Supporting Information

for

Two-State Three-Mode Parameterization of the Force Field of a Retinal Chromophore Model

Emanuele Marsili[†], Marwa H. Farag[¶], Xuchun Yang[%], Luca De Vico[†], and Massimo Olivucci[†,%,#,*]

†Dipartimento di Biotecnologie, Chimica e Farmacia, Università di Siena, via A. Moro 2, I-53100 Siena, Italy, ¶Department of Chemistry, University of Southern California, Los Angeles, CA 90089-0482, USA, %Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA and #Institut de Physique et Chimie des Matériaux de Strasbourg, UMR 7504 Université de Strasbourg-CNRS, F-67034 Strasbourg, France.

Contents:

S1. Rigid HOOP (Φ) scan

The parameter values shown in Table 1 in the main text are obtained by fitting the full dataset plus additional 144 dataset points obtained via 24 rigid scans along the Φ coordinate. To obtain the 144 dataset points we have employed the following protocol. (i) We have picked equally spaced points along the IRC for the MEPcis, MEPtrans, MEPct and MEPdir paths (see article, section 1). From the first two paths, we have selected 7 points, while for the last two, we have chosen 5 points (i.e. in total 24). These 24 starting geometries were employed to generate the HOOP scans. (ii) We have performed a rigid scan along the Φ coordinate leaving r and θ unchanged. For each starting geometry, we have modified the Φ reducing and increasing it of 7, 14 and 21 degrees, generating 6 new geometries that differ only for the Φ angle (see Fig. S1). Finally, the generated new points (144 in total) were added to the reference dataset (see point IV of section 2.4 in the main text) to ensure that the model describes properly the Φ dependence. Figures S1, S2, S3 and S4 show the relative energies for both ground (in blue), and excited (in red) states for MEPcis, MEPtrans, MEPct and MEPdir HOOP-scans, respectively. Each panel of figure S1-S4 has in abscissa the Φ value while in ordinates the corresponding energy (zero is assigned to *trans*-PSB3) of each geometry. The reference point (i.e. the starting geometries) is shown in the middle of the plot (notice that it does not always coincide with Φ = 0°). The circles show the computed energies while the squares correspond to the simulated energies. Figure S1 shows the HOOP-scans performed along the MEPcis reaction path. It starts with MEPcis000, which coincides with FCcis, and it ends with MEPcis024, a geometry close to the CIcis. Those points in-between belong to the IRC traced on the excited state. They have a structure that evolves gradually from a *cis*-PSB3-like till the first conical intersection (see Gozem, S. et al. Chem. Theory Comput. 2012, 8, 4069-4080 and Gozem, S. et al. J. Chem. Theory Comput. 2013, 9, 4495-4506). Figure S2 shows the HOOP-scans along MEPtrans reaction path, MEPtrans000 coincides with CItrans while MEPtrans024 is a structure close the FCtrans. Finally Figures S3 and S4 display the HOOP-scans of MEPct and MEPdir, respectively. MEPct008 and MEPdir008 are the two TSs while the other geometries corresponds with the evolution towards reactant and product structures. It should be noticed that only a small portion of the reaction path is computed, MEPct000 and MEPdir000 do not

correspond to the point of minima as well MEPct016 and MEPdir016. We have found that the simulated values fit reasonably well the point calculated at the XMCQDPT2 level of theory especially in the vicinity of the reference points and along the MEPcis and MEPtrans paths (figure S1 and S2). Slightly worse results are obtained for MEPct and MEPdir rigid scans (Figure S3 and S4) where some deviation of 5 kcal/mol and few of them above 7 kcal/mol (see section S3).

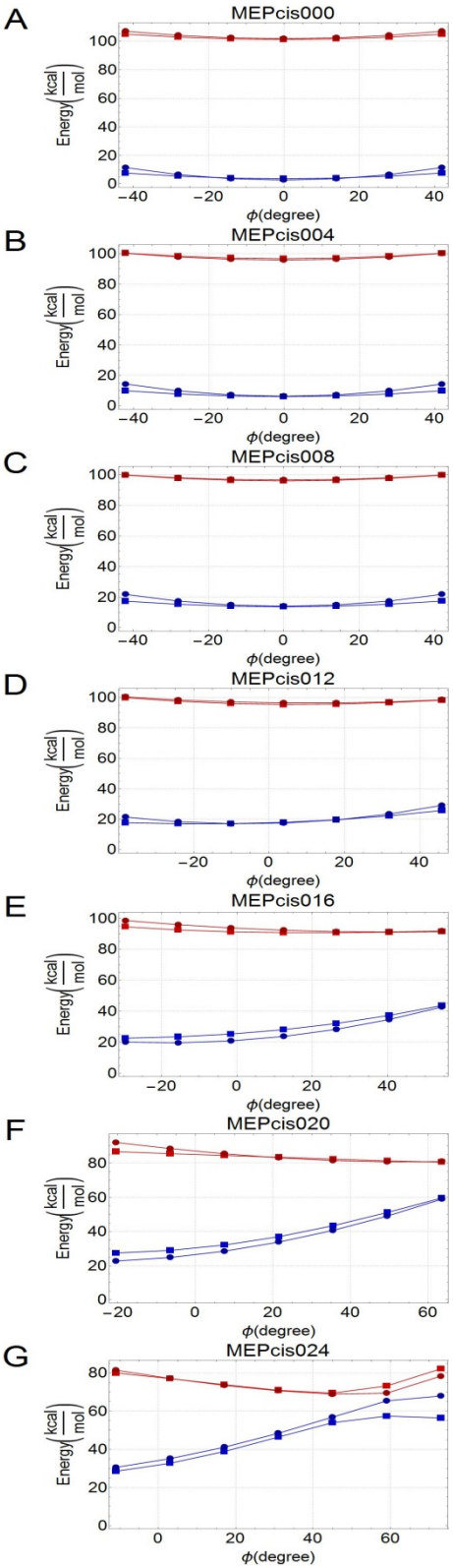

Figure S1. MEPcis HOOP scans. MEPcis000 (A) represent the FCcis while MEPcis024 (G) is just before the first conical intersection CIcis.

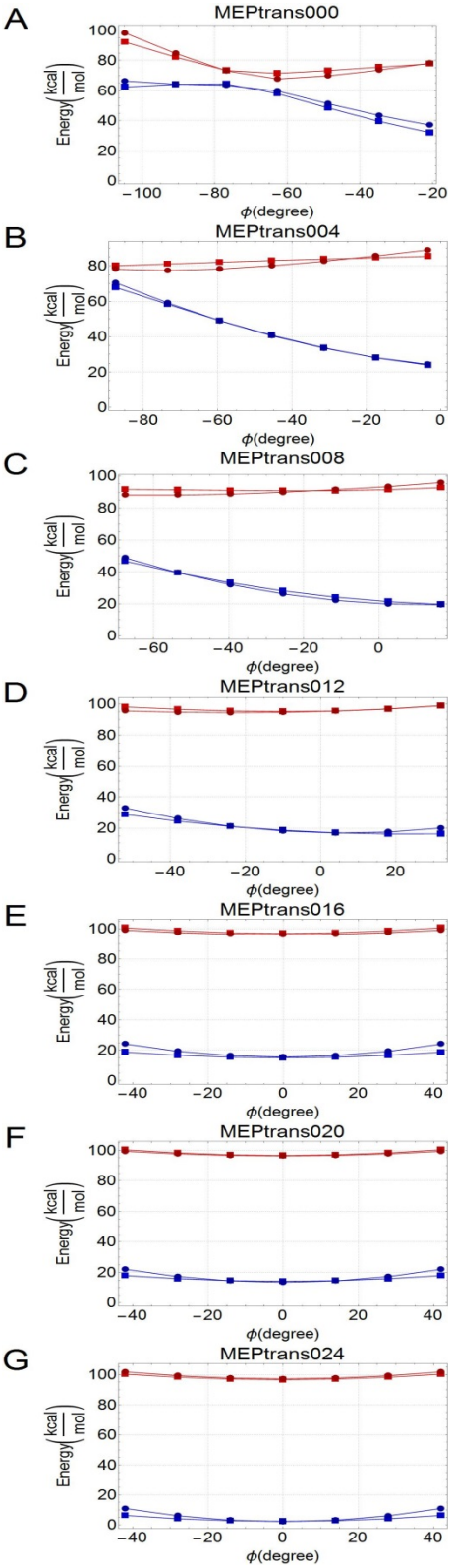

Figure S2. MEPtrans HOOP scans. MEPcis000 (A) represent the CItrans (noticeable is the energy degeneration between S_0 and S_1) while MEPcis024 (G) is just before the FCtrans

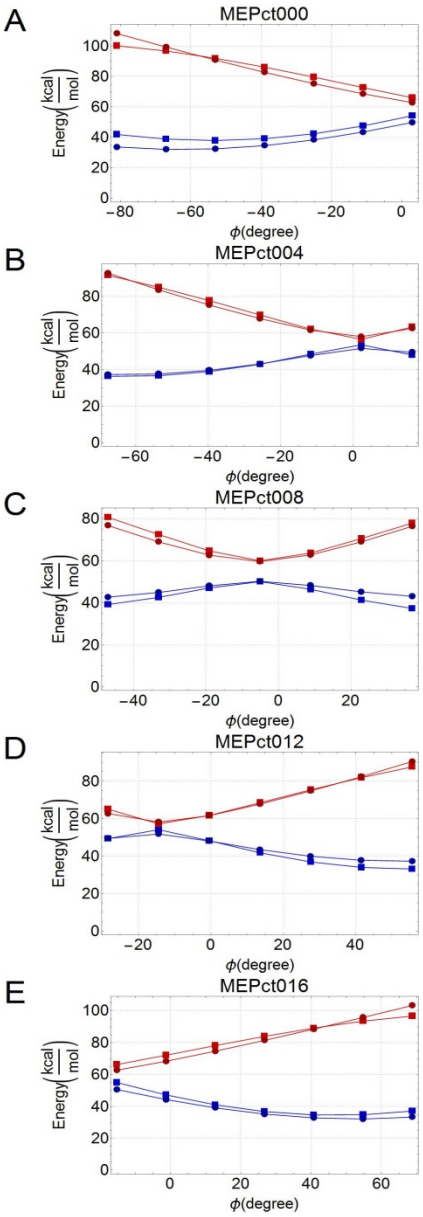

Figure S3. MEPct HOOP scans. MEPct000 (A) and MEPct016 (E) represent the ending of the path while MEPct008 (C) middle point is the TS

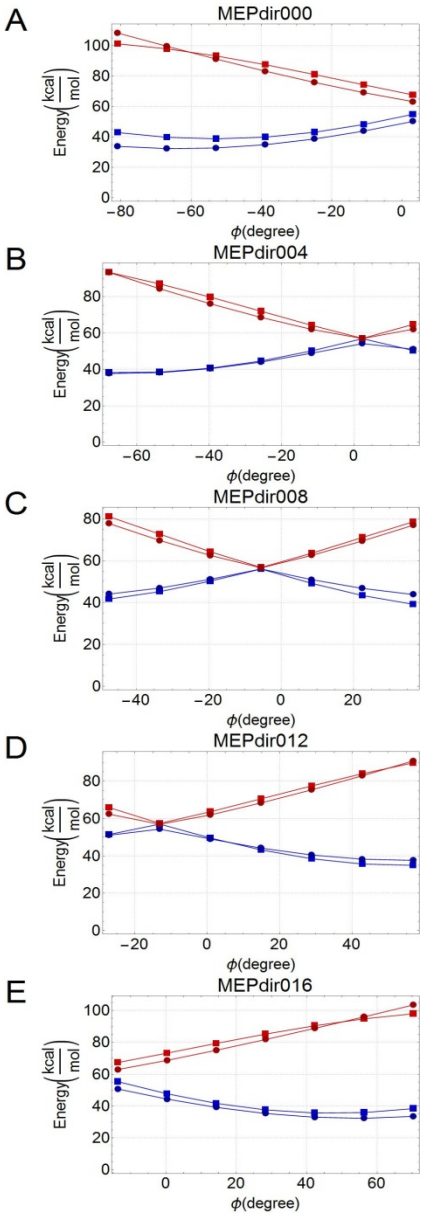

Figure S4. MEPdir HOOP scans MEPdir000 (A) and MEPdir016 (E) represent the ending of the path while MEPdir008 (C) is the TS, notice that in this case it coincides with the CI at minimum energy.

S2. Starting values for the fitting procedures

This test was made for ensure that the result obtained is solid and it does not change for small variations of the starting values. We report in Table S1 three sets of starting values, varying especially those belonging to Hcp, Hct_{corr} and $Hdir_{corr}$, that bring, after optimization via *FindFit* and *NonLinearModelFitting* the same optimized parameters shown in Table 1 (in the main text). Special care must be taken for the selection of hc_1, hc_2, hd_1 and hd_2. For instance, the use of negative starting parameters can alter the final fitting result (sometime obtaining very similar outcome but, in other cases, converging to a local solution yielding an higher standard deviation).

Table S1. Three different combination of starting parameters that bring the same result optimized parameter reported in Table 1 in the main article. Other combinations are possible but they are not reported.

	1	**2**	**3**
Hdir2D			
d_1	2300	-40	1200
d_2	50	200	-10
d_3	3.72	-10	12
d_4	-596	1000	-300
Hct2D			
c_1	500	1	100
c_2	6	100	36
c_3	0.01	-25	15
c_4	0.5	345	30
c_5	3.5	1	2
$Hdir_{corr}$			
hd_1	40	70	4
hd_2	20	150	60
Hct_{corr}			
hc_1	18	1	30
hc_2	60	10	7
Hcp			
k_1	20	10	50
k_2	30	1	-1

S3. Deviation between model and dataset.

In this chapter we discuss, briefly, the energy differences between the energy computed at the XMCQDPT2 and simulated energy values (residuals).

S3.1 Result for Hdir2D and Hct2D.

The first two diabatic functions are fitted using 36 dataset points dataset (see section 2.4 in the main text). These functions fit reliably as demonstrated by analyzing the residuals in Figure S5. Panel A and B show the residuals between computed data and Hdir2D and Hct2D, respectively. The absolute energy difference between dataset and simulated values is very low yielding a model with high accuracy making a small error of 1 kcal/mol.

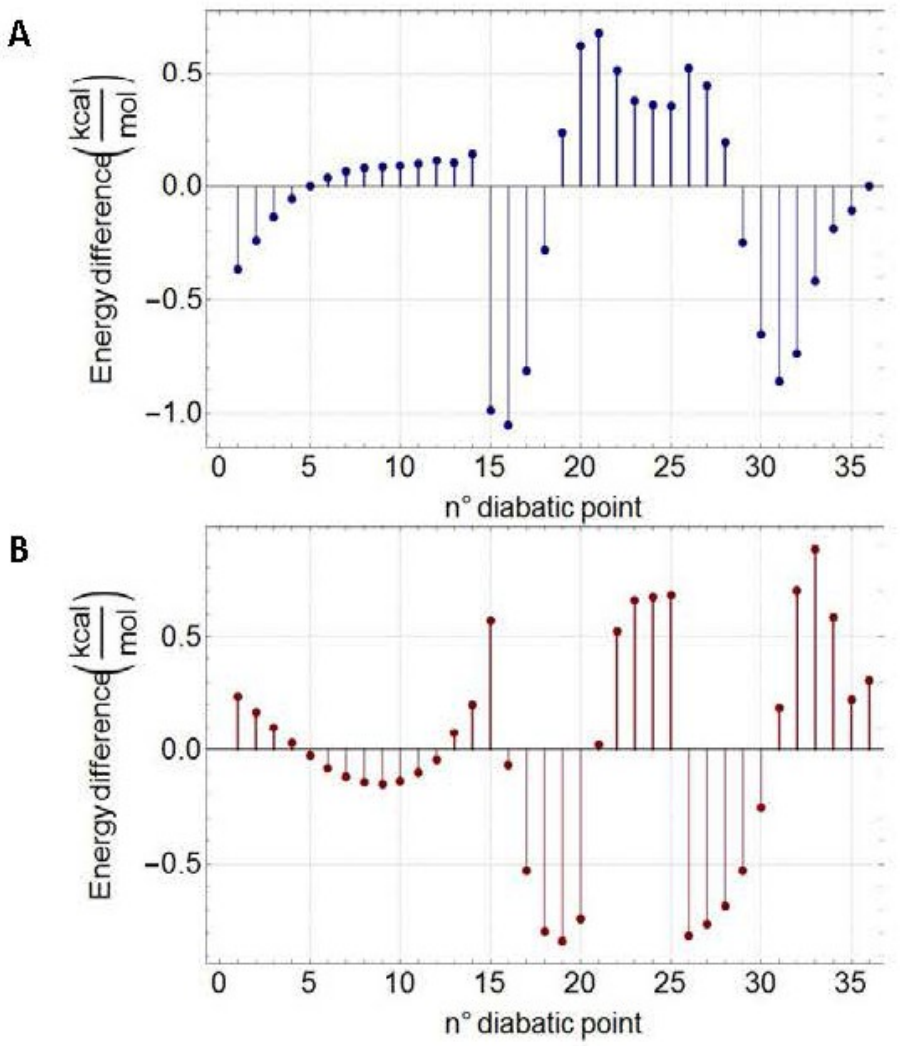

Figure S5. Energy residues between computed and simulated points belonging to the diabatic dataset. A and B is referring Hdir2D and Hct2D model respectively.

Table S2. Statistical values for the residues displayed in Figure S5.

	Hdir2D	**P-value**	*Hct2D*	**P-value**
SD	0.450	-	0.484	-
Variance	0.203	-	0.234	-
	d1	$9.44\ 10^{-8}$	c1	$1.62\ 10^{-11}$
	d2	$2.80\ 10^{-54}$	c2	$1.70\ 10^{-10}$
	d3	$4.05\ 10^{-22}$	c3	$4.50\ 10^{-11}$
	d4	$3.20\ 10^{-39}$	c4	$5.10\ 10^{-41}$
	-		c5	$1.04\ 10^{-9}$

Statistical values such as Standard Deviation and variance are collated together with reliability parameter (p-value) in Table S2. The p-values show the trustworthiness of each parameter: values lower than 0.05 should be considered accurate. We remind that, for the results interpretation, lower is the p-value, higher is the reliability of one parameter in the model.

S3.2. Result for S_0 and S_1 adiabatic states.

The two adiabatic functions are fitted using 257 geometries belonging to the reference dataset plus the 144 points of HOOP scans (see section 2.4, point IV in the main text). To validate the correctness between computed and simulated energies, we report in Figure S6 the residuals, while the same statistical considerations done in section S3.1 (Supporting Information), are reported in Table S3. Standard Error and t-statistic were also determinated but are not reported. The model fit with a good agreement the points belong to the six paths while it less accurate for the last 144 HOOP-scan points in both S_0 (Figure S6 panel A) and S_1 (Figure S6 panel B).

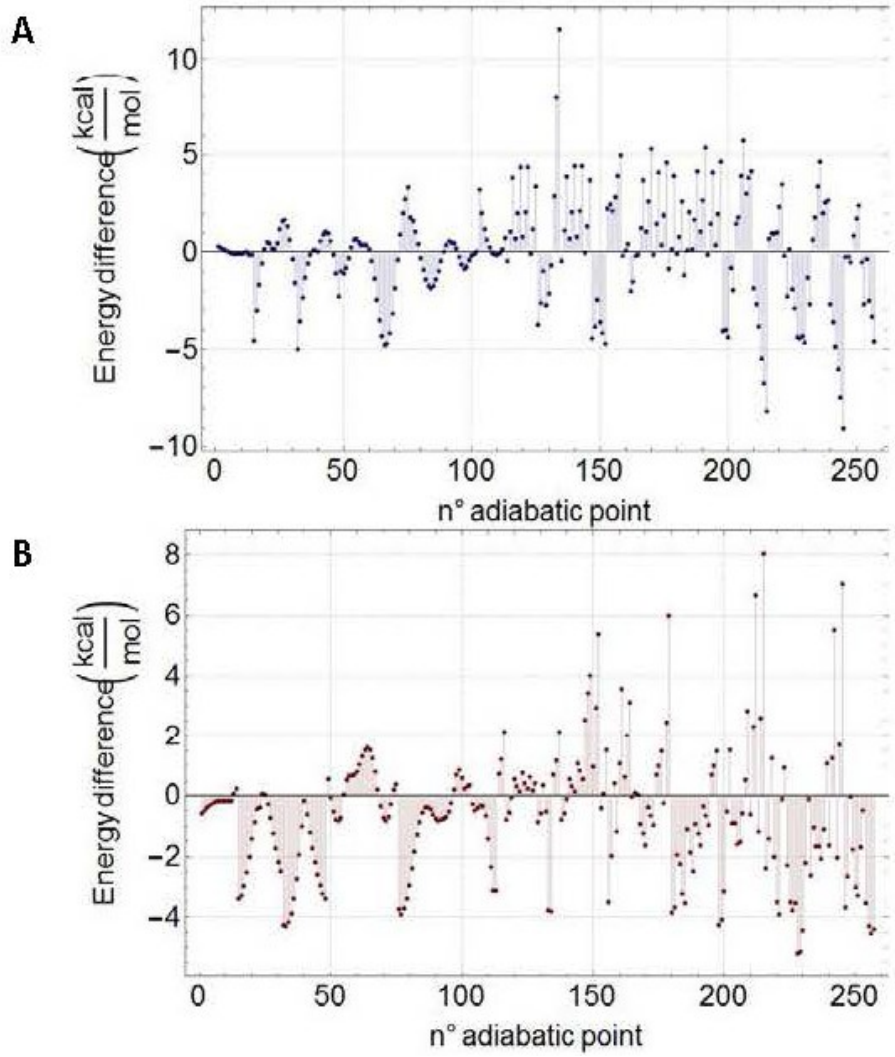

Figure S6. Energy residues between computed and simulated points belonging to the adiabatic dataset. The first 257 (panel A) and the last 257 (panel B) points are the residues between the computed energy reference and S_0 and S_1 adiabatic states simulated by the model, respectively. All points are used for the fitting simultaneously.

Table S3. Statistical values for the residues displayed in Figure S6.

	Adibatic States	**P-value**
SD	2.42	-
Variance	5.87	-
$Hdir_{corr}$	hd_1	$1.20\ 10^{-38}$
	hd_2	$3.27\ 10^{-65}$
Hct_{corr}	hc_1	$2.44\ 10^{-14}$
	hc_2	$1.22\ 10^{-25}$
Hcp	k_1	$4.15\ 10^{-59}$
	k_2	$5.46\ 10^{-18}$

S4. Energy profiles re-evaluation with expanded 3 states

The energy profiles along the six paths in this work (see Figure 3) were obtained at the 2-root state average XMCQDPT2//CASSCF/6-31G* level. To confirm that such results are not affected significantly by higher excited states, we re-evaluated the path energy profiles at the 3-root state average XMCQDPT2//CASSCF/6-31G* level. The obtained energy profiles, as displayed in Figure S7, shows that the second excited state S_2 is well separated from S_1 along all six paths, and there is no obvious difference of the S_0 and S_1 energy profiles between the 2-root and 3-root result, indicating that S_2 is not interfering with S_1 and S_0.

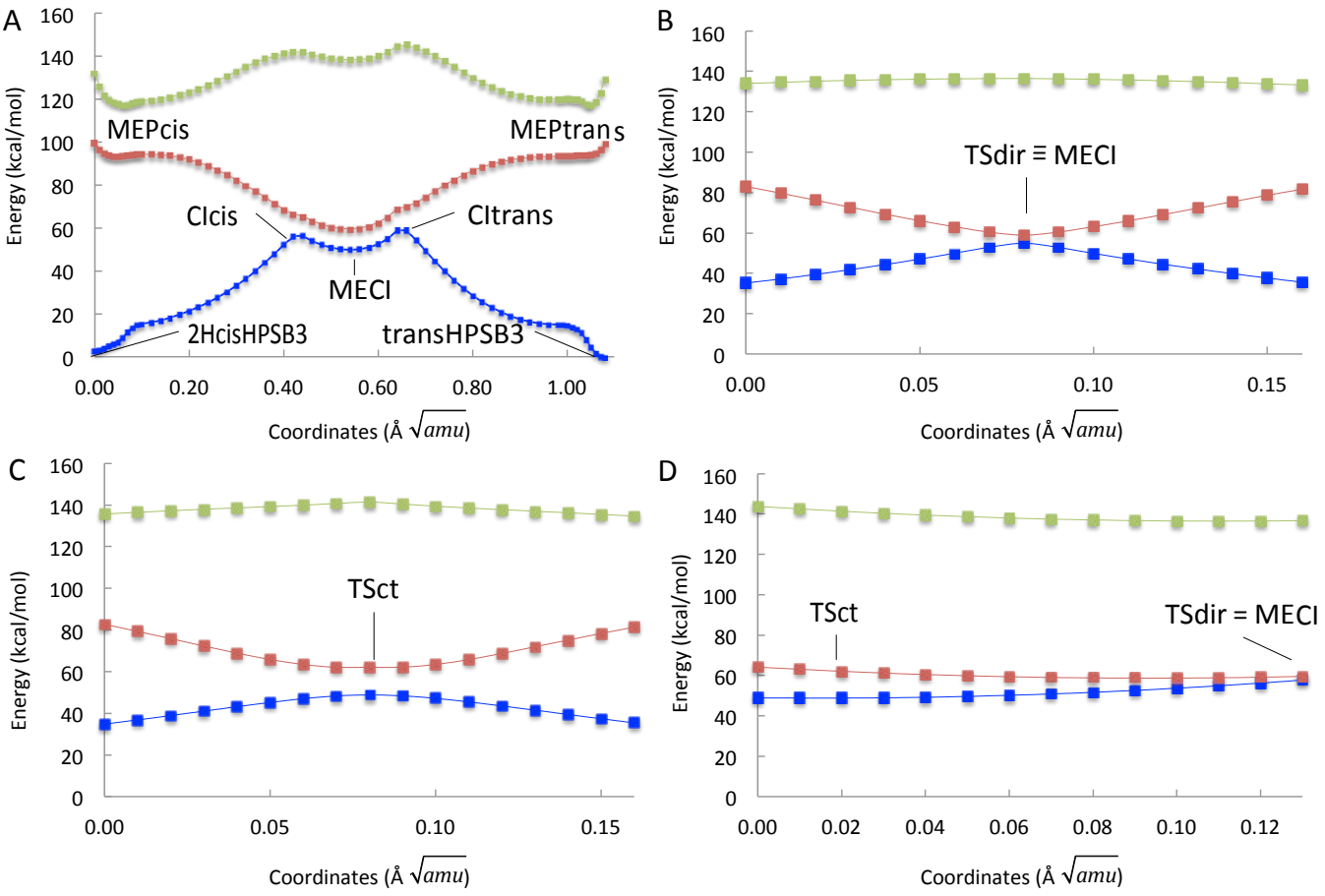

Figure S7. Simulated (squares) adiabatic energy profiles along the six paths considered in this work at 3-roots state average XMCQDPT2//CASSCF/6-31G* level. S_0, S_1 and S_2 profiles are in blue, red and green, respectively. A. MEPcis, MEPtrans and connecting IS profiles. The points include the 2-cis-PSB3 and all-trans-PSB3 equilibrium structures and the CIcis, CItrans and MECI conical intersections, B. MEPdir including the transition state TSdir. C. MEPct including the transition state TSct. D. BLAP energy including TSdir, TSct and MECI.