Supporting Information

A Theoretical Study on the UVR8 Photoreceptor:

Sensing Ultraviolet-B by Tryptophan and Dissociation of Homodimer

Xin Li,*,† Lung Wa Chung,§,‡ Keiji Morokuma,*,‡ and Guohui Li*,†

†State Key Lab of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Dalian 116023, China, ‡Fukui Institute for Fundamental Chemistry, Kyoto University, Kyoto 606-8103, Japan and §South University of Science and Technology of China, Shenzhen, Guangdong 518055, China

E-mail: lixindl@dicp.ac.cn; morokuma.keiji.3a@kyoto-u.ac.jp

Table of Contents

MD Simulation Details.

The molecular mechanics (MM) minimization was first performed to optimize the system. Then, the molecular dynamics (MD) simulations of 50 ps duration in the temperature ranges of 0-80, 80-160, 160-230, and 230-298 K (under constant volume conditions) were carried out to gradually heat the system. Additional 60 ps simulation was then performed in the NPT ensemble (at 298 K and 1 bar), followed by 300 ps simulation in the NVT ensemble. Subsequently, the long time-scale MD simulation in the NVT ensemble was carried out for data collection. The simulations have been conducted in periodic boundary conditions, with a cutoff radius of 12 Å for nonbonding interactions and the particle mesh Ewald method for calculating electrostatic interactions.[S1] Bond lengths involving hydrogen atoms were constrained by the SHAKE algorithm.[S2] The simulations were performed by using Leapfrog algorithm with a time step of 1 fs, and the Langevin thermostat algorithm were used to control the temperature.[S3] The protein structures taken from the production MD simulation (about 150-180 snapshots) were further refined by the ONIOM(QM:MM) geometry optimization.

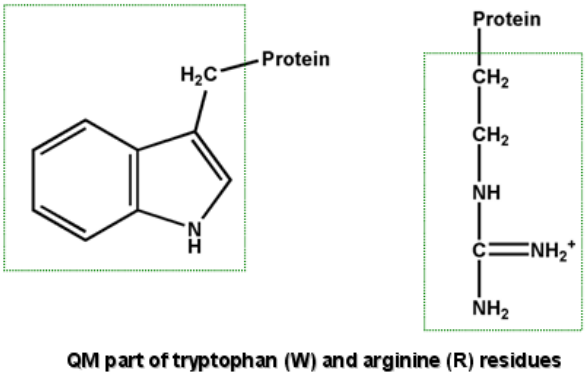

Figure S1. The QM parts of tryptophan (W) and arginine (R) residues in ONIOM(QM:MM) calculations.

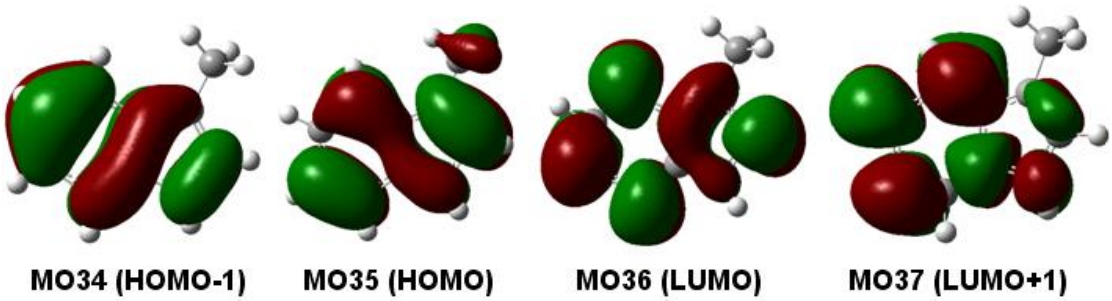

Figure S2. The key orbitals of 3-methylindole molecule in the gas phase, calculated by the HF/D95(d) method.

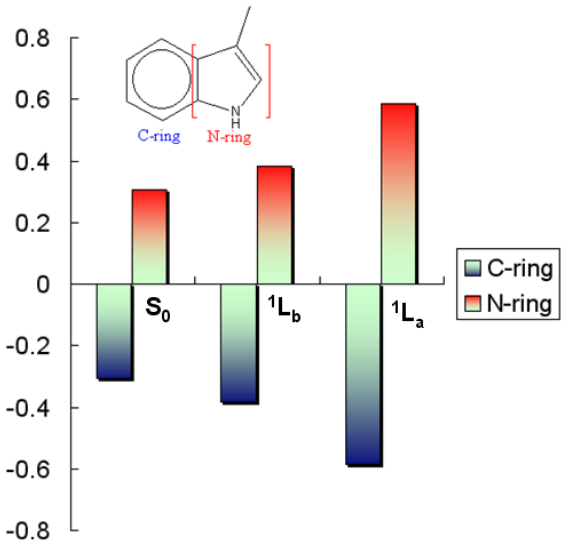

Figure S3. The SAC-CI calculated Mulliken charge distribution of 3-methylindole in the gas phase, based on the M06-2X optimized ground-state geometry.

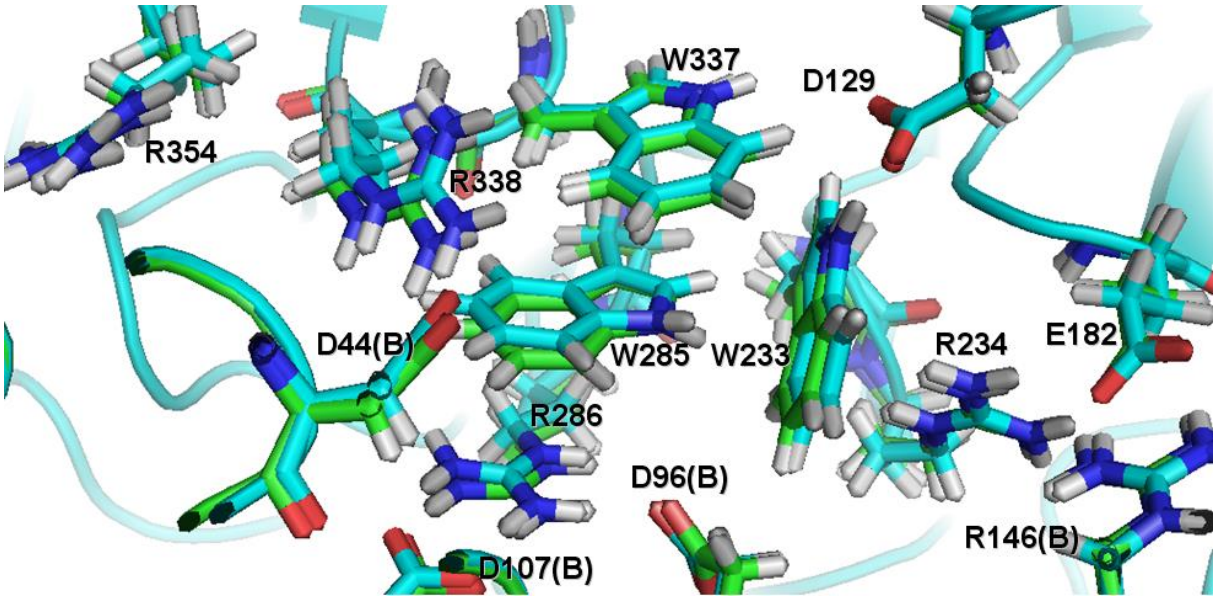

Figure S4. Superposition of the x-ray crystal structure (in green) and the ONIOM minimized structure of model **G** (in blue).

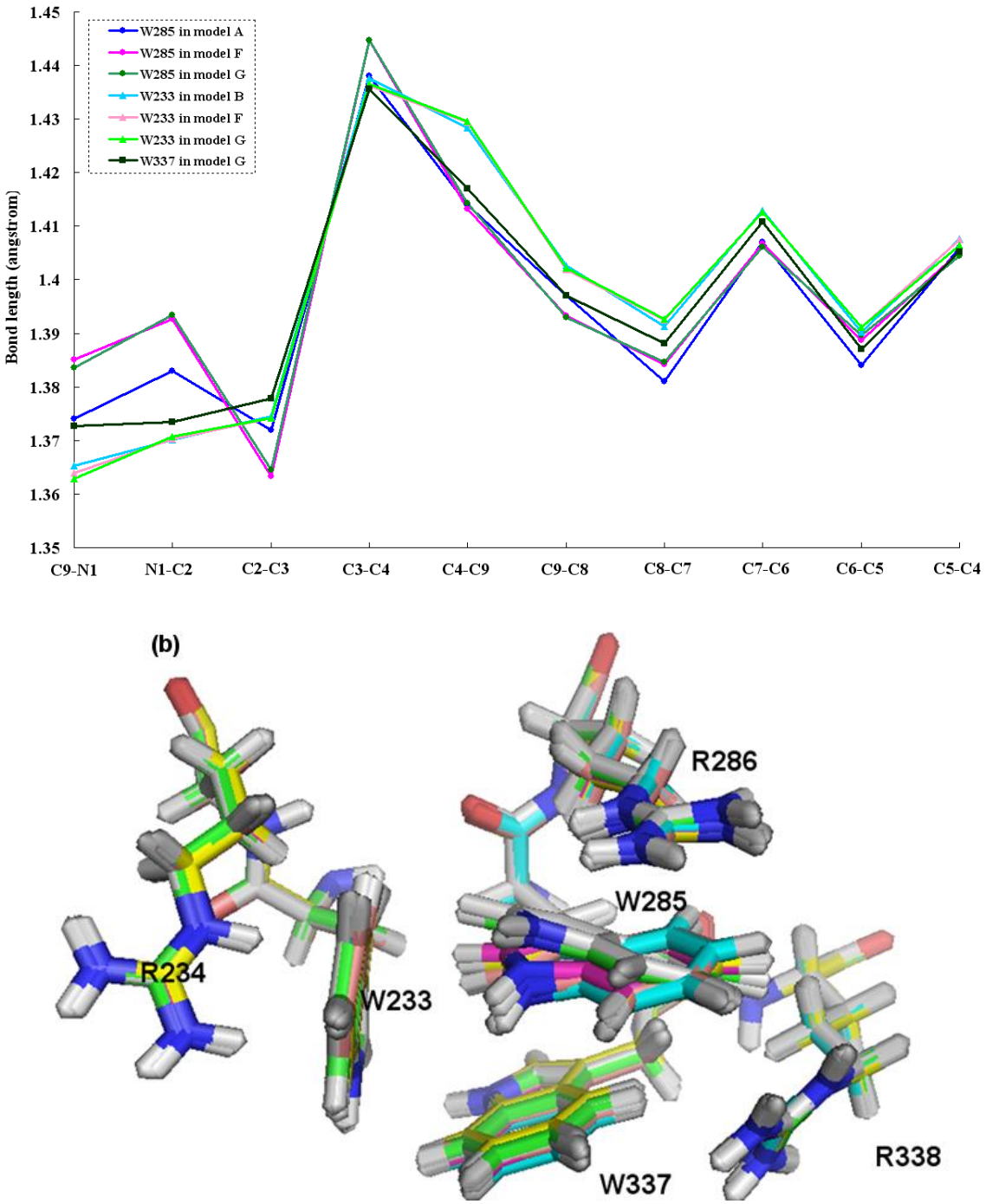

Figure S5. (a) Bond lengths of the ground-state indole moiety, and (b) the superposition of the local indole structures and their nearby arginine residues in UVR8, optimized by the ONIOM(M06-2X:AMBER)-EE method, for models **A** (with $W285^{qm}$, green), **B** (with $W233^{qm}$, cyan), **C**: (with $W285^{qm}$+$R286^{qm}$, pink), **D**: (with $W285^{qm}$+$R338^{qm}$, gray), **F**: (with $W285^{qm}$+$W233^{qm}$, magenta), and **G**: (with $W285^{qm}$+$W233^{qm}$+$W337^{qm}$, yellow).

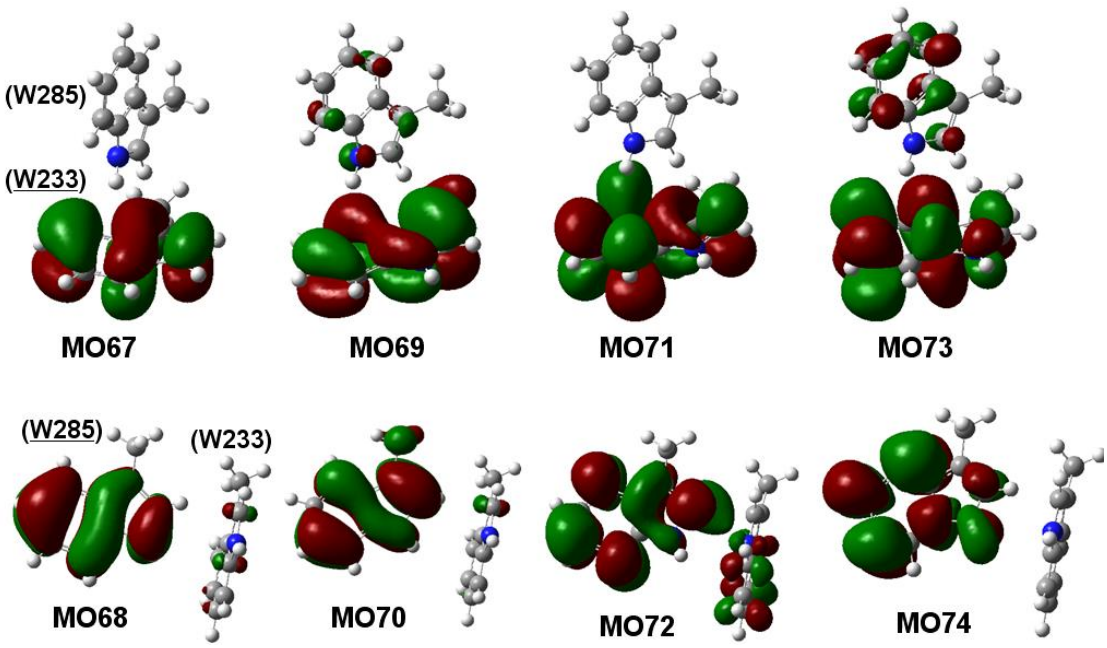

Figure S6. The HF/D95(d) calculated key orbitals of $W285^{qm}$+$W233^{qm}$ (model **F**) in UVR8.

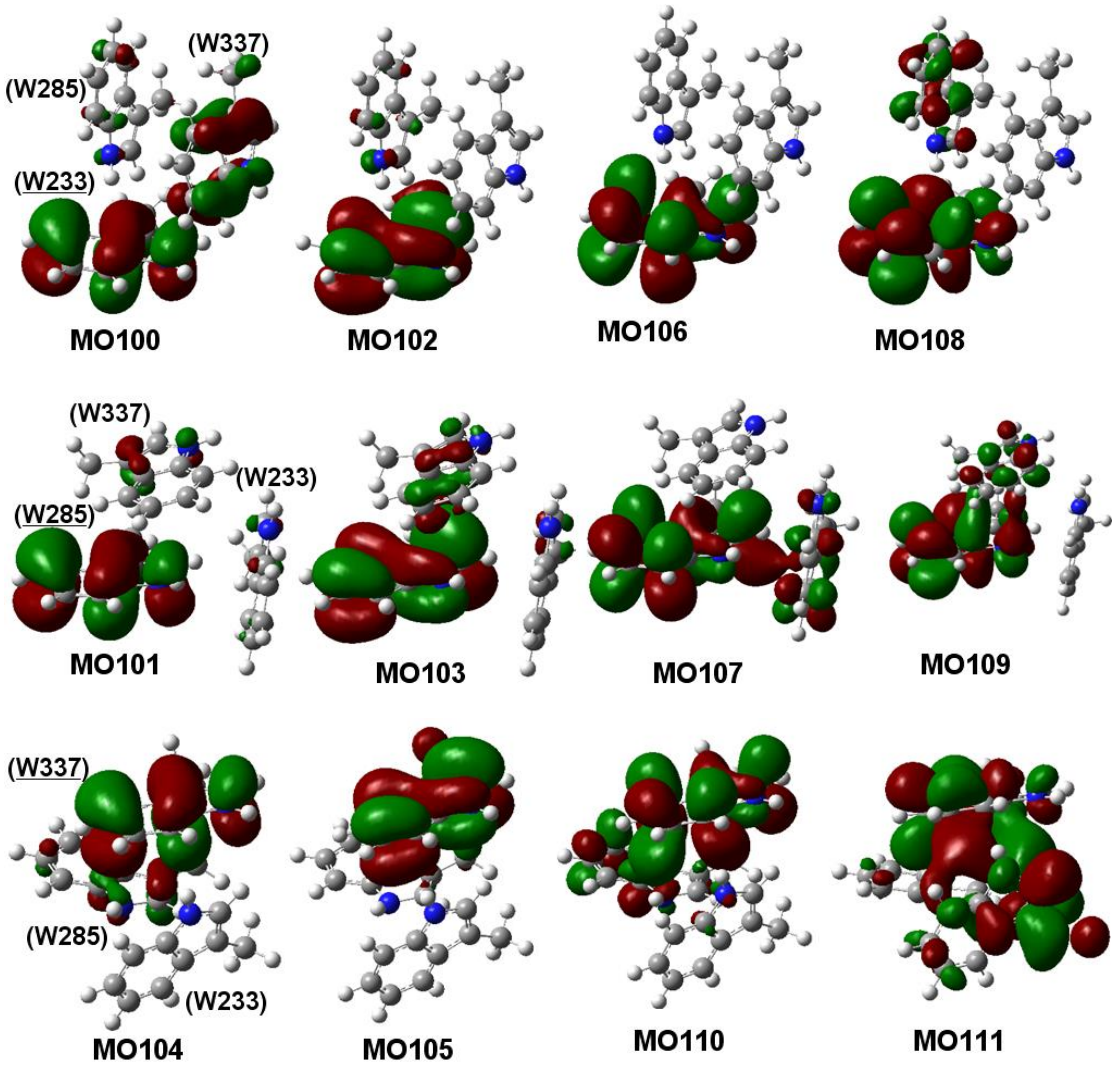

Figure S7. The HF/D95(d) calculated key orbitals of $W285^{qm}$+$W233^{qm}$+$W337^{qm}$ (model **G**) in UVR8.

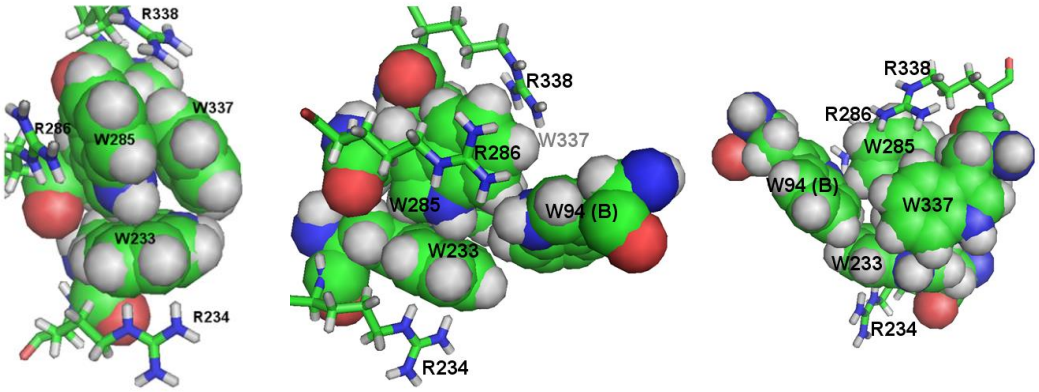

Figure S8. The local structure of triad tryptophans in the A-chain of UVR8.

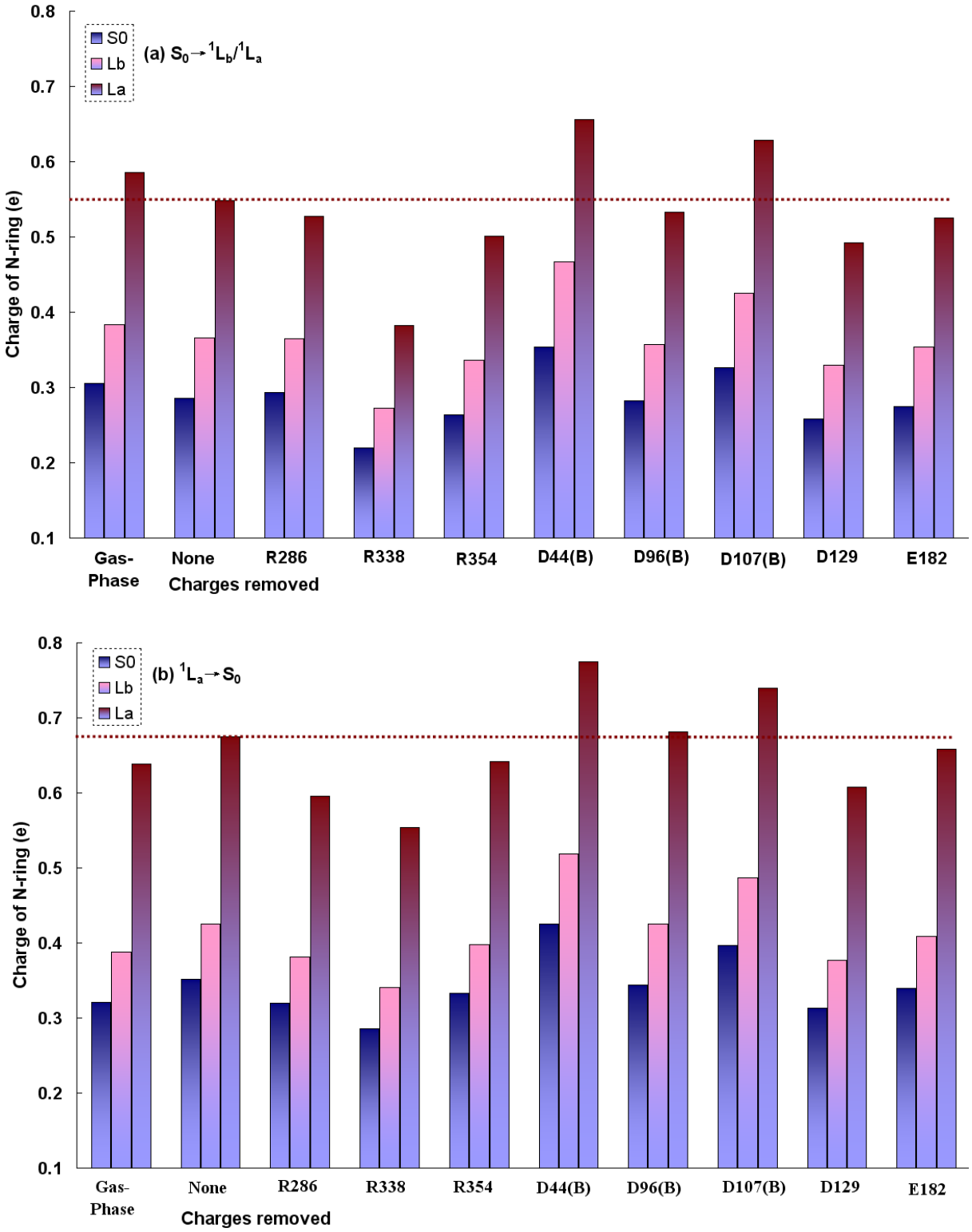

Figure S9. The effect of the key positively and negatively charged residues on the charge distribution of the pyrrole moiety (N-ring) of W285 in (a) the S_0 state and (b) the ^1L_a excited-state equilibrium geometries, compared with 3-methylindole in the gas phase.

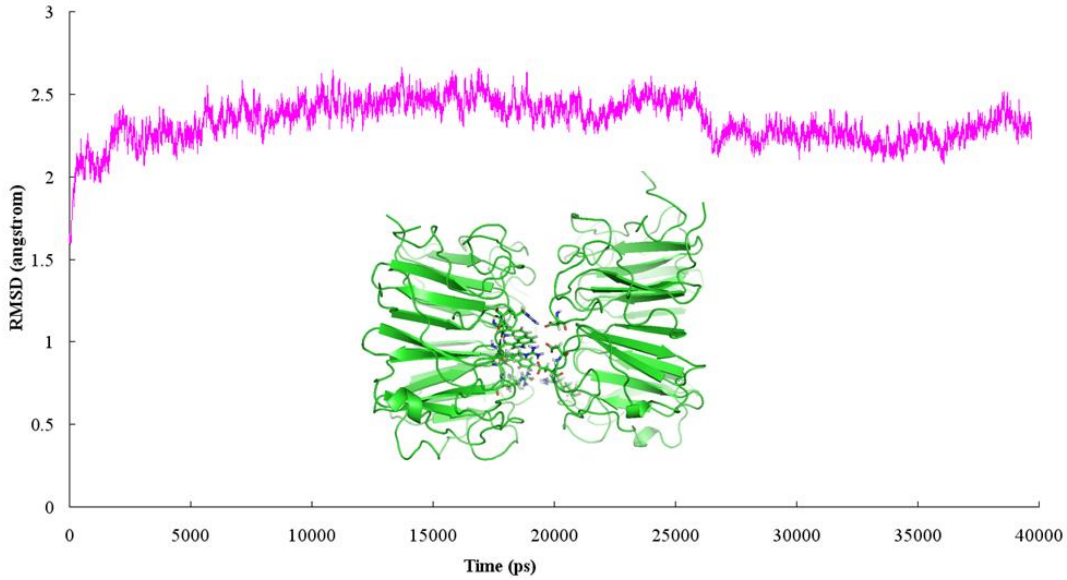

Figure S10. The time evolution of RMSD for all residues.

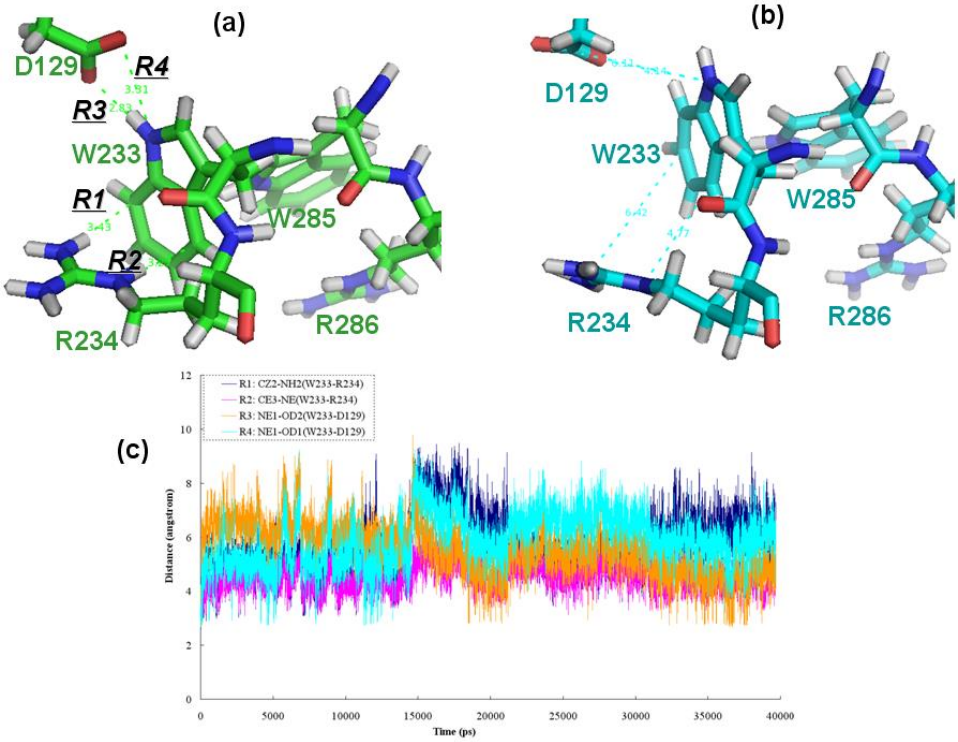

Figure S11. The local structure around W233 in the A chain of UVR8, (a) the initial structure (in green) and (b) one structure after more than 20 ns MD simulation (in cyan). (c) The atom-atom distances between W233 and R234/D129.

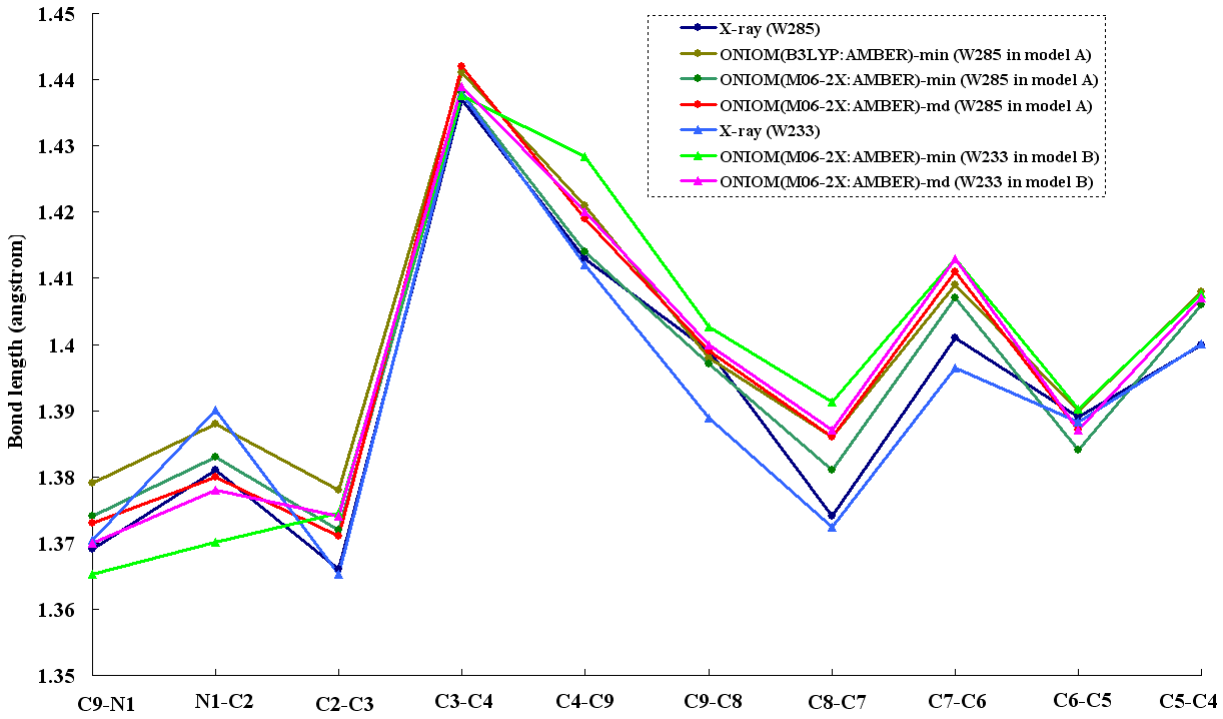

Figure S12. The comparison of bond lengths of W285 and W233 in the UVR8 crystal structure and ONIOM(QM:MM) optimized structures. “-min” denotes that the initial geometry is taken from the MM minimized protein structure. “-md” denotes the average bond lengths of tryptophan in the protein starting from a series of snapshots in MD simulations.

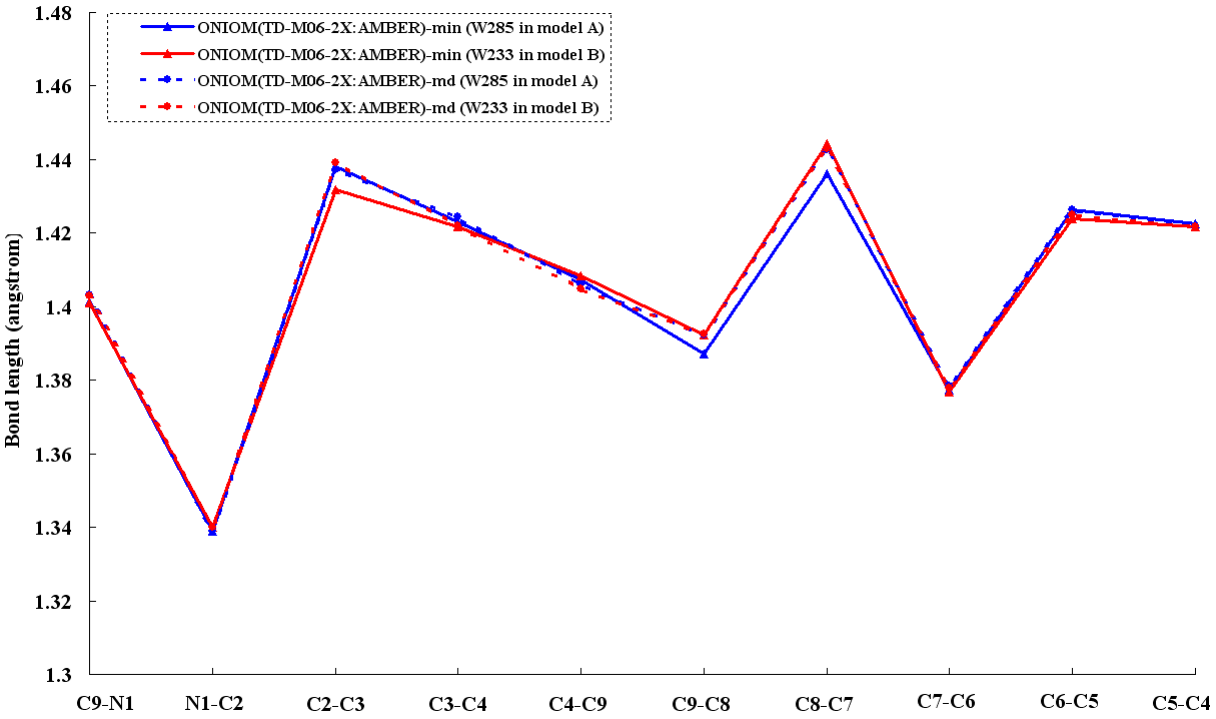

Figure S13. The ONIOM optimized ^1L_a excited-state bond lengths of W285 and W233 in UVR8. "-min" denotes that the initial geometry is taken from the MM minimized protein structure. "-md" denotes the average bond lengths of tryptophan in the protein starting from a series of snapshots in MD simulations.

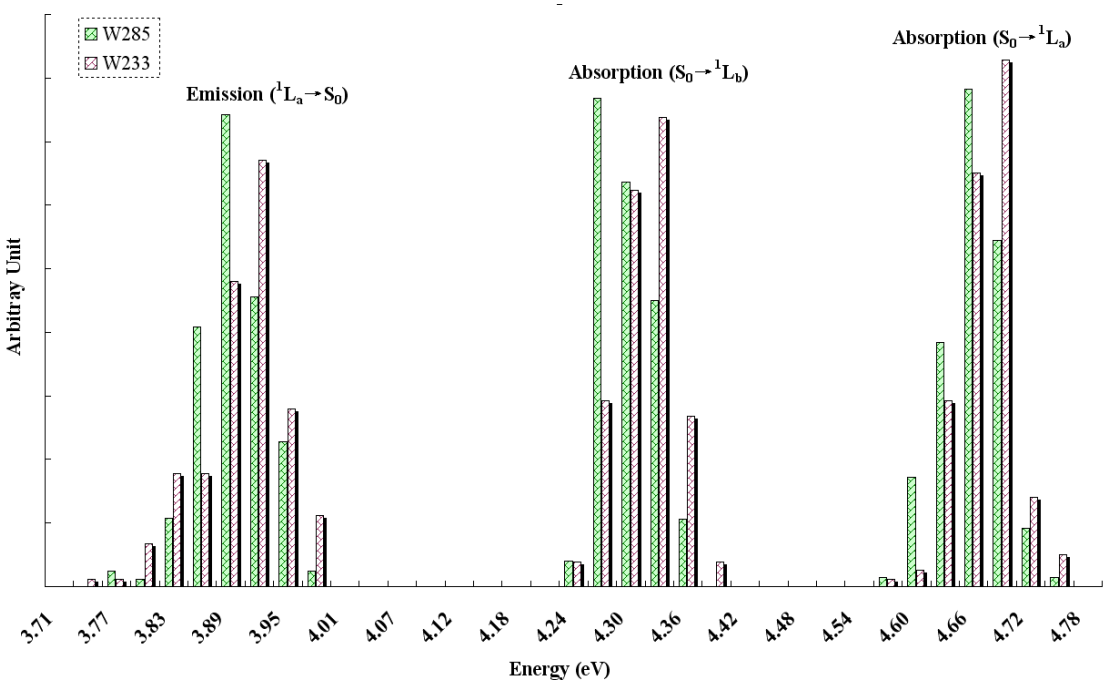

Figure S14. Histograms of distribution of the computed absorptions and emission of the bare W285 and W233 chromophores.

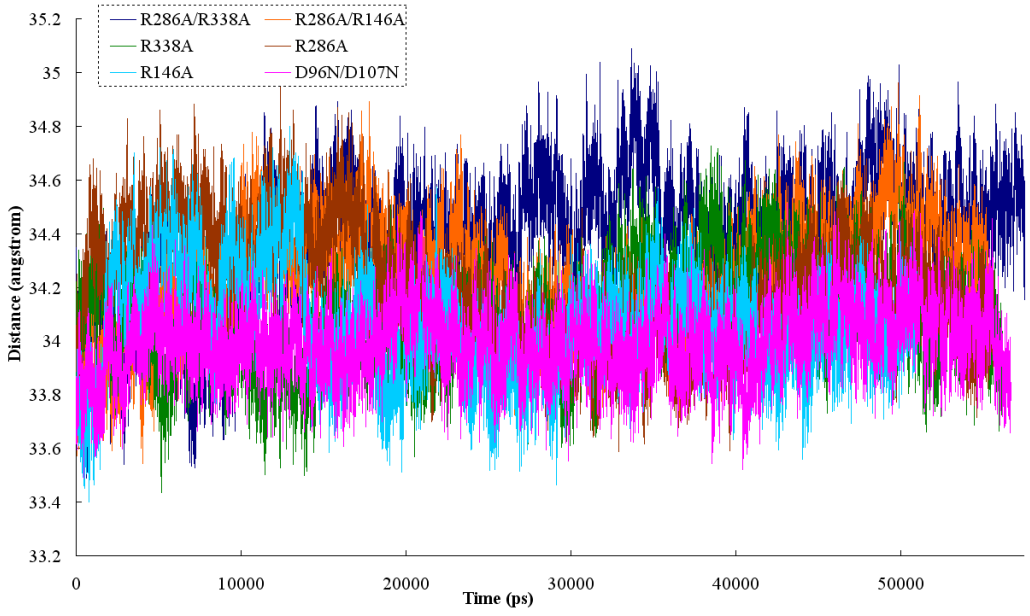

Figure S15. The time evolution of the center of mass (COM) distance between A and B monomers in key mutants.

Table S1. The computed absorption energies (ΔE) and oscillator strengths (f) of 3-methylindole in the gas phase.

Methods	^1L_b state		^1L_a state	
	ΔE (eV)	f	ΔE (eV)	f
TD-M06-L/BS1//B3LYP/BS	4.76	0.01	4.42	0.06
TD-M06/BS1//M06-2X/BS	4.75	0.03	4.50	0.08
TD-M06-2X/BS//M06-2X/BS	5.09	0.05	4.95	0.09
TD-M06HF/BS1//M06-2X/BS	4.75	0.03	4.50	0.08
TD-M05/BS1//M06-2X/BS	4.84	0.03	4.59	0.08
TD-M052X/BS1//M06-2X/BS	5.17	0.05	5.03	0.09
TD-SOGGA11/BS1//M06-2X/BS	4.63	0.01	4.25	0.04
TD-N12SX/BS1//M06-2X/BS	4.89	0.03	4.63	0.08
TD-M11/BS1//M06-2X/BS	5.13	0.07	5.02	0.08
TD-MN12L/BS1//M06-2X/BS	4.79	0.04	4.33	0.06
TD-B2PLYP-D/BS2//M06-2X/BS	4.93	0.07	4.85	0.08
TD-B2PLYP-D3/BS2//M06-2X/BS	4.93	0.07	4.85	0.08

Table S2. The calculated dipole moment (μ) for 3-methylindole in the gas phase.

Methods	μ (D)		
	S_0	^1L_b	^1L_a
SAC-CI/BS3//B3LYP/BS	2.01	2.42	5.80
SAC-CI/BS3//M06-2X/BS	2.01	2.45	5.85
SAC-CI/BS3//CASSCF/BS4	2.00	2.50	5.77
CASPT2/BS4//M06-2X/BS	1.82	1.70	6.09
CASPT2/BS5//M06-2X/BS	1.84	1.72	6.20
CASPT2/BS6//M06-2X/BS	1.84	1.74	6.25
CASPT2/BS4//CASSCF/BS4	1.80	1.68	6.00
CASPT2/BS5//CASSCF/BS4	1.84	1.72	6.14
CASPT2/BS6//CASSCF/BS4	1.83	1.74	6.21

Table S3. The MS-CASPT2 computed absorption energies (ΔE) and oscillator strengths (f) of 3-methylindole in the gas phase, with different zero-order Hamiltonians.[a]

MS-CASPT2//CASSCF/6-31G*	^1L_b state		^1L_a state	
	ΔE (eV)	f	ΔE (eV)	f
IMAG=0.2, IPEA=0.0	4.46	0.02	5.03	0.09
IMAG=0.0, IPEA=0.2	4.75	0.02	5.35	0.09
IMAG=0.1, IPEA=0.25	4.81	0.02	5.41	0.09
IMAG=0.1, IPEA=0.0	4.46	0.02	5.03	0.09

a. For MS-CASPT2/CASSCF calculations, two different zero-order Hamiltonians for 3-methylindole in the gas-phase have been tested. Compared to the experimental absorptions of indole (4.37 and 4.77 eV), the latest IPEA Hamiltonian[S4] has a larger error of the computed absorptions than an imaginary level shift.

Table S4. The SAC-CI computed vertical absorption energies (ΔE) and oscillator strengths (f) of the isolated QM models.[a]

QM models	^1L_b state		^1L_a state	
	ΔE (eV)	f	ΔE (eV)	f
W285^{bare} in **A**	4.31	0.03	4.66	0.11
W233^{bare} in **B**	4.24	0.02	4.57	0.12
W285^{bare} in **F**	4.48	0.02	4.77	0.17
W233^{bare} in **F**	4.42	0.03	4.80	0.14
W285^{bare} in **G**	4.84	0.04	5.17	0.16
W233^{bare} in **G**	4.78	0.04	5.22	0.14
W337^{bare} in **G**	4.88	0.04	5.27	0.21

a. “bare” means that the isolated QM calculation was performed for the chromophore with its geometry optimized in the different ONIOM models of UVR8.

Table S5. The MS-CASPT2 computed vertical absorptions (*ΔE*) and oscillator strengths (*f*) of the isolated QM models.[a]

Methods	^1L_b state		^1L_a state	
	ΔE (eV)	*f*	*ΔE* (eV)	*f*
W285^{bare} in **A**				
MS-CASPT2/BS4	4.47	0.02	4.99	0.09
MS-CASPT2/BS5	4.36	0.02	4.77	0.09
MS-CASPT2/BS6	4.23	0.02	4.57	0.10
W233^{bare} in **B**				
MS-CASPT2/BS4	4.41	0.02	4.95	0.10
MS-CASPT2/BS5	4.30	0.02	4.72	0.10
MS-CASPT2/BS6	4.18	0.02	4.52	0.10

a. "bare" means that the isolated QM calculation was performed for the chromophore with its geometry optimized in the ONIOM model **A** or **B** of UVR8.

Table S6. The ONIOM(MS-CASPT2:AMBER)-EE and MS-CASPT2 computed absorptions (*ΔE*) and oscillator strengths (*f*) of tryptophans in UVR8 and the isolated QM models.[a]

Methods	^1L_b state		^1L_a state	
	ΔE (eV)	*f*	*ΔE* (eV)	*f*
W285/W285^{bare} in **F**				
MS-CASPT2/BS4	4.51/4.46	0.01/0.02	5.20/5.06	0.10/0.10
MS-CASPT2/BS5	4.40/4.35	0.01/0.02	4.92/4.82	0.11/0.10
MS-CASPT2/BS6	4.29/4.23	0.02/0.02	4.68/4.61	0.12/0.10
W233/W233^{bare} in **F**				
MS-CASPT2/BS4	4.34[b]/4.41	0.01/0.02	4.45[b]/4.94	0.07/0.10
MS-CASPT2/BS5	4.31[c]/4.30	0.02/0.02	4.18/4.71	0.07/0.10
MS-CASPT2/BS6	4.19[c]/4.18	0.02/0.02	4.01/4.51	0.07/0.10

a. Based on the optimized geometry in the model **F**: W285^{qm}+W233^{qm}, the absorption energies of W285 or W233 in UVR8 and the isolated QM model were calculated. "bare" means that the isolated QM calculation was performed for the chromophore with its geometry optimized in the ONIOM model **F** of UVR8. b. The ^1L_b and ^1L_a states are mixed. c. The ^1L_b state is mixed with ^1L_a state.

Table S7. The ONIOM(SAC-CI:AMBER)-EE and SAC-CI calculated dipole moments (μ) of tryptophans in UVR8 and the isolated QM models.[a]

Protein	μ (D)			QM model	μ (D)		
	S_0	^1L_b	^1L_a		S_0	^1L_b	^1L_a
W285 in **A**[b]	3.14	3.21	5.78	W285^{bare} in **A**[d]	2.03	2.48	5.87
W233 in **B**[c]	5.66	7.00	10.16	W233^{bare} in **B**[e]	2.25	2.63	5.95

a. "bare" means that the isolated QM calculation was performed for the chromophore with its geometry optimized in the ONIOM model **A** or **B** of UVR8.

b. CASPT2 (with BS4, BS5 and BS6) dipole moments of S_0, ^1L_b and ^1L_a states are 2.81-3.19, 2.51-2.92 and 6.12-6.34 D, respectively.

c. CASPT2 (with BS4, BS5 and BS6) dipole moments of S_0, ^1L_b and ^1L_a states are 5.30-5.81, 5.38-6.02 and 9.85-10.17 D, respectively.

d. CASPT2 (with BS4, BS5 and BS6) dipole moments of S_0, ^1L_b and ^1L_a states are 1.80-1.83, 1.67-1.72 and 6.00-6.17 D, respectively.

e. CASPT2 (with BS4, BS5 and BS6) dipole moments of S_0, ^1L_b and ^1L_a states are 2.01-2.06, 1.82-1.88 and 6.10-6.29 D, respectively.

Table S8. The SAC-CI/BS3 and ONIOM(SAC-CI/BS3:AMBER)-EE computed vertical absorption energies (ΔE) and oscillator strengths (f) of the isolated QM models and tryptophans in UVR8, based on the ONIOM(B3LYP/BS:AMBER)-EE optimized geometries.[a]

	^1L_b state		^1L_a state	
	ΔE (eV)	f	ΔE (eV)	f
W285^{bare}/W285 in **A**	4.28/4.30	0.03/0.03	4.63/4.62	0.11/0.12
W285^{bare}/W285 in **F**	4.43/4.48	0.02/0.02	4.78/4.89	0.15/0.18
W233^{bare}/W233 in **F**	4.37/4.41	0.03/0.02	4.75/4.08	0.11/0.09
W285^{bare}/W285 in **G**	4.79/4.66	0.03/0.04	5.08/5.10	0.18/0.18
W233^{bare}/W233 in **G**	4.63/4.64	0.03/0.02	5.01/4.21	0.18/0.09
W337^{bare}/W337 in **G**	4.86/4.60	0.04/0.05	5.15/4.77	0.18/0.13

a. "bare" means that the isolated QM calculation was performed for the chromophore with its geometry optimized in different ONIOM models of UVR8.

Table S9. The ONIOM(SAC-CI:AMBER)-EE calculated Mulliken charge of W285 in UVR8, based on the ONIOM(M06-2X:AMBER)-EE optimized ground-state structure.[a]

		Charge		
		S_0	^1L_b	^1L_a
W285 in **A**	N-ring	0.286	0.366	0.548
	C-ring	-0.286	-0.366	-0.548
W285 in **F**	N-ring	0.326	0.425	0.574
	C-ring	-0.312	-0.404	-0.549
W285 in **G**	N-ring	0.362	0.445	0.596
	C-ring	-0.361	-0.441	-0.592

a. N-ring and C-ring denote the pyrrole and benzene rings of W285, respectively.

Table S10. The computed $^1L_a \rightarrow S_0$ transition energies (ΔE) and oscillator strengths (f) of the isolated QM models.[a]

QM methods	$^1L_a \rightarrow S_0$	
	ΔE (eV)	f
W285^{bare} in **A**		
SAC-CI/BS3	3.84	0.14
MS-CASPT2/BS4	4.31	0.11
MS-CASPT2/BS5	4.09	0.11
MS-CASPT2/BS6	3.91	0.12
W233^{bare} in **B**		
SAC-CI/BS3	3.92	0.15
MS-CASPT2/BS4	4.35	0.12
MS-CASPT2/BS5	4.14	0.12
MS-CASPT2/BS6	3.96	0.12

a. “bare” means that the isolated QM calculation was carried out for the chromophore with its geometry optimized in the ONIOM model **A** or **B** of UVR8.

Table S11. The ONIOM(SAC-CI:AMBER)-EE calculated charge of W233 in UVR8, based on the ONIOM optimized equilibrium structures in S_0 and ^1L_a states, with respect to the isolated QM model.[a]

ONIOM model **B**		Charge of N-ring		
		S_0	^1L_b	^1L_a
S_0 State Geometry				
$W233^{bare}$ in **B**		0.316	0.380	0.581
W233 in **B**		0.379	0.498	0.718
W233 in **B** with removal of side-chain charges	R146(B)	0.355	0.462	0.682
	E182	0.398	0.521	0.744
	D129	0.291	0.378	0.599
	R234	0.333	0.411	0.606
^1L_a State Geometry				
$W233^{bare}$ in **B**		0.324	0.392	0.623
W233 in **B**		0.426	0.535	0.787
W233 in **B** with removal of side-chain charges	R146(B)	0.407	0.521	0.771
	E182	0.435	0.545	0.797
	D129	0.334	0.426	0.683
	R234	0.359	0.438	0.681

a. N-ring denotes the pyrrole ring of W233. “bare” means that the isolated QM calculation was performed for the chromophore with its geometry optimized in the ONIOM model **B** of UVR8.

Table S12. The HF/D95(d) calculated key orbitals for tryptophan in different environments.[a]

Main configuration (weight>0.2)		$\pi\rightarrow\pi^*$ ^1L_b state	$\pi\rightarrow\pi^*$ ^1L_a state
3-methylindole in gas-phase		MO34→MO36(0.78), MO35→MO37(-0.51)	MO35→MO36(0.88), MO34→MO37(0.29)
$W285^{bare}$ in **A**		MO34→MO36(0.78), MO35→MO37(-0.50)	MO35→MO36(0.88), MO34→MO37(0.27)
W285 in **A**		MO34→MO36(-0.78), MO35→MO37(0.51)	MO35→MO36(0.89), MO34→MO37(0.26)
$W233^{bare}$ in **B**		MO34→MO36(-0.78), MO35→MO37(0.51)	MO35→MO36(0.89), MO34→MO37(0.26)
W233 in **B**		MO34→MO36(-0.70), MO35→MO37(0.55), MO35→MO36(-0.28)	MO35→MO36(0.84), MO34→MO36(-0.35), MO34→MO37(0.21)
W233 in **B** with removal of side-chain charges	D129	MO34→MO36(0.77), MO35→MO37(-0.53)	MO35→MO36(-0.89), MO34→MO37(-0.24)
	R234	MO34→MO36(0.77), MO35→MO37(-0.53)	MO35→MO36(-0.89), MO34→MO37(-0.27)

a. “bare” means that the isolated QM calculation was performed for the chromophore with its geometry optimized in the ONIOM model **A** or **B** of UVR8.

Table S13. The ONIOM(SAC-CI:AMBER)-EE and SAC-CI computed average charge of the pyrrole moiety of tryptophans in UVR8 as well as the isolated QM models.[a]

Charge	S_0 State Geometry			^1L_a State Geometry	
	S_0	^1L_b	^1L_a	S_0	^1L_a
W285 in **A**	0.332	0.437	0.629	0.370	0.714
$W285^{bare}$ in **A**	0.310	0.387	0.588	0.322	0.623
W233 in **B**	0.337	0.460	0.620	0.377	0.720
$W233^{bare}$ in **B**	0.312	0.391	0.587	0.324	0.623

a. “bare” means that the isolated QM calculation was carried out for the chromophore with its geometry optimized in the ONIOM model **A** or **B** of UVR8.

References

(S1) (a) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092. (b) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
(S2) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
(S3) (a) Pastor, R. W.; Brooks, B. R.; Szabo, A. *Mol. Phys.* **1988**, *65*, 1409-1419. (b) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523-535. (c) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. *J. Chem. Phys.* **2001**, *114*, 2090-2098.
(S4) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **2004**, *396*, 142-149.