Electronic Supplementary Information: The structural changes in the signaling mechanism of bacteriophytochromes in solution revealed by a multiscale computational investigation †

Veronica Macaluso,* Giacomo Salvadori, Lorenzo Cupellini, and Benedetta Mennucci*

Department of Chemistry and Industrial Chemistry, University of Pisa, Pisa, Italy

E-mail: veronica.macaluso@dcci.unipi.it; benedetta.mennucci@unipi.it

S1 Clustering

Due to the importance of the D-ring in the photo-isomerization of phytochrome, we performed a clustering analysis on 90000 structures extracted from the three MD replicas using a principal component analysis (PCA) based on H-bond donor-acceptor distances involving the chromophore D-ring (see Figure S1 (a)). These distances were transformed using the following function:

$$\mathrm{Dn}' = \begin{cases} \exp\left[-\left(\frac{x-2.5}{4.5}\right)^2\right] & \text{if } x \geq 2.5 \\ 1.0 \qquad \text{if } x < 2.5 \end{cases}$$

where x is the H-bond donor-acceptor distance.

*To whom correspondence should be addressed

We performed a principal component analysis (PCA) on those newly transformed quantities and we applied hierarchical hdbscan algorithm [1] on the first five principal components.

In Figure S1(b) we report the cluster-distribution and in Figure S1(c) we show the distances (used for the PCA) along the MD replica-frames (of the 1-4μs time interval of their production run). Cluster 6 is the most populated one in our MD, while crystal *Chain A* and *Chain B* would be part of cluster 0. From Figure S1(c) we can see that Cluster 6 and cluster 0 differ substantially for D0 (D-ring nitrogen-water molecule H-bond pair) and D1 (D-ring nitrogen-His201 H-bond pair). In particular, D0 is small for both clusters while D1 is small only for cluster 0. As we can see from the crystal-structures, His201 is at H-bond distance in the case of *Chain A*, and a little further for *Chain B*. In our simulation His201 leaves the H-bond site and it is replaced by a water molecule.

In order to understand if the environment may play a role in the D-ring distortion, we randomly extracted frames (about 10 per cluster), which were previously optimized and performed calculations of the spectroscopic properties. Cluster 0 was excluded from this analysis because it was ambiguously defined, as it includes structures with large differences in distances D2, D3, D5 and D6 .

We note that clusters may be separated into two groups, showing Pfr-like or red-shifted (Pr-like) photo-response, respectively (see Figure S2 (a,b)). This behavior is retained when repeating the calculations by replacing the atomistic environment (protein matrix and solvent) with a continuum solvation model (PCM) in water (see Figure S2 (c,d)).

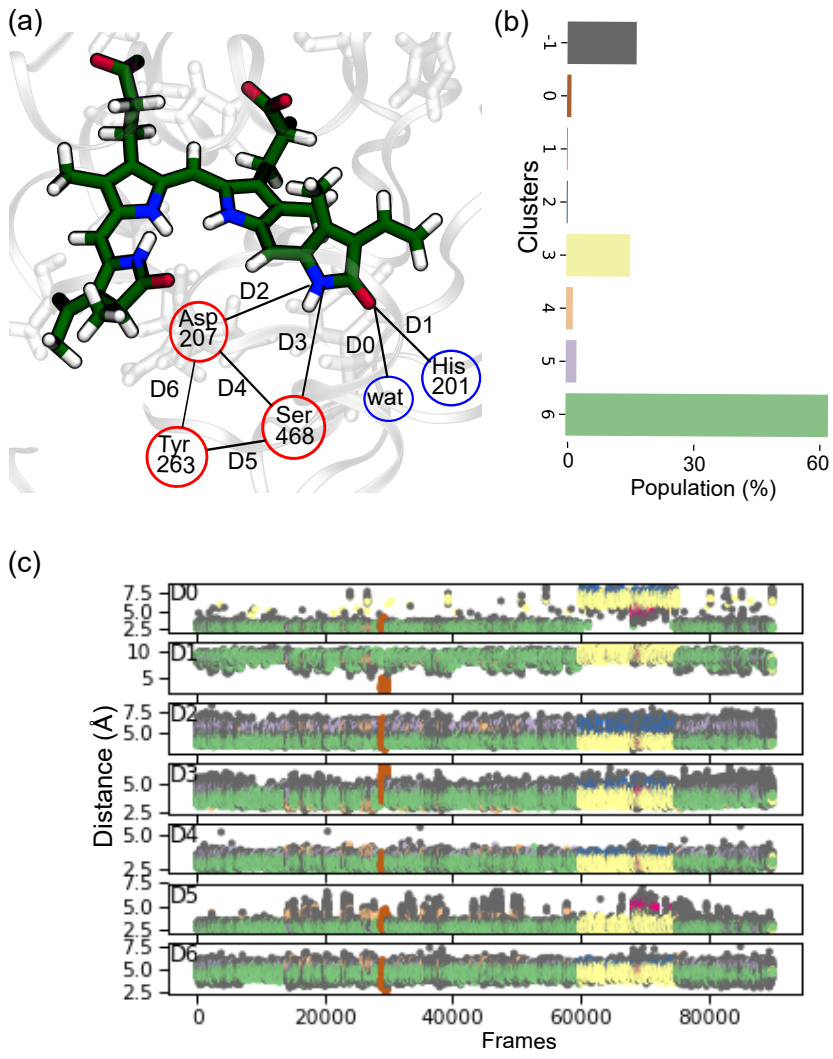


Figure S1: (a) Schematic representation of H-bond donor-acceptor distances involving the D-ring. (b) Clusters distribution over all MD replicas. (c) Clustering-based distances reported along the MD replica-frames. Frames 1 to 15000 and 15001 to 30000 correspond to Monomer A and Monomer B of Rep1, respectively; frames 30001 to 60000 to Monomer A and Monomer B of Rep2; frames from 60001 to 90000 to Monomer A and Monomer B of Rep3.

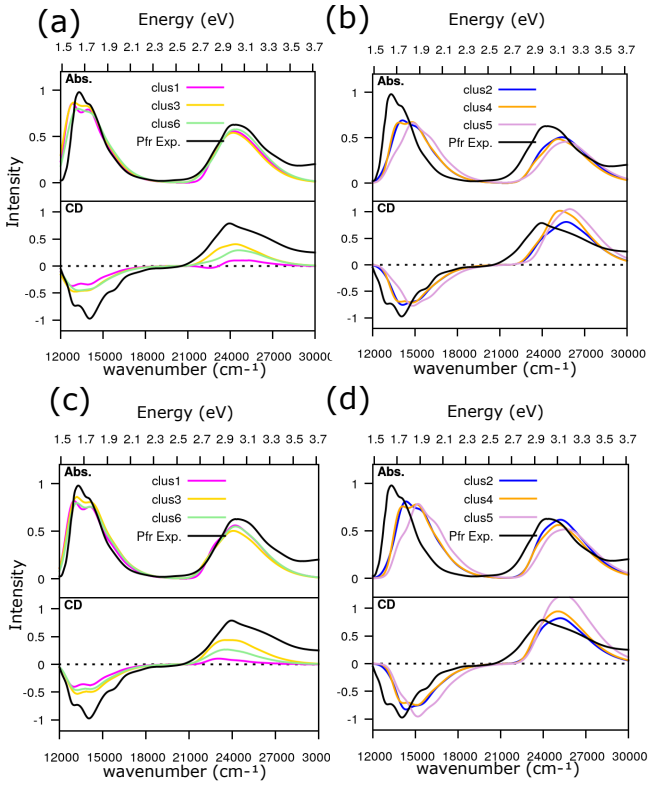


Figure S2: CAM-B3LYP/6-31+G(d) absorption and CD spectra computed for clusters 1 to 6 (see main text). (a,b) Spectra computed using a QM/MMPol scheme on QM/MM geometries and (c,d) spectra computed using PCM (water) on the same QM/MM geometries. All spectra have been shifted by -0.19 eV (as done for Pr calculated spectra[2] and for Pfr in the main text).

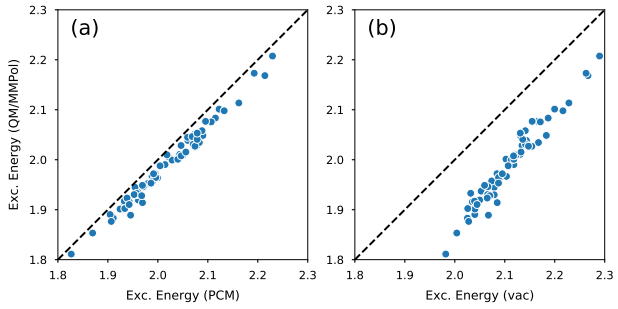


Figure S3: S_1 energies computed in (a) continuum PCM solvent (water) or (b) in vacuum, *versus* using a QM/MMPol description of the environment. All S_1 energies were calculated on the same QM/MM (ONIOM) optimized geometries, as described in the main text. For the calculations in vacuum the propionate-groups were excluded, by cutting the C 8^1 - C 8^2 and C 12^1 - C12^2 bonds.

S2 Additional figure

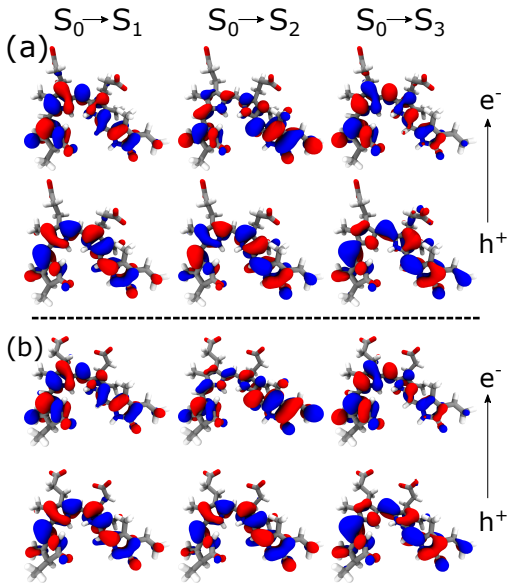
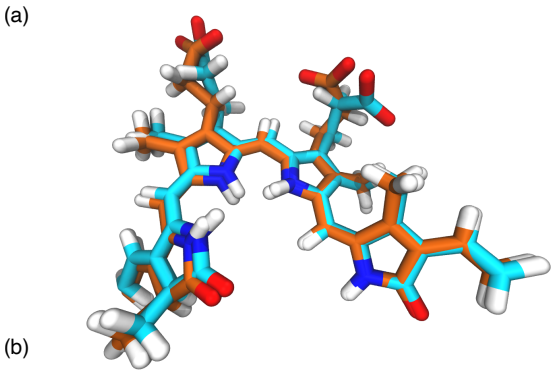


Figure S4: CAM-B3LYP NTO orbital pairs, "hole" (h) and "electron" (e^-, computed for (a) *Chain A* and (b) *Chain B* Pfr crystal structures for the $S_0 \rightarrow S_1$, $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$ transitions.



	φ_6	φ_5	φ_4	φ_3	φ_2	φ_1	C_{15}=C_{16}	C_{14}-C_{15}
Chain A	-154.53	-161.51	-7.08	-0.21	8.17	21.61	1.361	1.443
Chain B	-154.78	-162.79	-12.34	-4.61	13.99	25.70	1.359	1.445

Figure S5: (a) *Chain A* (blue) and *Chain B* (orange) Pfr crystal geometries optimized with B3LYP-D3/6-31G(d) using the ONIOM scheme. The dihedral angles (φ_x) are in degrees, while the bond lengths in angstrom.

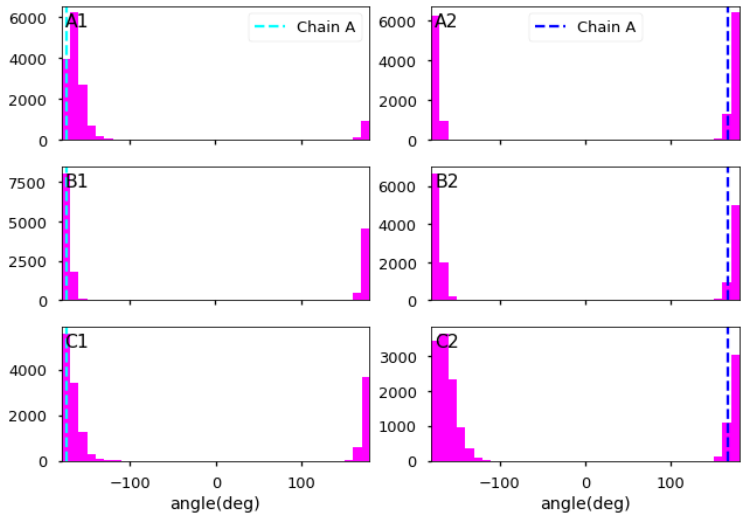


Figure S6: B-ring propionate dihedral angle C8-C8^1-C8^2-C8^3 (see Figure 1) distribution over all MD-repicas in the simulation range 1-4 μs. Those analysis are reported separately for Monomer A (left) and Monomer B (right). The force-field minimized initial structures (respectively Pfr crystal *Chain A* and *Chain B* for Monomer A and Monomer B) are indicated as dashed lines.

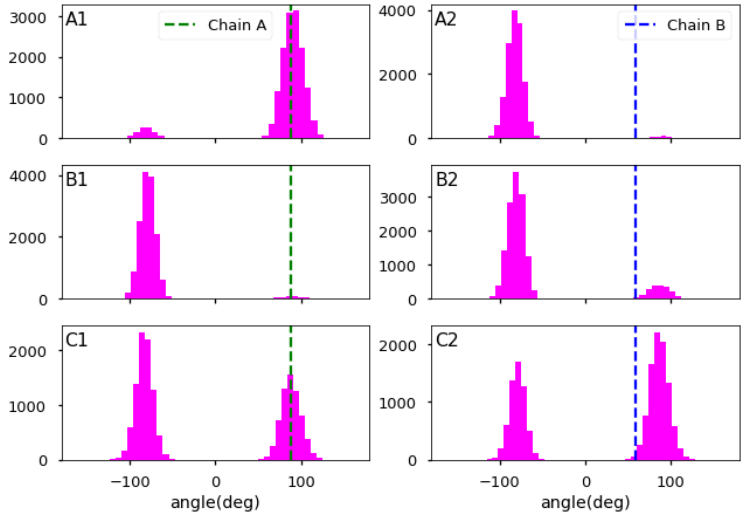


Figure S7: B-ring propionate dihedral angle C7-C8-C8^1-C8^2 (see Figure 1) distribution over all MD-repicas in the simulation range 1-4 μs. Those analysis are reported separately for Monomer A (left) and Monomer B (right). The force-field minimized initial structures (respectively Pfr crystal *Chain A* and *Chain B* for Monomer A and Monomer B) are indicated as dashed lines.

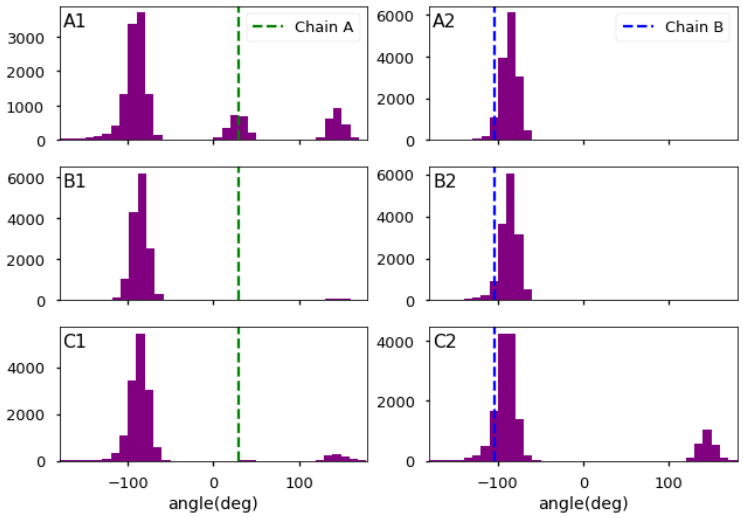


Figure S8: C-ring propionate dihedral angle C12-C12^1-C12^2-C12^3 (see Figure 1) distribution over the simulation range 1-4 μs of Rep1, Rep2, Rep3 (from top to bottom). Those analysis are reported separately for Monomer A (left) and Monomer B (right). The force-field minimized initial structures (respectively Pfr crystal *Chain A* and *Chain B* for Monomer A and Monomer B) are indicated as dashed lines.

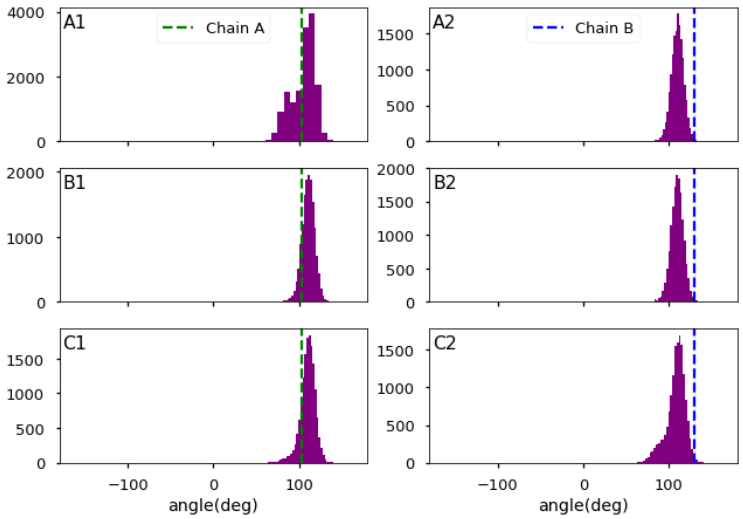


Figure S9: C-ring propionate dihedral angle C11-C12-C12^1-C12^2 (see Figure 1) distribution over the simulation range 1-4 μs of Rep1, Rep2, Rep3 (from top to bottom). Those analysis are reported separately for Monomer A (left) and Monomer B (right). The force-field minimized initial structures (respectively Pfr crystal *Chain A* and *Chain B* for Monomer A and Monomer B) are indicated as dashed lines.

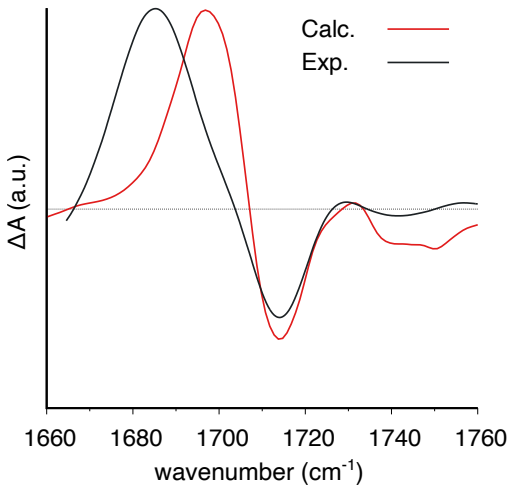


Figure S10: (a) Measured (black) and calculated (blue) difference IR spectra (Pfr-minus-Pr). The calculations were performed on previously QM/MM optimized MD snaphots and the resulting IR spectra were averaged. The calculated frequencies were scaled by 0.946 to match the negative Pr peak.

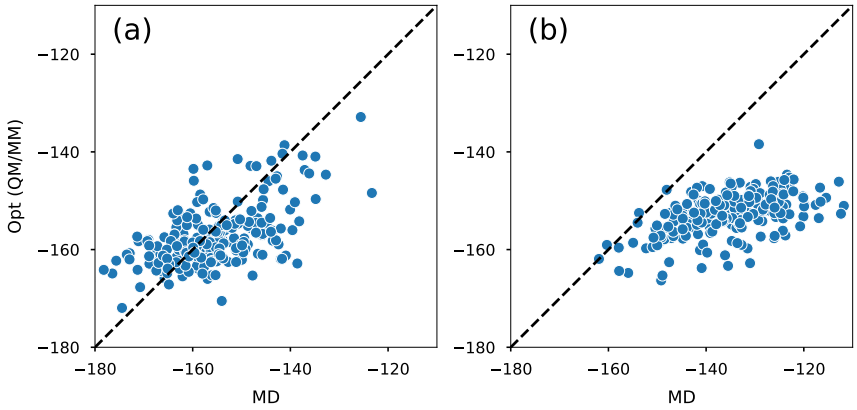


Figure S11: Comparison of φ_5 (panel a) and φ_6 (panel b) values in MD (non-optimized) and QM/MM optimized structures.

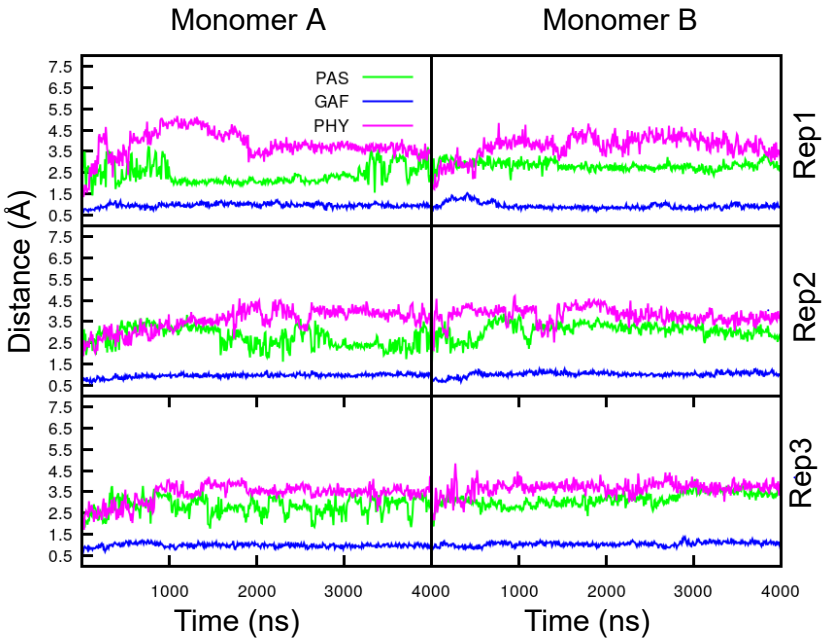


Figure S12: Root mean square displacement (RMSD) on the protein backbone (C,O,N atoms) for PAS, GAF and PHY domains calculated on Monomer A and Monomer B of replicas Rep1, Rep2, Rep3.

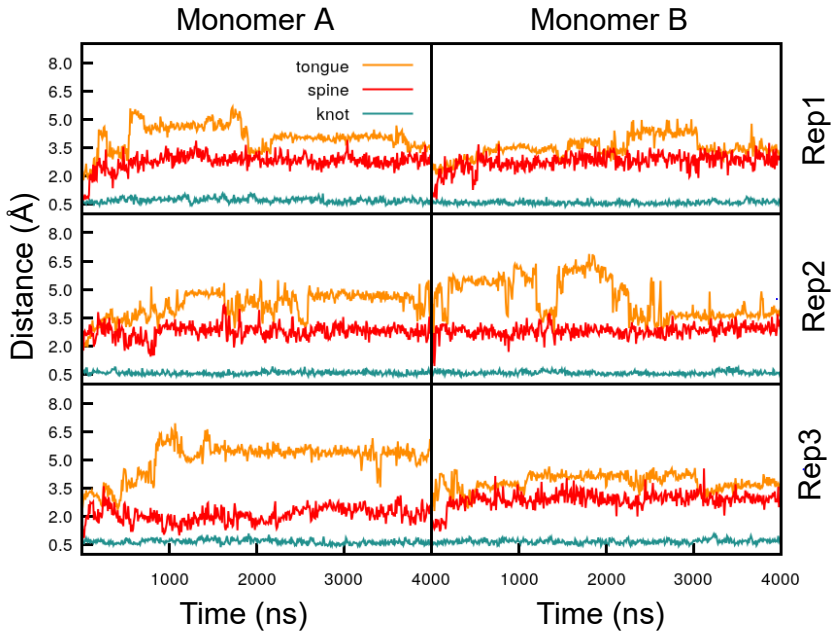


Figure S13: Root mean square displacement (RMSD) on the protein backbone (C,O,N atoms) for the helical-spine, tongue and knot calculated on Monomer A and Monomer B for replicas Rep1, Rep2, Rep3.

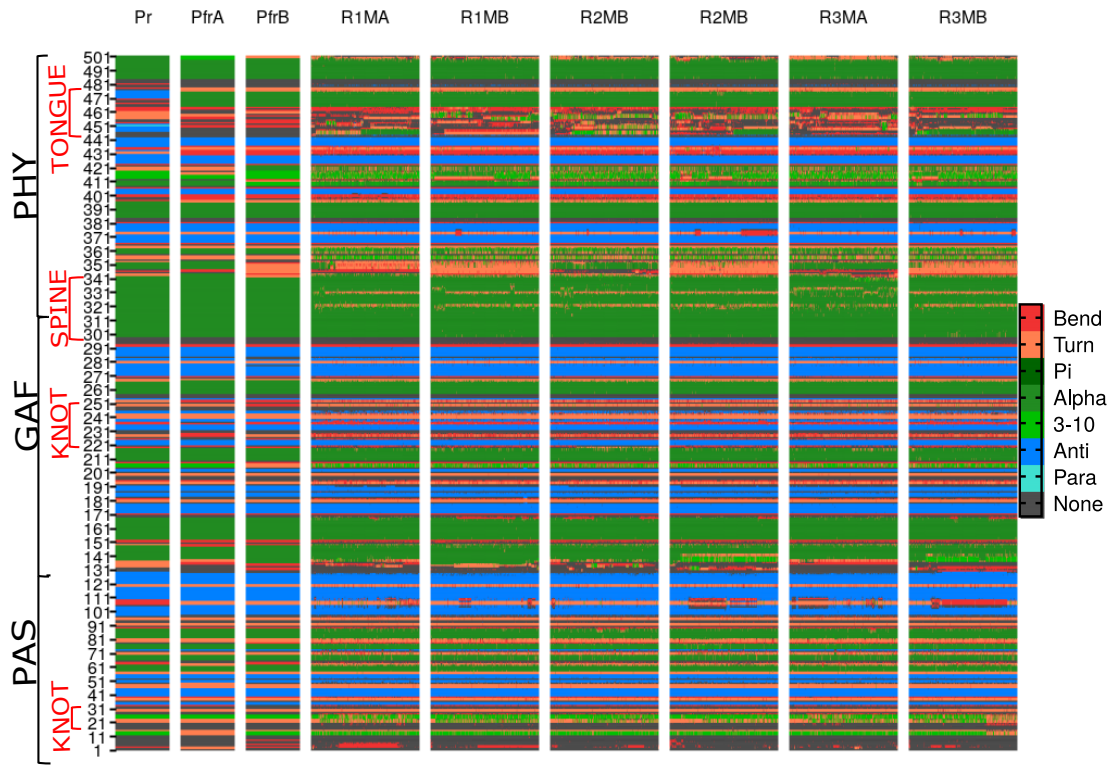


Figure S14: Secondary structure analysis on Pr, Pfr *Chain A* (PfrA), Pfr *Chain B* (PfrB) crystal structures (left part of the panel) and on production run for Rep1 (R1), Rep2 (R2) and Rep3 (R3) for Monomer A (MA) and Monomer B (MB). Initial structures for Monomer A and Monomer B are crystal *Chain A* and *Chain B*, respectively.

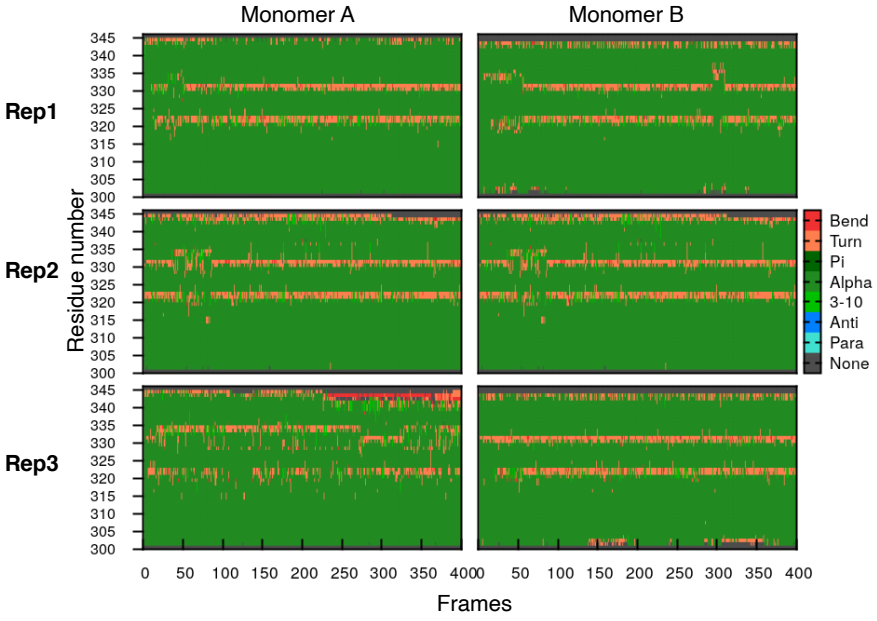


Figure S15: Zoom-in of the secondary structure analysis performed on the protein backbone (C,O,N atoms) for the helica-spine on Monomer A and Monomer B for replicas Rep1, Rep2, Rep3. Initial structures for Monomer A and Monomer B are crystal *Chain A* and *Chain B*, respectively.

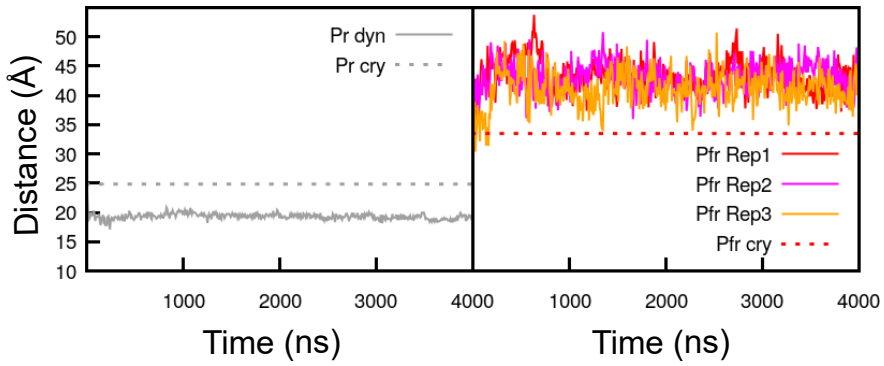


Figure S16: Distance between the PHY domains center-of-mass calculated on the production run of Pr (left panel) and Pfr (right panel) simulation runs. The dashed lines represent the crystal Pr and Pfr crystal structure values for the PHY domains distance.

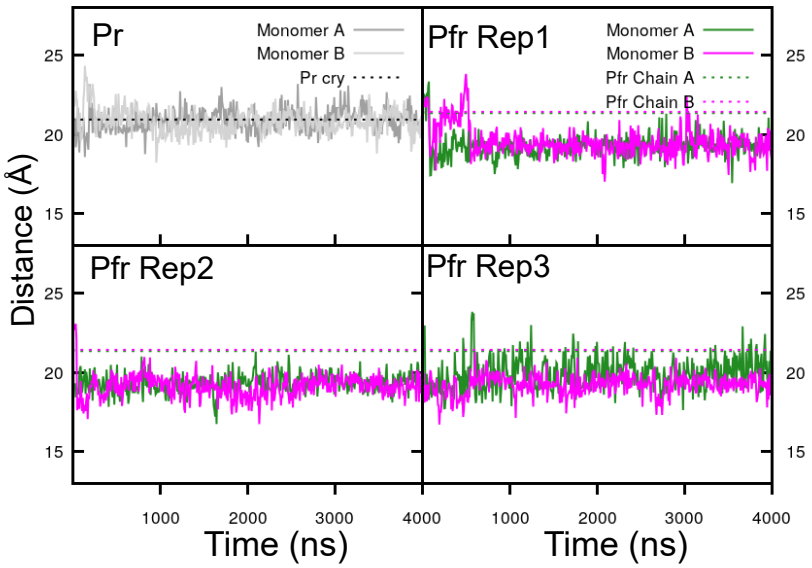


Figure S17: Distance between the GAF and PHY domain calculated on the production run of Pr and Pfr simulations. The same distance (dashed line) is also reported for Pr and Pfr crystal structures. The distances were calculated between the center-of-mass of residues 153 to 157 (for GAF) and 485 to 488 (for PHY). In the case of Pr MD simulations the initial structure (Pr cry) is the same for the two monomer- (Monomer A and Monomer B). The initial structures for Monomer A and Monomer B of Pfr MD simulations are respectively *Chain A* and *Chain B*.

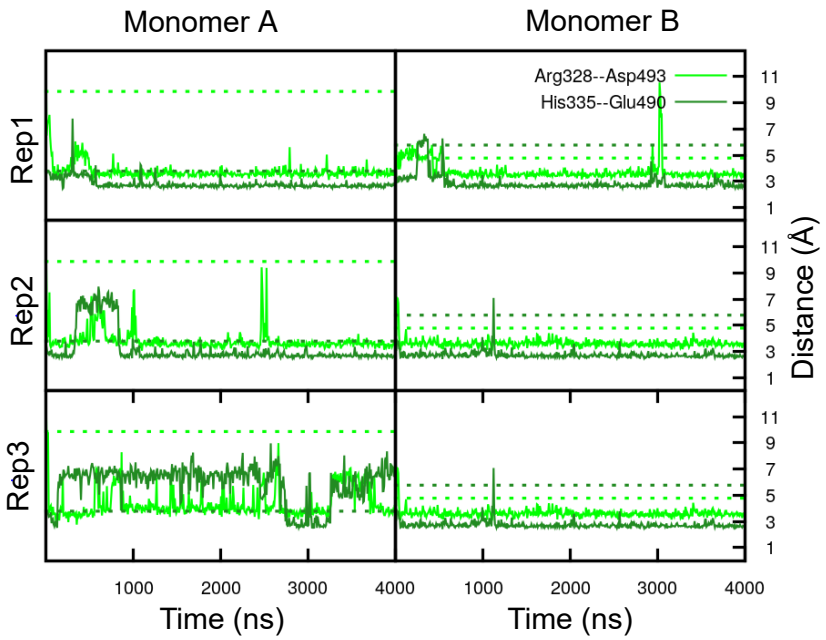


Figure S18: Distances between H-bond donor-acceptor atoms in proximity of the spine angle θ_1 (see Figure 5(b)), i.e. distance between Arg328 (amino-Nytrogens center-of-mass) and Asp493 (carbossylic Oxygens center-of-mass) and distance between His335 (δ-N) and Glu490 (carbossylic Oxygens center-of-mass). These analysis were performed on Pfr MD simulations and separately for Monomer A and Monomer B (left and right). The same distances are also reported for the Pfr crystal structures *Chain A* and *Chain B* (dashed lines, respectively left and right panel), which are respectively the initial structure for Monomer A and Monomer B).

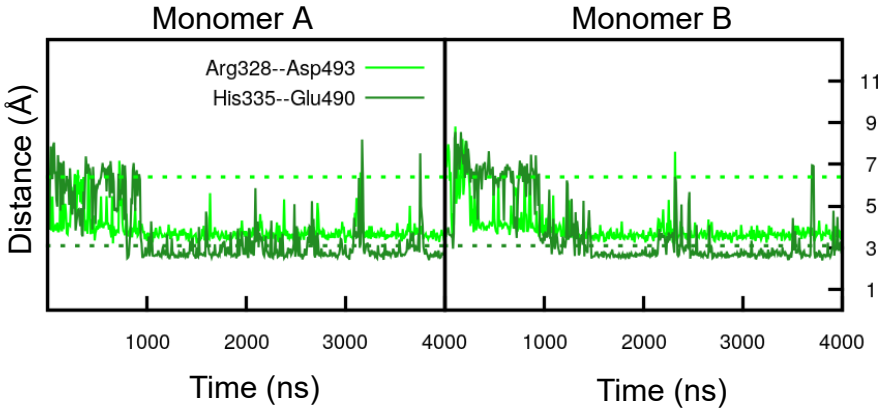


Figure S19: Distances between H-bond donor-acceptor atoms in proximity of the spine angle θ_1 (see Figure 5(b)), i.e. distance between Arg328 (amino-Nytrogens center-of-mass) and Asp493 (carbossylic Oxygens center-of-mass) and distance between His335 (δ-Nytrogen) and Glu490 (carbossylic Oxygens center-of-mass). These analysis were performed on Pr MD simulations and separately for Monomer A and Monomer B (left and right). The same distances are also reported for the Pr crystal structure (dashed line).

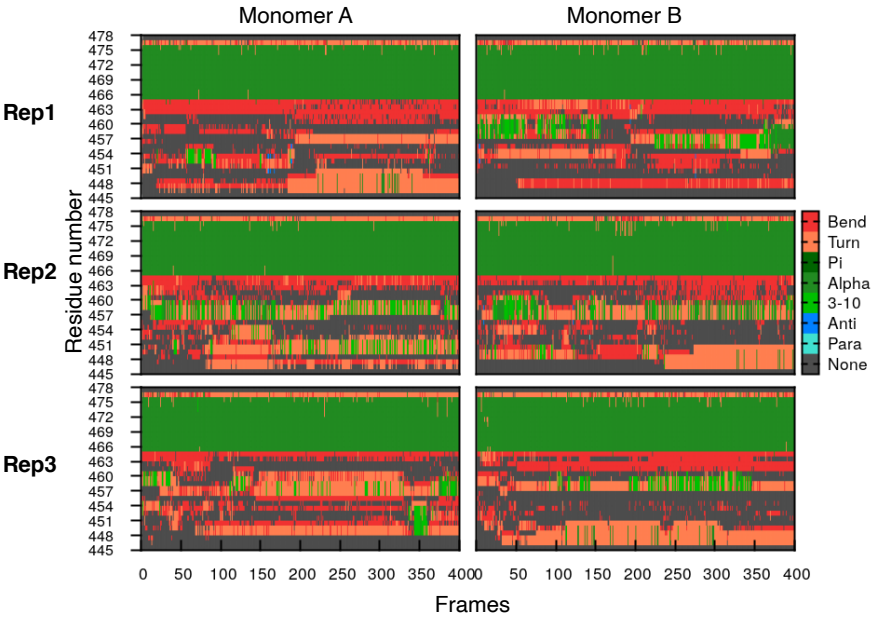


Figure S20: Zoom-in of the secondary structure analysis performed on the protein backbone (C,O,N atoms) for the tongue on Monomer A and Monomer B for replicas Rep1, Rep2, Rep3. Initial structures for Monomer A and Monomer B are crystal *Chain A* and *Chain B*, respectively.

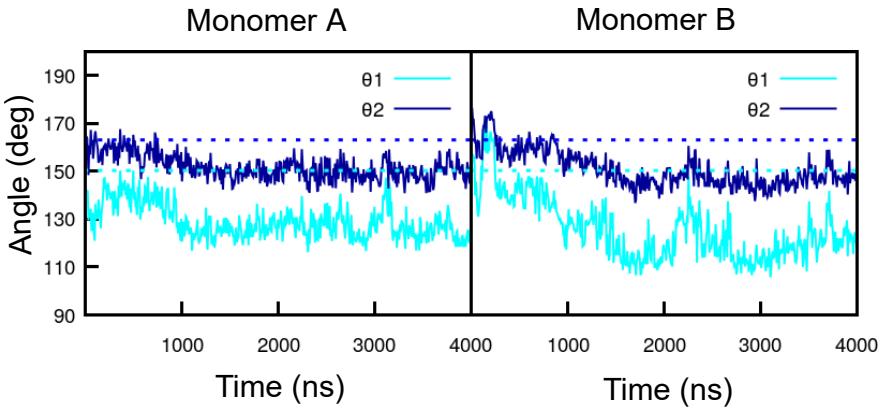


Figure S21: Spine-angles θ_1 and θ_2 are analyzed along the simulation time for Pr MD simulations and separately for Monomer A and Monomer B (left and right). The same angles are also calculated on the Pfr crystal structures *Chain A* and *Chain B* (dashed lines, respectively left and right panel), which are respectively the initial structure for Monomer A and Monomer B).

S3 Additional Tables

Table S1: The table below shows the structural elements of the Deinococcus Radiodurans bacteria-phytochrome.

	PAS	GAF	PHY	TONGUE	SPINE	KNOT
Residue number	$1-136$	$137-317$	$318-502$	$445-478$	$300-345$	$27-38, 228-256$

Table S2: Probabilities for H-bonds involving the spine and the PAS, GAF and PHY domains in Pr[2] and Pfr simulations. The Pfr analysis was performed on 150000 frames, while the Pr one on 60000 frames. Probabilities were calculated averaging over Monomer A and Monomer B of Rep1 and Rep2 and over Monomer B of Rep3 in the case of Pfr. Only the hydrogen bonds with a probability greater than 30% are represented. GAF′ and PHY′ are the GAF and PHY domains, respectively, reduced so as not to include spine residues. The AmberTools CPPTRAJ package[3] was used to perform the hydrogen bond analysis (with a distance cutoff of 3.2 Å and an angle one of 125Âř.

Spine-PAS			**Spine-GAF′**			**Spine-PHY′**		
Acceptor-Donor	**Pfr (%)**	**Pr (%)**	**Acceptor-Donor**	**Pfr (%)**	**Pr (%)**	**Acceptor-Donor**	**Pfr (%)**	**Pr (%)**
Asp300-Arg100	13	71	Ala150-Lys319	32	11	Ala341-Thr351	45	1
Thr135-Thr304	1	41	Glu148-Gln315	67	32	Asp323-His484	44	19
Asp300-Thr135	36	0	Glu148-Lys319	49	9	His344-Ser347	44	5
Glu306-Arg100	39	2	Leu297-Arg302	71	53	Ala356-Arg337	26	42
Glu127-Arg302	92	100	Gln317-Gly280	11	42	Asp493-Arg328	99	98
			Glu320-Lys177	16	41	Glu489-Arg328	36	31
			Glu320-Trp284	64	84	Glu490-Arg328	98	99
			Pro151-Lys319	16	36	Glu490-His335	99	93
			Pro298-Leu301	25	41	Leu357-His334	76	77
			Ser313-Val278	29	54			
			Pro298-Arg302	96	91			
			Pro299-Thr303	87	96			

Table S3: Probabilities for H-bonds involving the tongue and the GAF and PHY domains in Pr[2] and Pfr simulations. The Pfr analysis was performed on 150000 frames, while the Pr one on 60000 frames. Probabilities were calculated averaging over Monomer A and Monomer B of Rep1 and Rep2 and over Monomer B of Rep3 in the case of Pfr. Only the hydrogen bonds with a probability greater than 30% are represented. PHY″ is the PHY domain, which was reduced so as not to include tongue residues. The AmberTools CPPTRAJ package[3] was used to perform the hydrogen bond analysis (with a distance cutoff of 3.2 Å and an angle one of 125Âř.

Tongue-GAF			**Tongue-PHY″**		
Acceptor-Donor	**Pfr (%)**	**Pr (%)**	**Acceptor-Donor**	**Pfr (%)**	**Pr (%)**
Ala450-Arg202	0	97	Glu444-Arg446	84	57
Arg202-Gly452	0	92	Glu444-Glu448	35	0
Asp207-Arg466	0	100	Glu444-Leu447	69	0
Arg466-Asn266	0	77	Lys476-Tyr479	73	12
Thr471-His201	0	36	Glu444-Gly478	96	83
Asp207-Ser468	100	0	Tyr479-Leu445	95	90
Ser468-Tyr263	88	0			
Asp181-Lys476	64	19			
Leu463-Asn266	73	0			
Tyr472-Arg202	51	0			

References

(1) Schubert, E.; Sander, J.; Ester, M.; Kriegel, H. P.; Xu, X. *ACM Trans. Database Syst.* **2017**, *42*.

(2) Macaluso, V.; Cupellini, L.; Salvadori, G.; Lipparini, F.; Mennucci, B. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8585–8594.

(3) Roe, D. R.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.