Electronic Supplementary Information

Potential Energy Interpolation with Target-Customized Weighting Coordinates: Application to Excited-State Dynamics of Photoactive Yellow Protein Chromophore in Water

Seung Soo Kim and Young Min Rhee*

Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST),

Daejeon 34141, Korea

*E-mail: ymrhee@kaist.ac.kr

Supporting Text: Determining the confidence radii for two-part weight functions

In line with an earlier report,[1] the confidence radii for the two-part weight functions, namely $\mathrm{crad}(n)$ in eq 7 of the main text, can be assigned by adopting a Baysian analysis to the interpolation database. In brief, the Baysian analysis statistically estimates the confidence radii, each of which is the distance where the expansion error at a given data point does not exceed a tolerance value, E_{tol}. The estimation is from the nearest M neighboring configurations, usually collected from other interpolation data points. Within a simple assumption to the relationship between the distance from the data point and its expansion error, the Baysian confidence radius for a given data point can be written as

$$\mathrm{crad}(n)^{-6} = \frac{1}{M} \sum_{l=1}^{M} \frac{\left[T_n(\mathbf{Z}_l^{\mathrm{nei}(n)}) - U_l^{\mathrm{nei}(n)} \right]^2}{E_{\mathrm{tol}}^2 \left[d_n(\mathbf{X}_l^{\mathrm{nei}(n)}) \right]^6} \tag{S1}$$

where $\mathbf{Z}_l^{\mathrm{nei}(n)}$ and $\mathbf{X}_l^{\mathrm{nei}(n)}$ are the geometry of the l-th neighboring configuration for the n-th data point geometry, as represented in the Z-matrix internal coordinates and in the Cartesian coordinates, respectively. $U_l^{\mathrm{nei}(n)}$ is the quantum chemical potential energy obtained at the l-th neighboring configuration, $\mathbf{X}_l^{\mathrm{nei}(n)}$ among the nearest M neighboring configurations. Here, the superscript "nei(n)" is used to emphasize the quantities are from the point belonging to the set of the neighboring configurations for the n-th data point. Although eq S1 is written for the case where interpolation is conducted on a single electronic state, extending for multistate cases is also straightforward.[2]

One can notice that the confidence radii thus calculated are nothing but the averages of the estimated values from the neighboring configurations. Thus, the confidence radii will likely be more reliable as more neighboring configurations are considered, and they can be simultaneously updated with additional sampling of the interpolation data points. However, our earlier study showed that when the interpolation is conducted for a large molecule, there sometimes should be additional sampling of neighboring configurations to render the confidence radii reliable. In this work, to avoid an additional computational cost associated with that, we modified eq S1 as follows:

$$\mathrm{crad}(n)^{-6} = \max\left\{ \frac{\left[T_n(\mathbf{Z}_l^{\mathrm{nei}(n)}) - U_l^{\mathrm{nei}(n)}\right]^2}{E_{\mathrm{tol}}^2 \left[d_n(\mathbf{X}_l^{\mathrm{nei}(n)})\right]^6} \right\} \quad l = 1, 2, \ldots, M \tag{S2}$$

With the target-customized weighting coordinate defined in the main text, the equation can be further modified as

$$\mathrm{crad}(n)^{-6} = \max\left\{ \frac{\left[T_n(\mathbf{Z}_l^{\mathrm{nei}(n)}) - U_l^{\mathrm{nei}(n)}\right]^2}{E_{\mathrm{tol}}^2 \left[d'_n(\mathbf{X}_l^{\mathrm{nei}(n)}, \mathbf{z}_l^{\mathrm{nei}(n)})\right]^6} \right\} \quad l = 1, 2, \ldots, M \tag{S3}$$

One may wonder that directly using eq S2 or S3 in dynamics simulations may potentially be harmful as these equations systematically underestimate the confidence radii by taking the maxima rather than the averages. However, we experienced that sensitively assigning the radii through these equations was particularly useful during our "static" calculations toward analyzing the effect of the weighting coordinates. For the values of M and E_{tol}, we choose M = 100, and E_{tol} = 1 kJ/mol, similar with the earlier report.[3] The obtained radii were used to calculate the corresponding contour maps with the two-part weighting functions shown in Figure 5c of the main text.

Supporting Tables

Table S1. Z-matrix internal coordinates adopted toward the interpolation.[a]

Bonds	Angles	Torsions
C_2-C_1		
C_3-C_2	C_3-C_2-C_1	
C_4-C_3	C_4-C_3-C_2	C_4-C_3-C_2-C_1
C_5-C_4	C_5-C_4-C_3	C_5-C_4-C_3-C_2
C_6-C_5	C_6-C_5-C_4	C_6-C_5-C_4-C_3
H_7-C_1	H_7-C_1-C_2	H_7-C_1-C_2-C_6
H_8-C_6	H_8-C_6-C_5	H_8-C_6-C_5-C_1
H_9-C_3	H_9-C_3-C_2	H_9-C_3-C_2-C_4
H_{10}-C_4	H_{10}-C_4-C_5	H_{10}-C_4-C_5-C_3
C_{11}-C_5	C_{11}-C_5-C_4	C_{11}-C_5-C_4-C_6
C_{12}-C_{11}	C_{12}-C_{11}-C_5	C_{12}-C_{11}-C_5-C_4
H_{13}-C_{11}	H_{13}-C_{11}-C_{12}	H_{13}-C_{11}-C_{12}-C_5
C_{14}-C_{12}	C_{14}-C_{12}-C_{11}	C_{14}-C_{12}-C_{11}-C_5
H_{15}-C_{12}	H_{15}-C_{12}-C_{11}	H_{15}-C_{12}-C_{11}-C_{14}
O_{16}-C_{14}	O_{16}-C_{14}-C_{12}	O_{16}-C_{14}-C_{12}-C_{11}
O_{17}-C_2	O_{17}-C_2-C_3	O_{17}-C_2-C_3-C_1
S_{18}-C_{14}	S_{18}-C_{14}-C_{12}	S_{18}-C_{14}-C_{12}-O_{16}
C_{19}-S_{18}	C_{19}-S_{18}-C_{14}	C_{19}-S_{18}-C_{14}-O_{16}
H_{20}-C_{19}	H_{20}-C_{19}-S_{18}	H_{20}-C_{19}-S_{18}-C_{14}
H_{21}-C_{19}	H_{21}-C_{19}-S_{18}	H_{21}-C_{19}-S_{18}-C_{14}
H_{22}-C_{19}	H_{22}-C_{19}-S_{18}	H_{22}-C_{19}-S_{18}-C_{14}

[a] Atom designations can be found in Figure 1 of the main text.

Table S2. RESP diabatic charges at the S_0-optimized geometry.

atom[a]	diabatic atomic partial charge[b]		
	P state	Q state	PQ coupling
C_1	−0.5179	−0.2995	0.0602
C_2	0.6994	0.8495	−0.0560
C_3	−0.5661	−0.4247	0.0243
C_4	−0.1132	−0.0028	−0.0761
C_5	−0.0891	0.1645	0.1267
C_6	−0.1764	−0.2069	−0.0649
H_7	0.1652	0.1266	−0.0048
H_8	0.1481	0.1334	0.0092
H_9	0.1742	0.1511	0.0000
H_{10}	0.1322	0.0986	0.0116
C_{11}	0.0756	0.0250	−0.0835
C_{12}	−0.4232	−1.0386	0.1047
H_{13}	0.0874	0.0934	−0.0094
C_{14}	0.8351	0.6803	−0.0512
H_{15}	0.2238	0.2908	−0.0128
O_{16}	−0.5802	−0.6552	0.0081
O_{17}	−0.8186	−0.6905	0.0104
S_{18}	−0.3927	−0.4343	0.0005
C_{19}	0.0119	0.0257	0.0067
H_{20}	0.0347	0.0300	−0.0007
H_{21}	0.0550	0.0535	−0.0022
H_{22}	0.0347	0.0299	−0.0007

[a] Atom designations can be found in Figure 1 of the main text.

[b] Diabatic state designations can be found in Figure 2b of the main text.

Table S3. Detailed simulation conditions of the restrained sampling for the IM database.

Bin index	Initial [α, β]	Restrained angle [a]
6	[70, 180]	α
7	[50, 180]	
8	[30, 180]	
9	[10, 180]	
10	[–10, 180]	
11	[–30, 180]	
12	[–50, 180]	
13	[–70, 180]	
14	[0, 250]	β
15	[0, 230]	
16	[0, 130]	
17	[0, 110]	
18	[0, 210]	α, β
19	[0, 190]	
20	[0, 170]	
21	[0, 150]	

[a] All the restraining centers were the same as the initial angle values. The restraining force constant was 200 kJ/mol/rad^2, excepting the restraint on α from bin indexes 18 – 21, where the force constant was 100 kJ/mol/rad^2 instead.

Table S4. Pearson's correlation coefficients between the squared effective distance from the closest data point and the state-averaged interpolation error for all bond stretching coordinates.[a]

Index	Bond[b]	Coefficient
1	C_1-C_2	–0.0417
2	C_1-C_6	–0.0017
3	C_1-H_7	0.0253
4	C_2-C_3	–0.0343
5	C_2-O_{17}	0.3326
6	C_3-C_4	0.1368
7	C_3-H_9	0.0264
8	C_4-C_5	0.0351
9	C_4-H_{10}	0.0289
10	C_5-C_6	0.0135
11	C_5-C_{11}	0.3074
12	C_6-H_8	0.0548
13	C_{11}-C_{12}	0.2383
14	C_{11}-H_{13}	0.0222
15	C_{12}-C_{14}	0.1012
16	C_{12}-H_{15}	–0.0011
17	C_{14}-O_{16}	–0.0568
18	C_{14}-S_{18}	–0.1096
19	S_{18}-C_{19}	0.1341
20	C_{19}-H_{20}	0.0594
21	C_{19}-H_{21}	0.0406
22	C_{19}-H_{22}	0.0630

[a] The effective distance was calculated as in eq 9 of the main text without any tolerance parameter, and the state-averaged interpolation error was defined as the RMS value of the errors from both S_0 and S_1 states. Each correlation coefficient was obtained with the test configurations.

[b] Atom designations can be found in Figure 1 of the main text.

Table S5. Pearson's correlation coefficients between the squared effective distance from the closest data point and the state-averaged interpolation error for all angle bending coordinates.[a]

Index	Angle[b]	Coefficient
1	C_1-C_2-C_3	0.2362
2	C_1-C_2-O_{17}	0.0810
3	C_1-C_6-C_5	0.0272
4	C_1-C_6-C_8	–0.0088
5	C_2-C_1-C_6	0.0874
6	C_2-C_1-H_7	0.0643
7	C_2-C_3-C_4	0.1990
8	C_2-C_3-H_9	0.0922
9	C_3-C_2-O_{17}	0.1600
10	C_3-C_4-C_5	–0.0053
11	C_3-C_4-H_{10}	0.0625
12	C_4-C_3-H_9	–0.0366
13	C_4-C_5-C_6	0.1594
14	C_4-C_5-C_{11}	0.0934
15	C_5-C_4-H_{10}	0.0942
16	C_5-C_6-H_8	0.0721
17	C_5-C_{11}-C_{12}	0.0490
18	C_5-C_{11}-H_{13}	0.1095
19	C_6-C_1-H_7	0.0205
20	C_6-C_5-C_{11}	0.0900
21	C_{11}-C_{12}-C_{14}	–0.0066
22	C_{11}-C_{12}-H_{15}	0.0562
23	C_{12}-C_{11}-H_{13}	–0.0096
24	C_{12}-C_{14}-O_{16}	0.2134
25	C_{12}-C_{14}-S_{18}	0.0695
26	C_{14}-C_{12}-H_{15}	–0.0713
27	C_{14}-S_{18}-C_{19}	0.2325
28	O_{16}-C_{14}-S_{18}	0.2454
29	S_{18}-C_{19}-H_{20}	0.0834
30	S_{18}-C_{19}-H_{21}	–0.0429
31	S_{18}-C_{19}-H_{22}	0.0239
32	H_{20}-C_{19}-H_{21}	0.0119
33	H_{20}-C_{19}-H_{22}	0.0360
34	H_{21}-C_{19}-H_{22}	–0.0722

[a] The effective distance was calculated as in eq 9 of the main text without any tolerance parameter, and the state-averaged interpolation error was defined as the RMS value of the errors from both S_0 and S_1 states. Each correlation coefficient was obtained with the test configurations.

[b] Atom designations can be found in Figure 1 of the main text.

Table S6. Pearson's correlation coefficients between the squared effective distance from the closest data point and the state-averaged interpolation error for all proper dihedral coordinates.[a]

Index	Dihedral Angle[b]	Coefficient	Index	Dihedral Angle[b]	Coefficient
1	C_1-C_2-C_3-C_4	0.0595	26	H_7-C_1-C_6-H_8	0.0416
2	C_1-C_2-C_3-H_9	0.0800	27	H_8-C_6-C_5-C_{11}	0.0153
3	C_1-C_6-C_5-C_4	0.1203	28	H_9-C_3-C_2-O_{17}	0.0643
4	C_1-C_6-C_5-C_{11}	0.0425	29	H_9-C_3-C_4-H_{10}	0.0710
5	C_2-C_1-C_6-C_5	0.0841	30	H_{10}-C_4-C_5-C_{11}	0.0125
6	C_2-C_1-C_6-H_8	0.0095	31	C_{11}-C_{12}-C_{14}-O_{16}	0.0279
7	C_2-C_3-C_4-C_5	0.2057	32	C_{11}-C_{12}-C_{14}-S_{18}	0.1322
8	C_2-C_3-C_4-H_{10}	0.0316	33	C_{12}-C_{14}-S_{18}-C_{19}	0.1420
9	C_3-C_4-C_5-C_6	0.2331	34	H_{13}-C_{11}-C_{12}-C_{14}	0.2962
10	C_3-C_4-C_5-C_{11}	0.1383	35	H_{13}-C_{11}-C_{12}-H_{15}	0.3723
11	C_4-C_3-C_2-O_{17}	0.0682	36	C_{14}-S_{18}-C_{19}-H_{20}	0.1427
12	C_4-C_5-C_6-H_8	0.0627	37	C_{14}-S_{18}-C_{19}-H_{21}	0.1568
13	C_4-C_5-C_{11}-C_{12}	–0.0810	38	C_{14}-S_{18}-C_{19}-H_{22}	0.0944
14	C_4-C_5-C_{11}-H_{13}	0.2084	39	H_{15}-C_{12}-C_{14}-O_{16}	0.2133
15	C_5-C_4-C_3-H_9	0.0904	40	H_{15}-C_{12}-C_{14}-S_{18}	0.1061
16	C_5-C_{11}-C_{12}-C_{14}	0.1245	41	O_{16}-C_{14}-S_{18}-S_{19}	0.1605
17	C_5-C_{11}-C_{12}-H_{15}	0.2780			
18	C_6-C_1-C_2-C_3	0.0883			
19	C_6-C_1-C_2-O_{17}	–0.0262			
20	C_6-C_5-C_4-H_{10}	0.1237			
21	C_6-C_5-C_{11}-C_{12}	0.0449			
22	C_6-C_5-C_{11}-H_{13}	0.2163			
23	H_7-C_1-C_2-C_3	0.0481			
24	H_7-C_1-C_2-O_{17}	0.0874			
25	H_7-C_1-C_6-C_5	0.0313			

[a] The effective distance was calculated as in eq 9 of the main text without any tolerance parameter, and the state-averaged interpolation error was defined as the RMS value of the errors from both S_0 and S_1 states. Each correlation coefficient was obtained with the test configurations.

[b] Atom designations can be found in Figure 1 of the main text.

Table S7. Pearson's correlation coefficients between the squared effective distance from the closest data point and the state-averaged interpolation error for the selected improper dihedral coordinates.[a]

Index	Dihedral Angle[b]	Coefficient
1	C_1-C_3-C_2-O_{17}	0.0983
2	C_1-C_5-C_6-H_8	−0.0034
3	C_2-C_4-C_3-H_9	0.1179
4	C_3-C_5-C_4-H_{10}	0.0390
5	C_4-C_6-C_5-H_{11}	0.0472
6	C_5-C_{12}-C_{11}-H_{13}	0.2367
7	H_7-C_1-C_6-C_2	0.0584
8	C_{12}-O_{16}-C_{14}-S_{18}	0.2310
9	C_{14}-C_{11}-C_{12}-H_{15}	0.2163

[a] The effective distance was calculated as in eq 9 of the main text without any tolerance parameter, and the state-averaged interpolation error was defined as the RMS value of the errors from both S_0 and S_1 states. Each correlation coefficient was obtained with the test configurations.

[b] Atom designations can be found in Figure 1 of the main text.

Table S8. RMS errors in interpolated energies (eV) for the hopping geometries from IM/MM NAMD simulations.

Level of theory	S_0 state	S_1 state	S_0-S_1 gap
SA-CASSCF	0.074	0.098	0.056
α-CASSCF	0.042	0.044	0.020
MS-CASPT2[a]	0.131	0.146	0.033

[a] In the case of the MS-CASPT2, the dual-interpolation scheme was adopted, with Hessians from α-CASSCF.

Supporting Figures

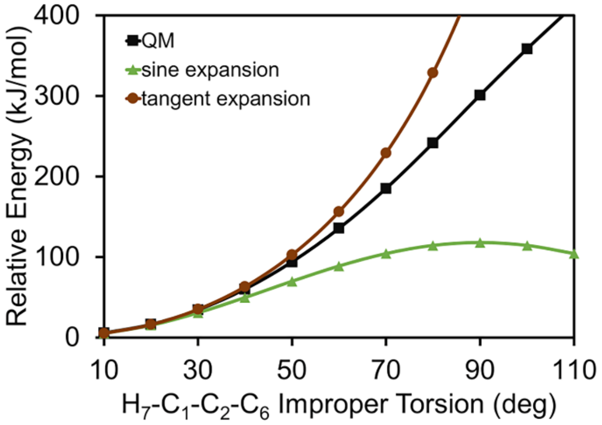

Figure S1. Taylor expansions calculated around the S_0-optimized geometry using different displacement function for improper torsions. The reference quantum chemical calculation results with SA2-CAS(4,3)SCF are also shown for comparison.

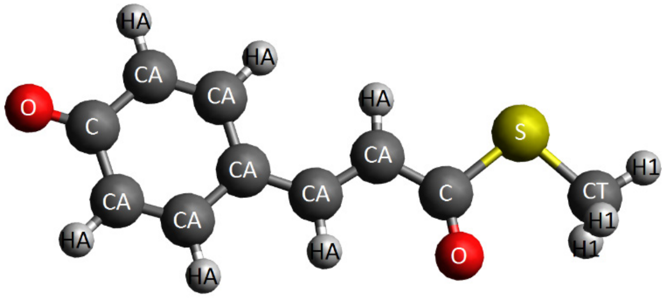

Figure S2. Assigned AMBER99SB bond atom types[4] for the stabilizer function of the PYP chromophore.

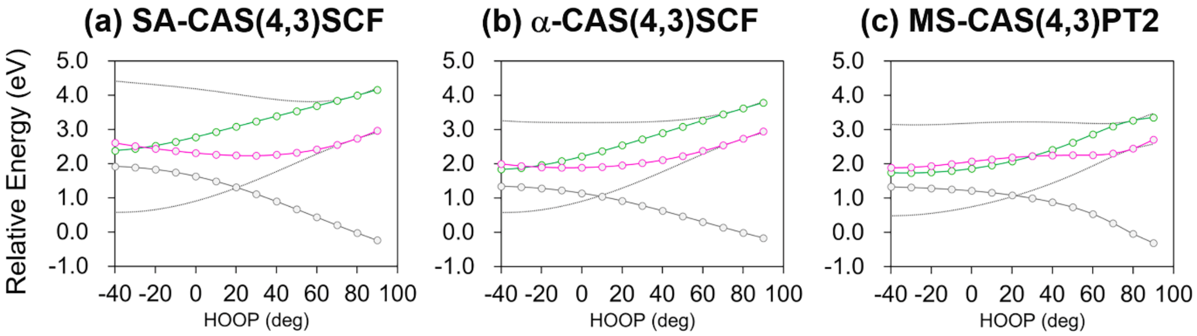

Figure S3. Diabatic potential energy curves along the HOOP coordinate, obtained with (a) SA-CASSCF, (b) α-CASSCF, and (c) MS-CASPT2 levels of theories. The initial geometry (corresponding to HOOP = 0 deg) was generated from the S_0-optimized one by rigidly rotating the β torsional angle to 130 deg. Other details are the same as in Figure 3 of the main text.

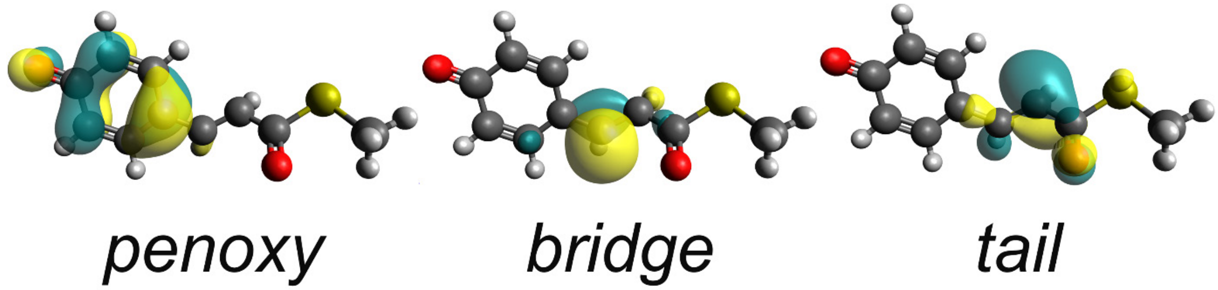

Figure S4. Boys-localized active space orbitals in the HOOP-twisted geometry. The geometry was generated from the S_0-optimized one by rigidly rotating β and HOOP torsional angles into 130 deg and 90 deg, respectively. The definition of β and the HOOP torsional angle can be found in Figure 1 of the main text.

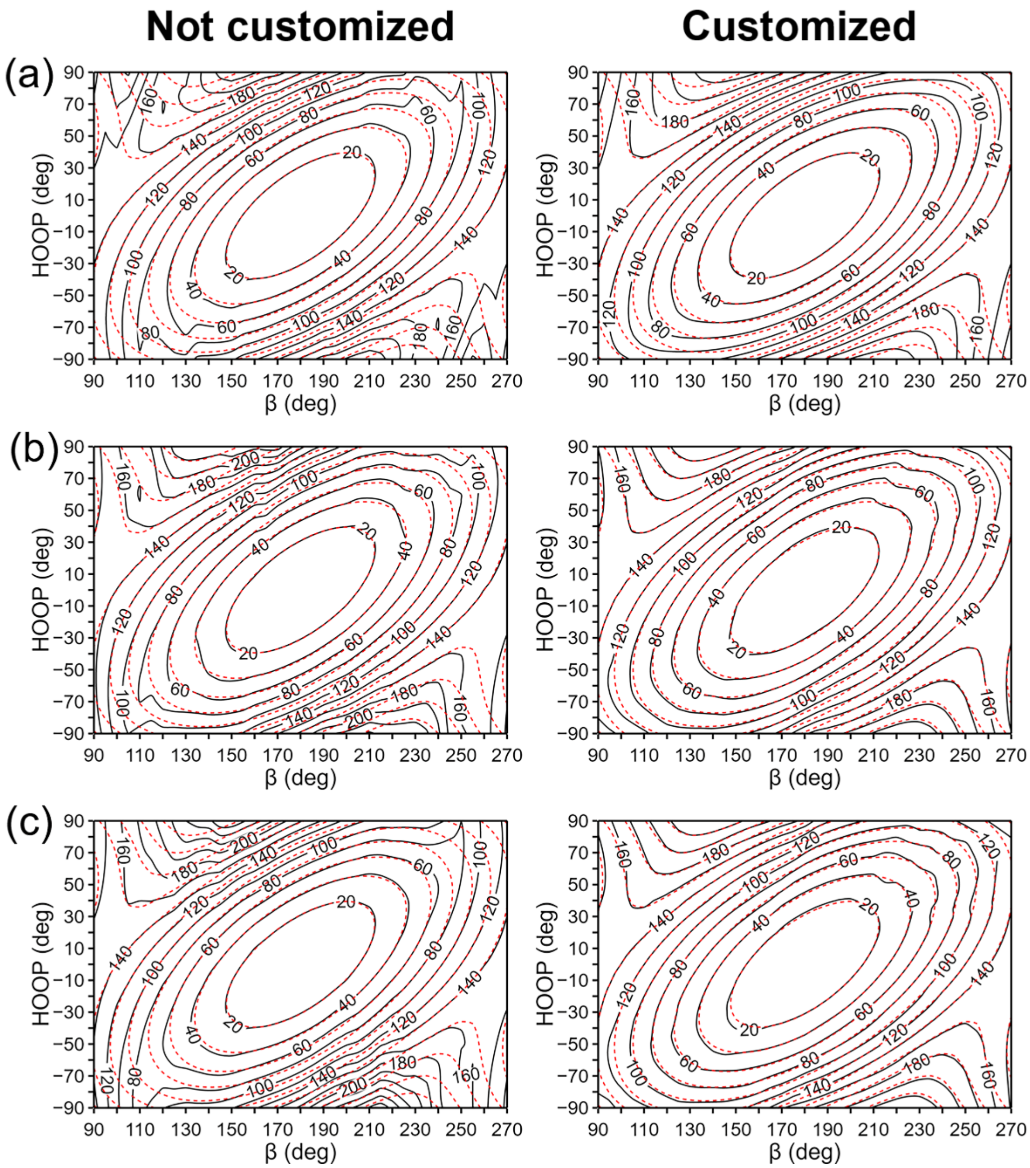

Figure S5. Contour plots of the interpolated (solid black) and the reference QM (dashed red) S_0 PES's in kJ/mol around the Franck-Condon region, without using the target-customized weighting coordinate (left panels) and with using it (right panels). Other details are the same as in Figure 5 of the main text.

Supporting References

1. R. P. A. Bettens and M. A. Collins, *J. Chem. Phys.*, 1999, **111**, 816-826.
2. C. R. Evenhuis, X. Lin, D. H. Zhang, D. Yarkony and M. A. Collins, *J. Chem. Phys.*, 2005, **123**, 134110.
3. C. W. Kim and Y. M. Rhee, *J. Chem. Theory Comput.*, 2016, **12**, 5235-5246.
4. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct. Funct. Bioinform.*, 2006, **65**, 712-725.