Supporting Information

Theory of Adaptive Optimization for Umbrella Sampling

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. DATA FOR THE TOY MODEL FROM ALL THE TESTED PARAMETER SETS

S1.1. The Kullback-Leibler divergence D_{KL}

Here, we present the D_{KL} (eq 18 in the main text) averaged over the all the sets of the initial configurations for each parameter set. In the calculation of D_{KL}, for numerical stability, we set $P(y|x) = \max\{\epsilon, P(y|x)\}$, where $\epsilon = 10^{-13}$. From the average D_{KL}, the relaxation time of $D_{\mathrm{KL}}(t)$, τ, is estimated by the numerical fit to $a + b\exp(-t/\tau)$. The results are shown in Figure S1 and summarized in Table S1.

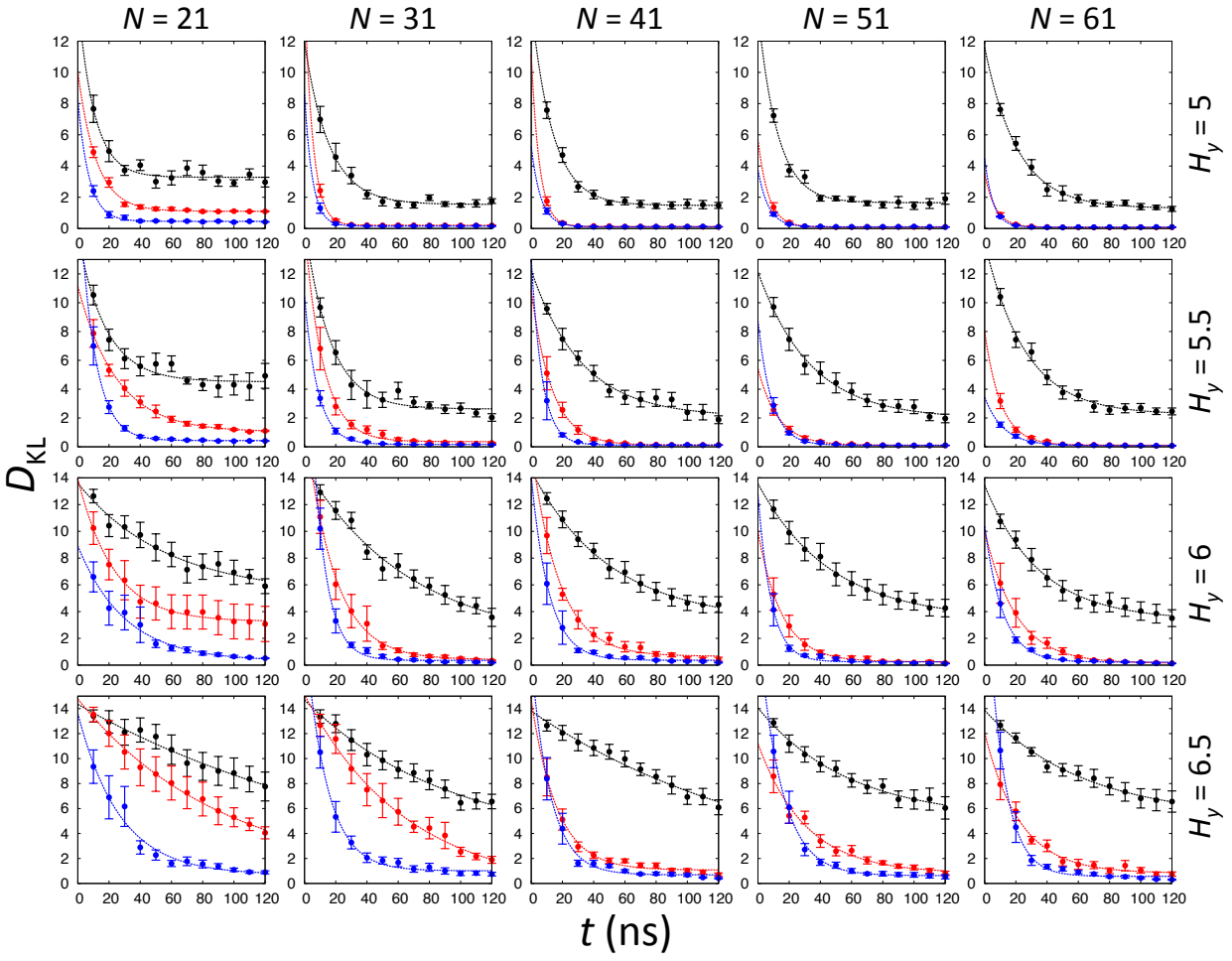

FIG. S1: The average Kullback-Leibler divergence D_{KL} for $P(y|x)$ for all the tested parameter sets. Data from USMD, WEUSMD, and aWEUSMD are shown in black, red, and blue circles with error bars, respectively. Dashed lines are the numerical fit of D_{KL} to $a + b\exp(-t/\tau)$. The error bars represent the standard error.

TABLE S1: The average relaxation time τ of D_{KL} for $P(y|x)$

H_{y}	Simulation	N (Number of windows)				
		21	31	41	51	61
5	USMD	10.5 ± 2.1[a]	15.9 ± 1.5	13.7 ± 0.7	12.4 ± 1.5	21.0 ± 1.2
	WEUSMD	11.8 ± 0.9	5.21 ± 0.05	5.23 ± 0.17	6.73 ± 0.13	6.90 ± 0.19
	aWEUSMD	7.19 ± 0.53	5.01 ± 0.16	6.15 ± 0.14[b]	6.64 ± 0.23	5.34 ± 0.23
5.5	USMD	17.8 ± 3.4	16.6 ± 2.3	32.2 ± 3.9	33.6 ± 3.2	26.1 ± 2.2
	WESUMD	24.7 ± 0.9	11.8 ± 0.9	13.2 ± 0.3	12.9 ± 0.2	10.5 ± 0.5
	aWEUSMD	9.70 ± 0.13	8.58 ± 0.31	6.96 ± 0.18	8.76 ± 0.28	12.0 ± 0.3[b]
6	USMD	57.0 ± 15.9	68.5 ± 14.4	50.6 ± 3.6	46.4 ± 3.4	35.3 ± 3.2
	WEUSMD	22.9 ± 1.9	18.6 ± 1.1	17.6 ± 1.2	15.3 ± 0.6	19.9 ± 1.0
	aWEUSMD	30.2 ± 3.8[b]	8.91 ± 0.47	11.4 ± 0.7	8.49 ± 0.93	11.7 ± 0.6
6.5	USMD	200 ± 11[c]	92.7 ± 22.9	160 ± 8[c]	58.0 ± 9.2	63.2 ± 10.2
	WEUSMD	92.8 ± 14.3	65.2 ± 9.2	16.4 ± 1.3	30.3 ± 4.2	22.6 ± 2.0
	aWEUSMD	26.9 ± 4.1	13.9 ± 0.8	12.4 ± 1.1	14.4 ± 0.8	10.6 ± 0.6

[a]The number following $\pm$ sign is the asymptotic standard error.

[b]Although τ is larger, D_{KL} is smaller than that from WEUSMD (see Figure S1).

[c]data from the fit to $b\exp(-t/\tau)$

S1.2. The number of y-barrier crossing N_{cross}

Here, we provide the average number of y-barrier crossing N_{cross} for all the tested parameter sets. For each parameter set, the number of y-barrier crossing was calculated from the trajectory of each replica for each set of initial configurations and then averaged, from which N_{cross} and its standard error were estimated. The data are shown in Figure S2 and summarized in Table S2.

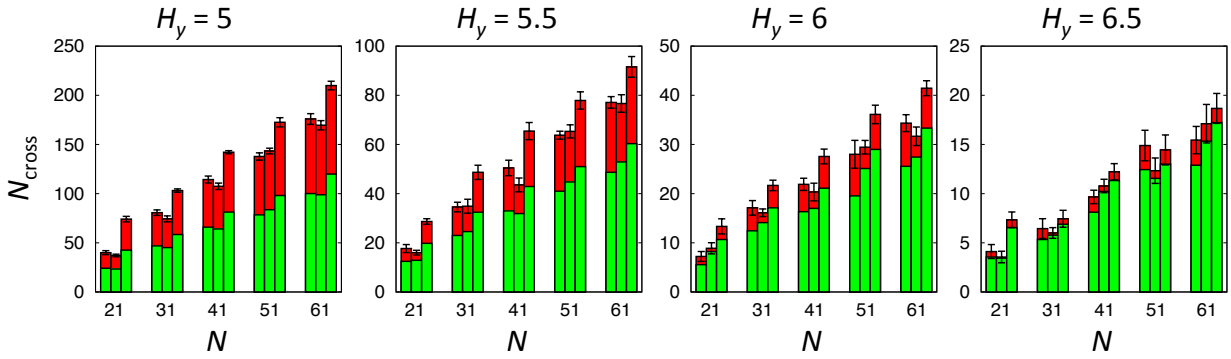

FIG. S2: The average number of y-barrier crossing N_{cross}. For each N, the left, center, and right bars are the data calculated fromUSMD, WEUSMD, and aWEUSMD, respectively. The green boxes represent the number of y-barrier crossing from $y > 0$ to $y < 0$ and the red boxes represent those in the opposite direction. The error bars are the standard error of N_{cross}. As H_y increases, the improvement becomes less obvious.

TABLE S2: The number of crossing y-barrier N_{cross}

H_y	Simulation	N (Number of windows)				
		21	31	41	51	61
5	USMD	40.0 ± 2.0[a]	80.7 ± 2.8	114 ± 4	138 ± 4	176 ± 5
	WEUSMD	37.1 ± 1.3	74.3 ± 3.1	107 ± 3	143 ± 3	170 ± 5
	aWEUSMD	74.1 ± 2.8	103 ± 2	142 ± 2	173 ± 5	210 ± 4
5.5	USMD	17.7 ± 1.6	34.6 ± 2.0	50.4 ± 3.2	63.8 ± 1.6	77.1 ± 2.4
	WESUMD	16.0 ± 0.9	34.9 ± 2.8	43.6 ± 2.8	65.3 ± 2.7	76.7 ± 3.6
	aWEUSMD	28.7 ± 1.1	48.7 ± 2.9	65.4 ± 3.5	77.9 ± 3.5	91.6 ± 4.2
6	USMD	7.22 ± 1.00	17.1 ± 1.5	21.9 ± 1.2	28.0 ± 2.8	34.3 ± 1.7
	WEUSMD	8.89 ± 1.14	16.1 ± 0.8	20.3 ± 1.8	29.4 ± 1.4	31.7 ± 1.9
	aWEUSMD	13.3 ± 1.5	21.7 ± 1.1	27.6 ± 1.5	36.1 ± 1.9	41.4 ± 1.5
6.5	USMD	4.11 ± 0.71	6.44 ± 1.01	9.67 ± 0.68	14.9 ± 1.6	15.4 ± 1.4
	WEUSMD	3.56 ± 0.59	6.00 ± 0.54	10.8 ± 0.7	12.3 ± 1.3	17.1 ± 2.0
	aWEUSMD	7.33 ± 0.80	7.44 ± 0.86	12.2 ± 0.83	14.4 ± 1.5	18.7 ± 1.5

[a]The number following $\pm$ sign is the standard error.

S1.3. The random walk of replicas in WEUSMD and aWEUSMD

Here, we provide the data representing the quality of random walk of replicas along window space, which is shown by the fraction of replicas that have visited the lowest-index window most recently, $f(i)$, in eq 19 in the main text. For each parameter set, the fraction was calculated for each set of initial configurations and then averaged, from which $f(i)$ and its standard error were estimated (see Figure S3).

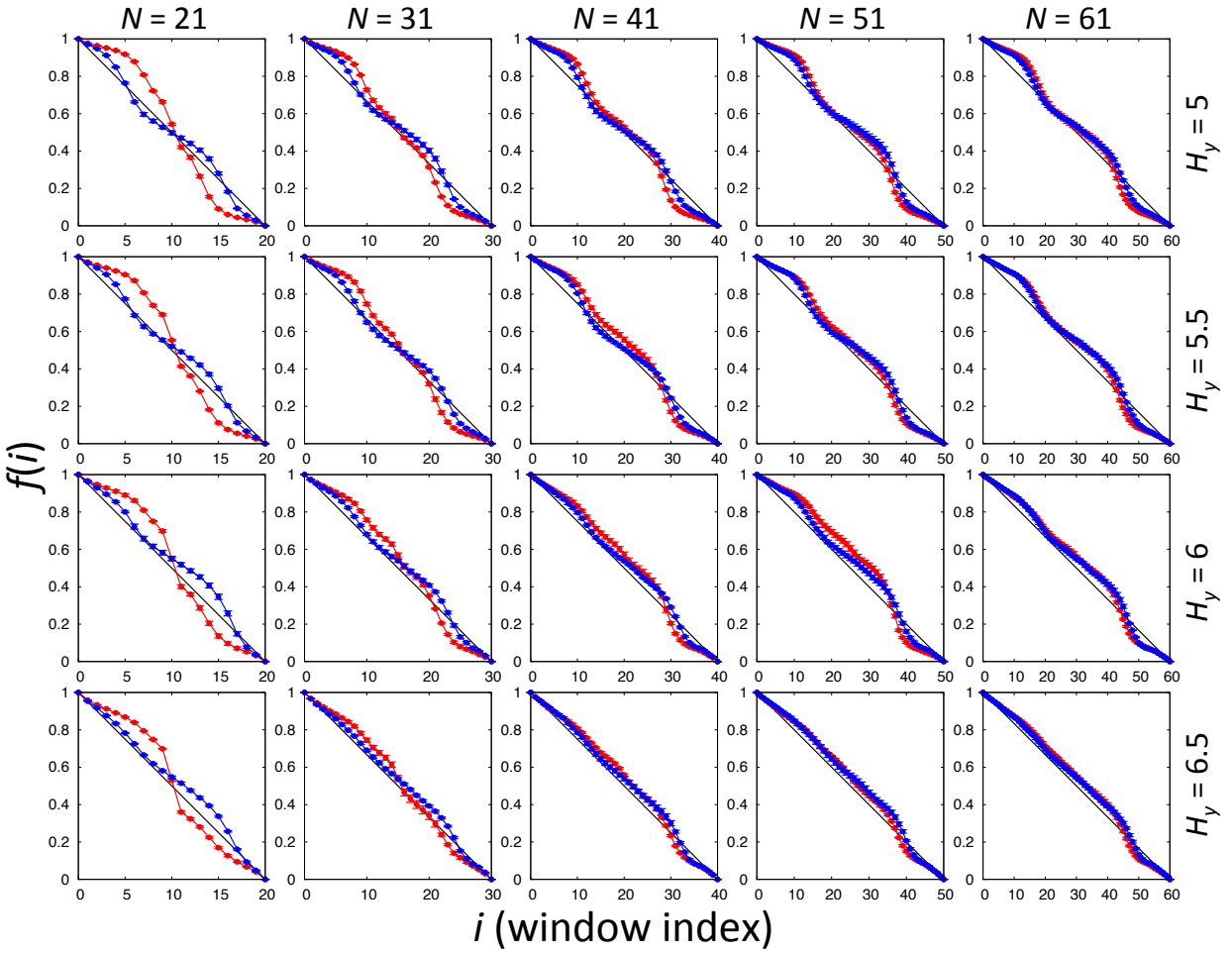

FIG. S3: The fraction of replica that have visited the lowest-index window most recently, $f(i)$, for all the tested parameter sets. Data from WEUSMD and aWEUSMD are shown in red and blue circles with error bars. The black line is $f(i)$ for the ideal random walk situation. The error bars represent the standard error. The improvement in the quality of random walk is more apparent for smaller number of windows. As N increases, the improvement becomes less obvious.

S2. ADDITIONAL DATA FOR THE GPA-TM ASSEMBLY

S2.1. The average RMSD $R_a(r_{HH}, \Omega)$ for IS1 and IS2

Here, we provide the average RMSD of sampled conformations, $R_a(r_{HH}, \Omega)$, from WEUSMD and aWEUSMD for IS1 and IS2. The RMSD with respect to the NMR structure was calculated for each conformation at (r_{HH},Ω) and then averaged over the simulation time to obtain $R_a(r_{HH}, \Omega)$. As shown in Figure S4, aWEUSMD was able to sample wider conformations along Ω, especially, at $r_{HH} < 9$ Å. The sampled region by aWEUSMD for IS1 includes the NMR structure-like conformations, which supports the improved sampling efficiency by aWEUSMD over WEUSMD.

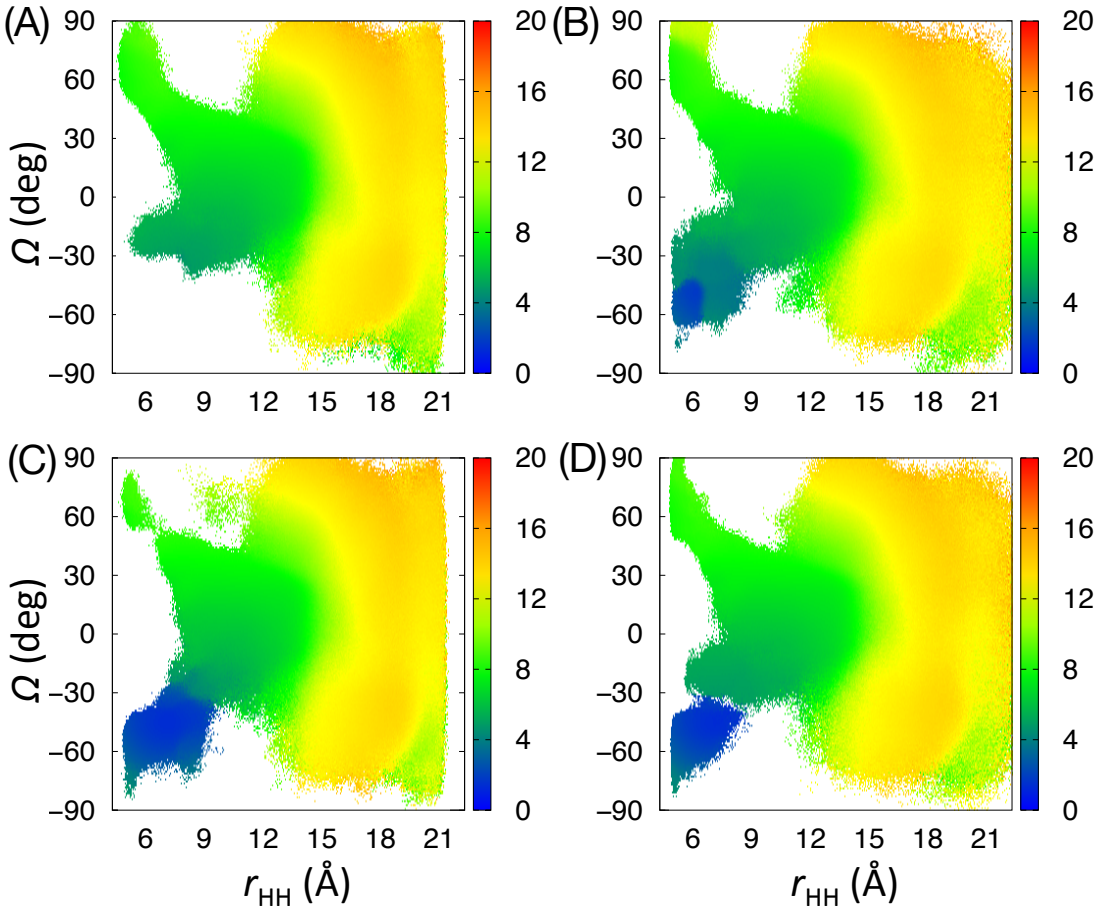

FIG. S4: (A)-(B) The average RMSD of sampled conformations with repeat to the NMR structure, $R_a(r_{HH}, \Omega)$, calculated from the results of (A) WEUSMD and (B) aWEUSMD for the IS1. (C)-(D) $R_a(r_{HH}, \Omega)$, calculated from the results of (C) WEUSMD and (D) aWEUSMD for the IS2. In RMSD calculations, C_α and C_β atoms were used and then averaged over the simulation time.

S2.2. Data from WEUSMD and aWEUSMD for IS3

Here, we present the results from WEUSMD and aWEUSMD starting with the initial configurations with the left-handed helix-dimer interfaces (IS3). The PMFs obtained from the results of WEUSMD and aWEUSMD differ within the error bar (Figure S5A). The average relaxation time τ of window parameters from IS3 aWEUSMD are about 38 ns and 3.8 ns for the bias force constants and window centers, respectively, which are consistent with those obtained from the IS1 and IS2 aWEUSMD.

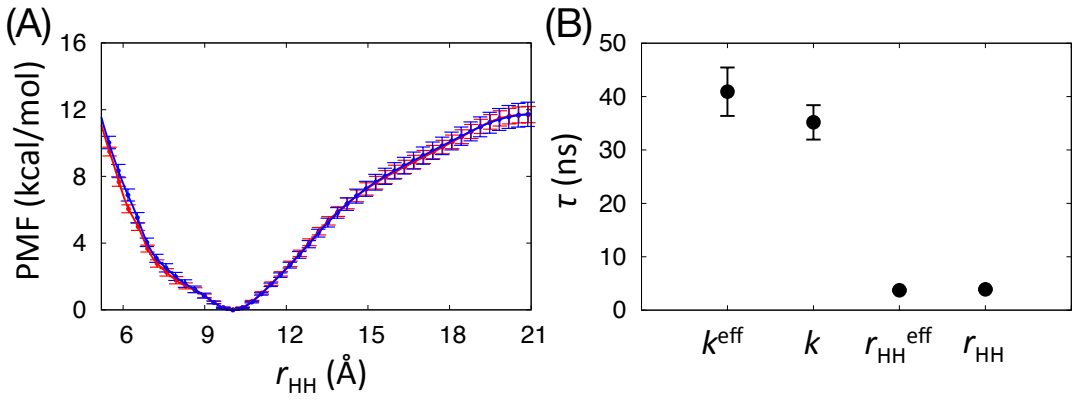

FIG. S5: (A) The PMF as a function of r_{HH}, $\mathcal{W}(r_{\mathrm{HH}})$, calculated from the results of WEUSMD (red) and aWEUSMD (blue) starting from the initial configurations with parallel helix-dimer interfaces (IS3). For the PMF calculation, we used trajectories from 40 ns to 200 ns. The error bars represent the standard deviations calculated from 16 10-ns block average PMFs. (B) The average relaxation time τ of window parameters for IS3. The error bars are the standard errors of the average τ over windows.

As shown in Figure S6, the sampling power of WEUSMD and aWEUSMD were comparable for IS3, which is different from the results of aWEUSMD for IS1 and IS2. This suggests that WEUSMD starting with the initial configurations with parallel helix-dimer interfaces can be a good choice for the TM assembly of unknown interfaces.

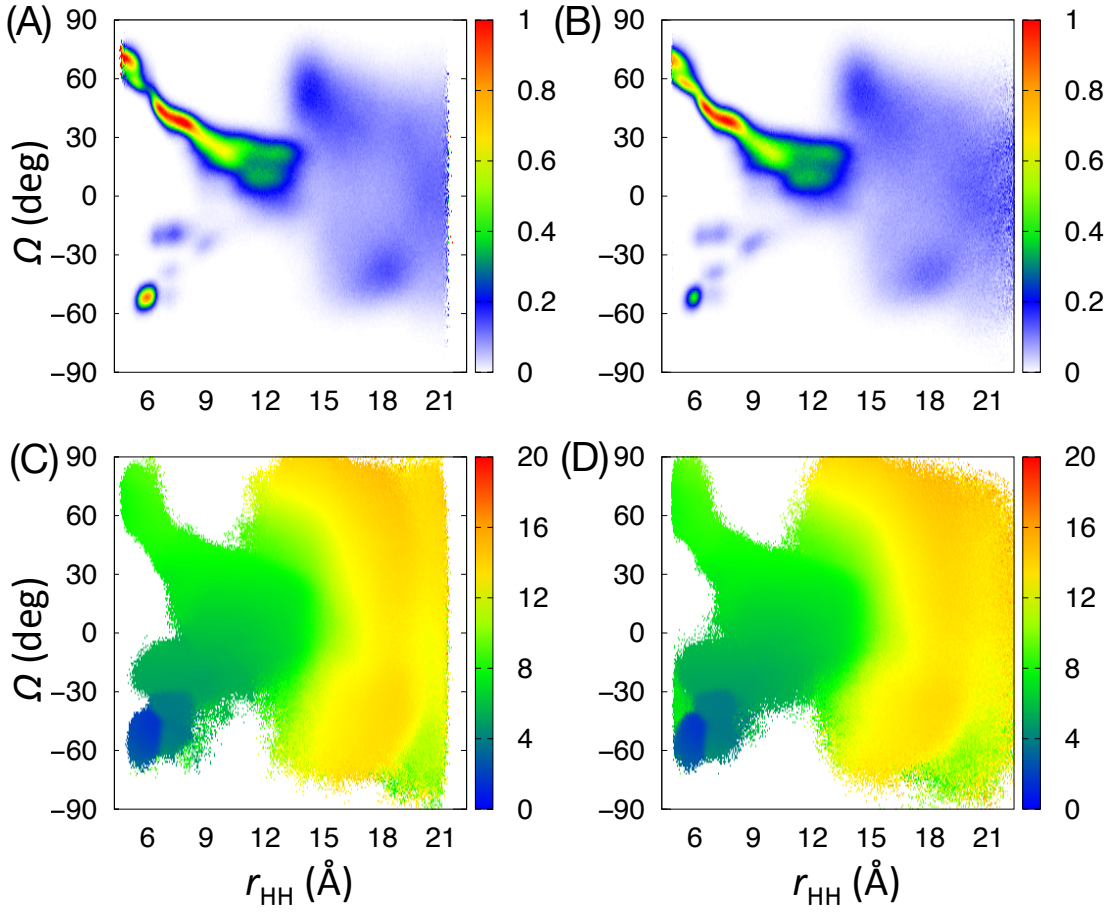

FIG. S6: The conditional probability, $P(\Omega|r_{\mathrm{HH}})$, calculated from the results of (A) WEUSMD and (B) aWEUSMD for IS3. The average RMSD of sampled conformations with respect to the NMR structure, $R_{\mathrm{a}}(r_{\mathrm{HH}}, \Omega)$, calculated from the results of (C) WEUSMD and (D) aWEUSMD for IS3.