD.

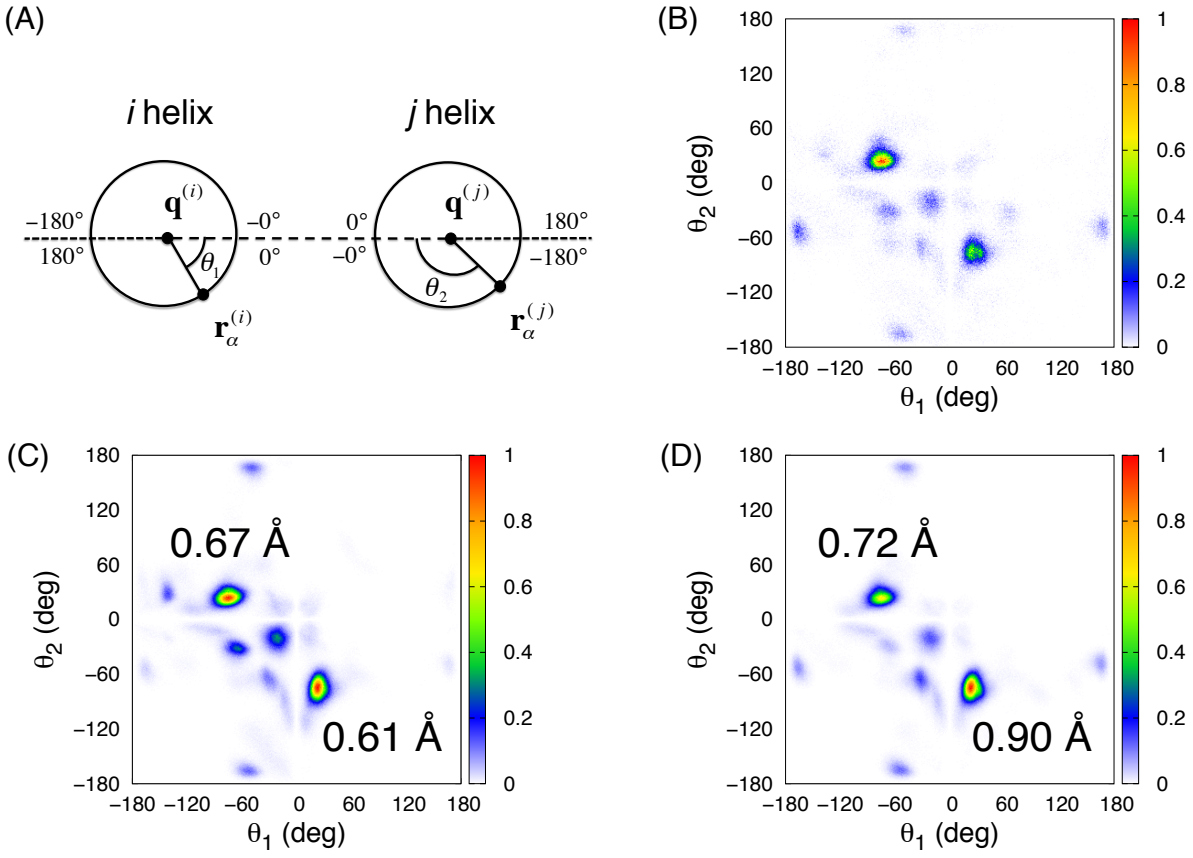

Figure S1. (A) Definition of rotation angles θ_1 and θ_2. The population plots of θ_1 and θ_2, $P(\theta_1,\theta_2)$, from (B) TREXMD, (C) IS1 1D-WEUSMD, and (D) IS2 1D-WEUSMD. In (C) and (D), the values of C_α RMSD between the representative structures from 1D-WEUSMD and from TREXMD are also shown.

S2. Exchange bottlenecks in 1D-WEUSMD

In the 1D-WEUSMD simulations, our hypothesis was that the IS1 and IS2 PMFs should be comparable if the WEUSMD has sufficient sampling power, i.e., the results should not depend on the initial configurations. However, as shown in Figure 1A, the discrepancy between the IS1 and IS2 PMFs implies that the sampling at r_{HH} < 8 Å is incomplete. To find out the cause, we analyzed replicas' walk along r_{HH} by making histograms of replicas' position along r_{HH}. As shown in Figure S2, bottlenecks of window exchange in 1D-WEUSMD exist around r_{HH} = 8 Å for both IS1 and IS2, which lowers the WEUSMD performance. Consequently, the PMF converged slowly for the IS1 because of more severe bottlenecks than in IS2 (Figure S3A). And, the sampling is incomplete at short r_{HH} during the 200-ns simulation time (Figures S3C and S3D). The origin of such bottlenecks can be attributed to the high barriers along the helix-helix crossing angle (Ω) at short r_{HH}.

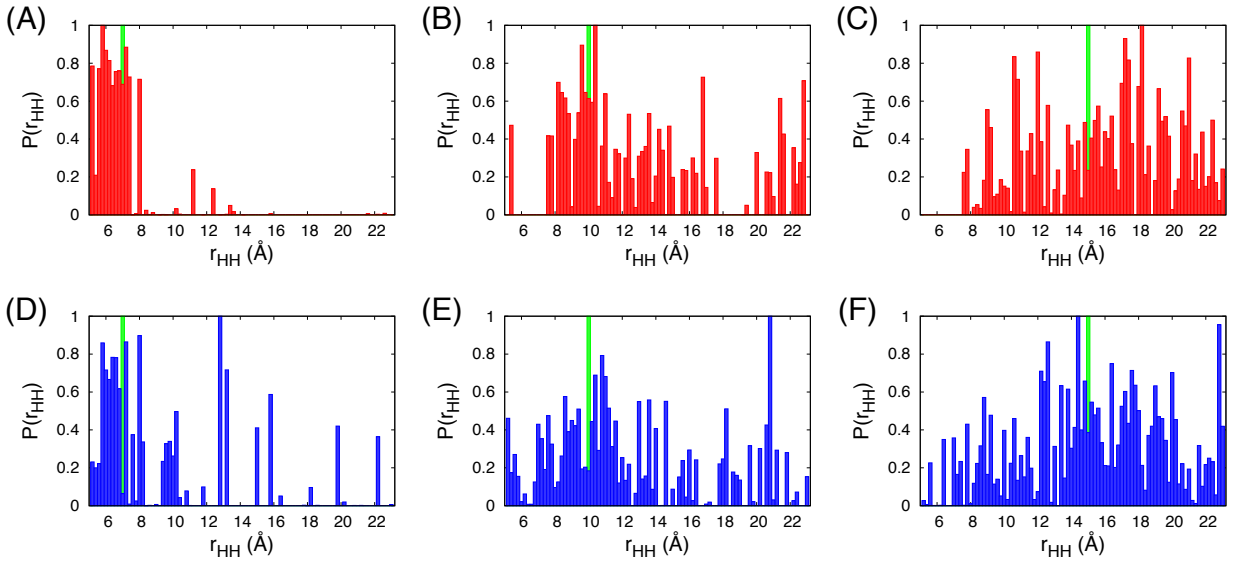

Figure S2. The histograms of the initial locations of the replicas (at t = 80 ns) that visited the windows at r_{HH} = 7, 10, and 15 Å (green bars) for (A-C) IS1 and (D-F) IS2, respectively.

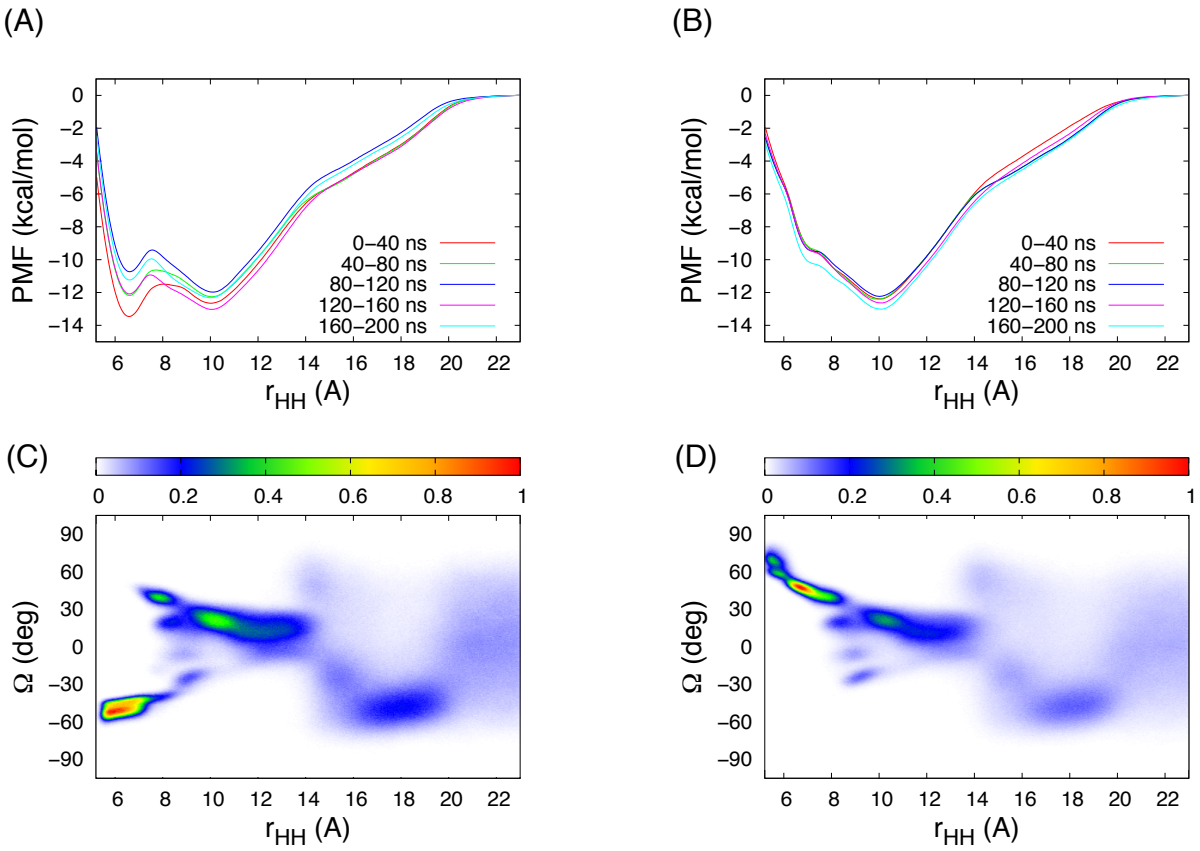

Figure S3. The block-averaged PMFs from (A) IS1 and (B) IS2 1D-WEUSMDs. The population in r_{HH} and Ω, $P(r_{HH},\Omega)$, from (C) IS1 and (D) IS2 1D-WEUSMDs. The inconsistency between the results from the IS1 and IS2 WEUSMDs can be attributed to the high barriers along Ω at short r_{HH}.

S3. Additional results from 2D-WEUSMD

Here, we provide additional results from 2D-WEUSMD: (i) the block-averaged PMFs for two time intervals, (ii) the RMSD obtained from the last seven 10-ns block-averaged PMFs, (iii) the PMF along Ω at several r_{HH} (≥ 9 Å), and (iv) the overlay of the boundary contour lines for B1 and B2 basins (defined based on the 2D-PMF) with $P(r_{HH},\Omega)$ from TREXMD. The time evolution of the 2D-PMF can be monitored by the series of the block-averaged 2D-PMFs. As shown in Figures S4A and S4B, the initially strong PMF minimum corresponding to the NMR structure became weaker and the other PMF minima corresponding to $P(r_{HH},\Omega)$ from TREXMD developed as the simulation time increased. The 2D-PMF converged after 30 ns, as shown in Figure S4C, faster than that for 1D-WEUSMD. Figure S4D shows the 1D-PMFs along Ω at several r_{HH} (shown in legends), where the barrier along Ω from right-handed to left-handed (and vice versa) conformations of GpA-TM is negligible at r_{HH} > 10 Å. The density in r_{HH} and Ω, $\varrho(r_{HH},\Omega)$, from 2D-WEUSMD (Figure S4E) shows a good agreement with $P(r_{HH},\Omega)$ from TREXMD (Figure S4F), where the latter was excellently represented by the B1 basin.

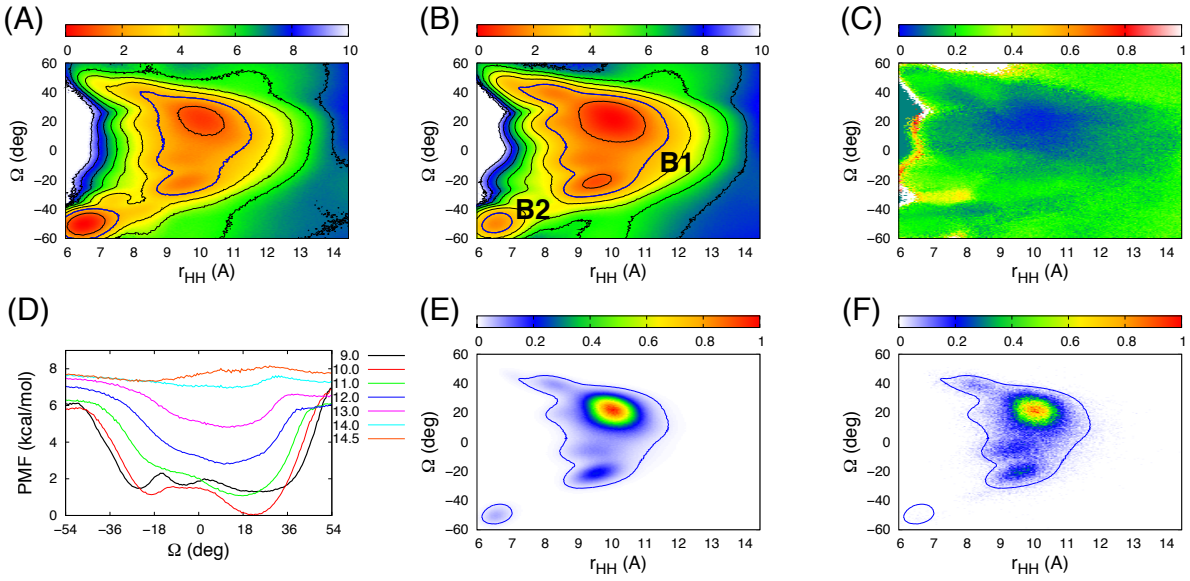

Figure S4. The block-averaged PMFs from 2D-WEUSMD for (A) 0-30 ns and (B) 30-100 ns. (C) The RMSD of the 2D-PMF (in kcal/mol) calculated from the last seven 10-ns block-averaged PMFs. (D) The PMFs along Ω at several r_{HH} (shown in legends). The overlay of the boundary of the basins B1 and B2 on (E) $\varrho(r_{HH},\Omega)$ from 2D-WEUSMD and (F) $P(r_{HH},\Omega)$ from TREXMD at T =300 K. The contour lines are shown at every 1.2 kcal/mol in (A) and (B).

S4. Two most probable pathways along r_{HH} and Ω

Based on the 2D-PMF, two most probable pathways L_1 and L_2 along r_{HH} and Ω (Ω < 0° or Ω > 0° at short r_{HH}, respectively) were obtained by the modified string method,[2] as described below. Initially, the value of Ω for the end point at r_{HH} = 14.5 Å was set to 0° and those at r_{HH} = 6 Å were set to –60° and 60° for the L_1 and L_2, respectively. Then, a total of 51 points were distributed evenly along the initial pathways given by the lines connecting these end points. During Monte Carlo (MC) simulations, these points were allowed to move along both r_{HH} and Ω except the end points that were allowed to move only along Ω. The points were relocated along the pathway evenly after each MC step. After 20,000 MC steps, we obtained a final pathway. A total of 1,000 such final independent pathways for the two pathways were used to obtain the average pathways L_1 and L_2 (shown in Figure 2A). These pathways are also shown in Figure S5A. The per-point deviation (PPD) of the each pathway from the average, $d(i)$, was calculated by

$$d(i) = \frac{1}{N}\sqrt{\sum_{k=1}^{N}\left[\left(\frac{r_{HH,k}(i) - \bar{r}_{HH,k}}{\delta r_{HH}}\right)^2 + \left(\frac{\Omega_k(i) - \bar{\Omega}_k}{\delta\Omega}\right)^2\right]} \qquad \text{(S1)}$$

where N = 51 is the total number of data points along the path and i is the index number for the independent MC runs (i = 1, 2, …, 1000) with 20,000 step each. $\bar{r}_{HH,k}$ and $\bar{\Omega}_k$ are the positions of the k-th data point along the average path, and δr_{HH} and δΩ are the intervals between the grid points along r_{HH} and Ω, respectively. We defined the average path ensemble as a set of paths with $d(i) < 0.8\ d_{mp}$, where the factor 0.8 is determined empirically and d_{mp} is the most probable PPD obtained from the histogram of $d(i)$'s (shown in Figure S5B). Statistical analysis for the average path PMFs was performed on the basis of these ensembles.

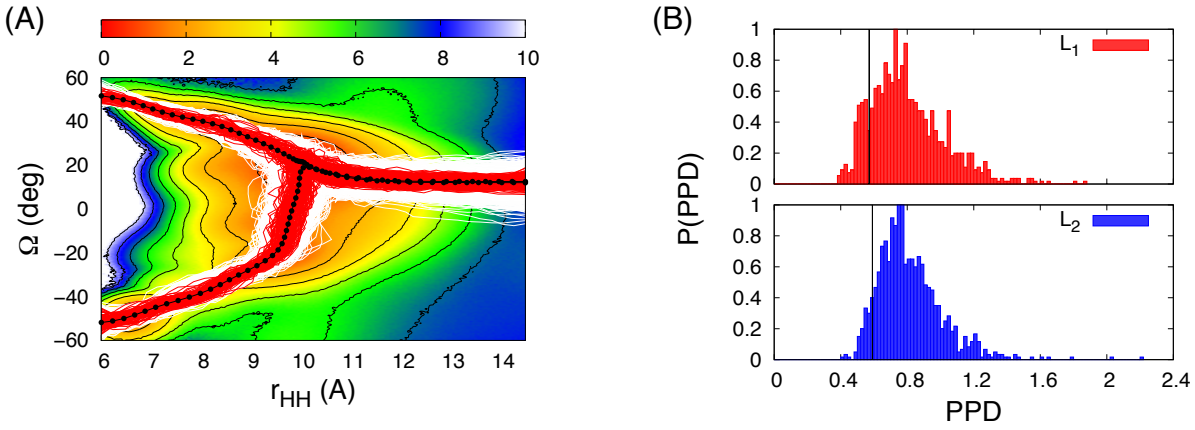

Figure S5. (A) Two sets of 1,000 independent GpA-TM assembly pathways (white) are mapped on the 2D-PMF. Also shown are the average pathways L_1 and L_2 (black) and the corresponding ensemble of each path (red). The 2D-PMF contour lines are shown at every 1.2 kcal/mol. (B) The histograms of the per-point deviation (PPD) for the sampled pathways for the L_1 and L_2. The black vertical lines are the cutoff PPD values, i.e., $d_{cut} = 0.8\ d_{mp}$.

REFERENCES

(1) Park, S.; Kim, T.; Im, W. Transmembrane helix assembly by window exchange umbrella sampling. *Phys. Rev. Lett.* **2012**, *108*, 108102.

(2) Jo, S.; Rui, H.; Lim, J. B.; Klauda, J. B.; Im, W. Cholesterol flip-flop: insights from free energy simulation studies. *J. Phys. Chem. B* **2010**, *114*, 13342-13348.