Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information

Exploring Structural Dynamics and Optical Properties of UnaG Fluorescent Protein upon N57 Mutations

Mohammad Asad, and Adèle D. Laurent*

Nantes Université, CNRS, CEISAM, UMR 6230, F-44000 Nantes, France.

*Corresponding Author

Adèle D. Laurent: Email: Adele.Laurent@univ-nantes.fr; Phone Number: +33(0)251 125743

1. Structural details

Table ESI-1: Detailed set-up of simulations for UnaG-WT and two mutants.

System	Protein Residues	Water Molecules	Na^+	Total Atoms	Simulation Time [ns]
WT	139	10557	2	33957	500
N57A	139	10205	2	32894	500
N57Q	139	9762	2	31575	500

Table ESI-2**:** Detailed information of each protein domain and sequence length.

Domain	Number of residues	Sequence Length	Sequence	Domain	Number of residues	Sequence Length	Sequence
H2	9	16-24	NFGEYLKAI	B7	8	96-103	KLVYVQKW
H3	11	27-37	PKELSDGGDAT	B8	10	106-115	KETTYVREIK
B1	9	7-15	GTWKIADSH	B9	8	118-125	KLVVTLTM
B2	7	40-46	TLYISQK	B10	8	128-135	VVAVRSYR
B3	7	50-56	KMTVKIE	L1	7	57-63	NGPPTFL
B4	7	64-70	DTQVKFK	L2	4	71-74	LGEE
B5	3	75-77	FDE	L3	8	78-85	FPSDRRKG
B6	8	86-93	VKSVVNLV				

2. Structural analysis for the two replicates

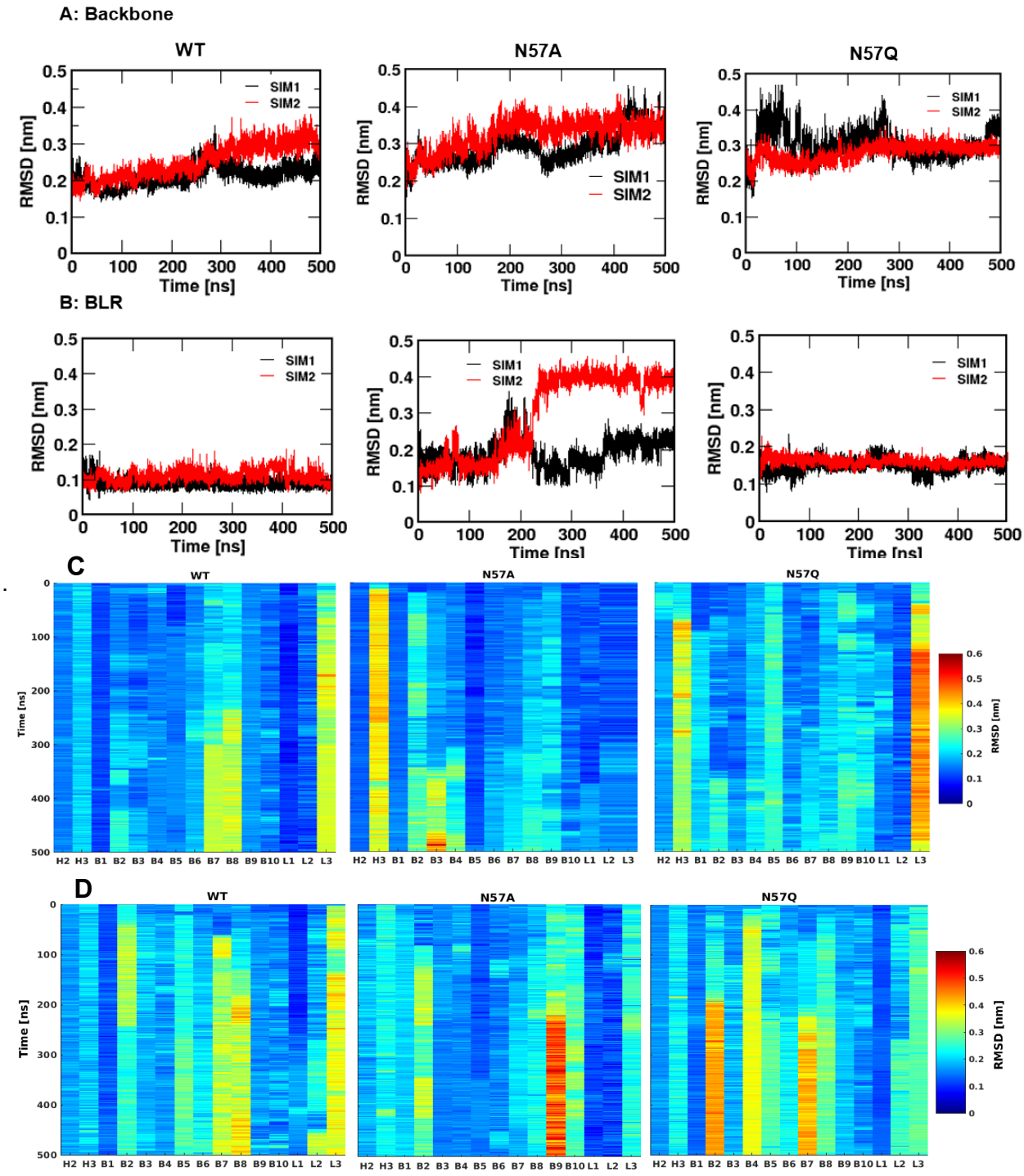

Figure ESI-1: Evolution of the protein backbone RMSD values along the simulations (SIM1 and SIM2) are presented on Figure ESI-1A and ESI-1B as well as RMSD values for the all-atom BLR ligand. For a more detailed inspection of the RMSD values for each secondary structure element, we have split RMSD values (for SIM1 ans SIM2) per helices, β sheets (B), and loops (L) (Figure ESI-1C and ESI-1D) following the above mentioned nomenclature (Table ESI-2).

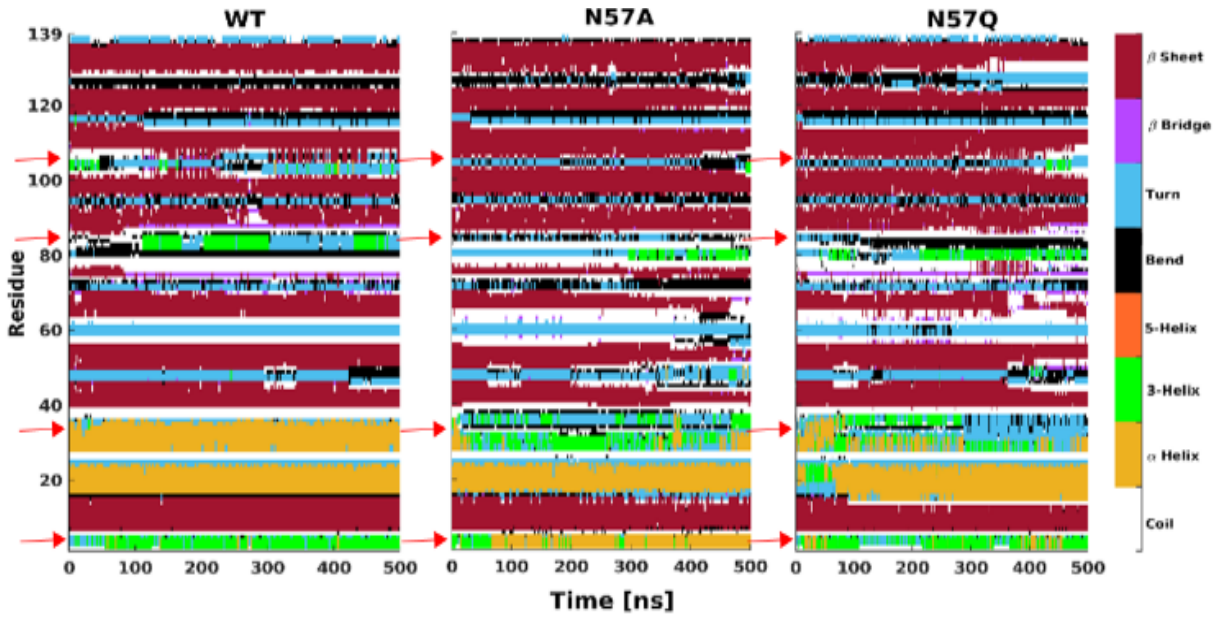

Figure ESI-2: Evaluation of the secondary structure along the simulation time for each systems

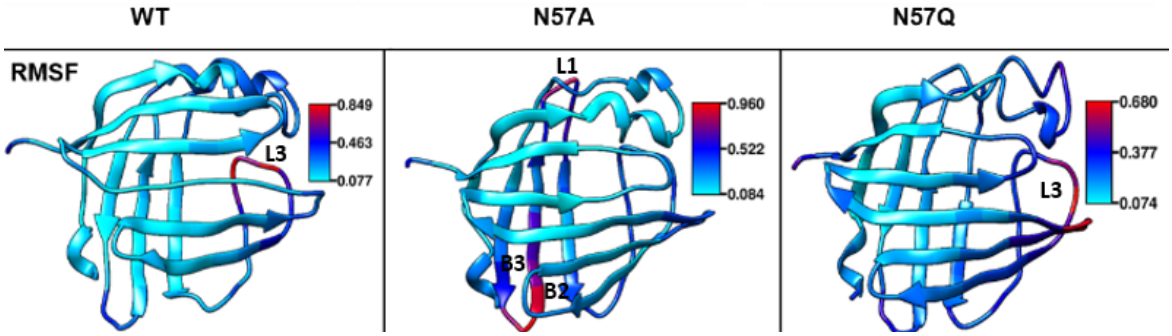

Figure ESI-3: RMSF mapped on to the 3D structure of each system. Highly fluctuated regions shown in red color.

Principal Component Analysis (PCA) was performed using Bio3D[1]. Firstly, the gromacs trajectories were converted into the Bio3D compatible format (dcd) as input for the PCA calculations. A PDB format trajectory was then produced to interpolate between the most dissimilar structures in

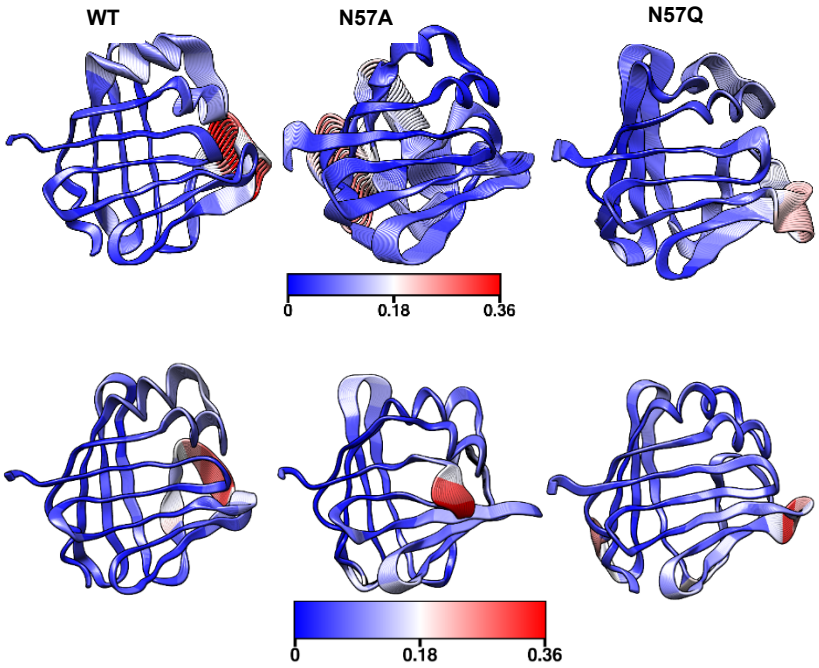

Figure ESI-4: Top, PCA for three systems WT, N57A and N57Q from simulation 1, bottom from simulation 2. Color scale from blue to red depicts low to high atomic displacements.

the distribution along a given principal component.

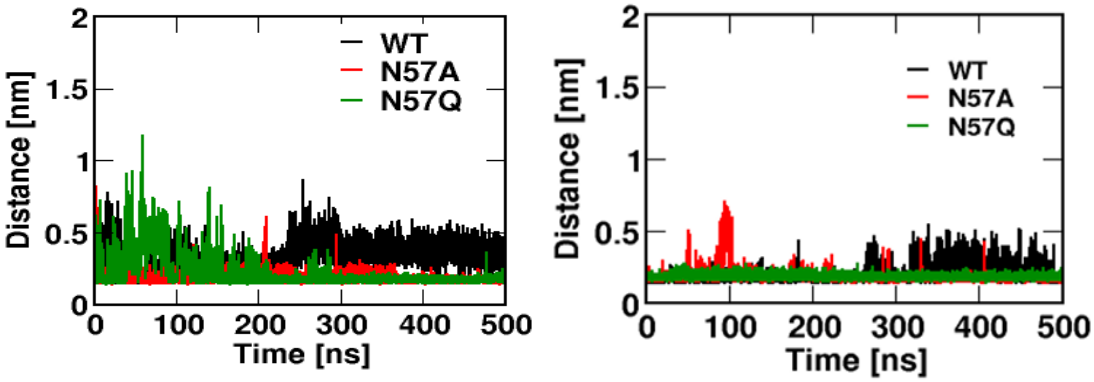

Figure ESI-5: Distance between the BLR center of mass and L3 center of mass. Left is for simulation 1 and right for simulation 2.

3. Water analysis

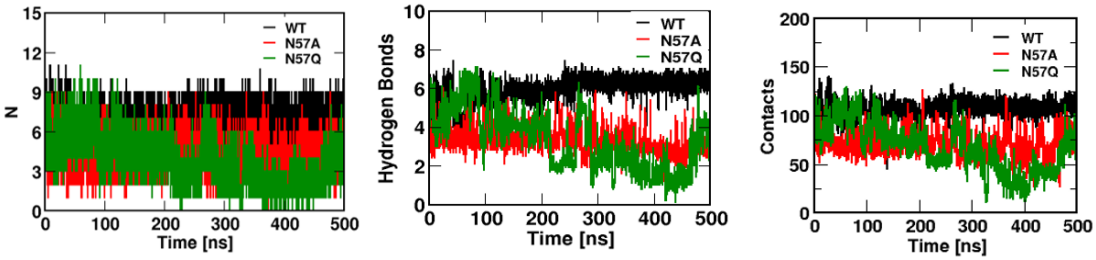

Figure ESI-6: Left, number of water molecule around BLR within 4 Å radius. Center: Number of H-bond between water molecules and BLR. Right, number of contacts between water and BLR. H-bond and contacts were defined If the distance between two atoms ≤ 3.5 Å in each system.

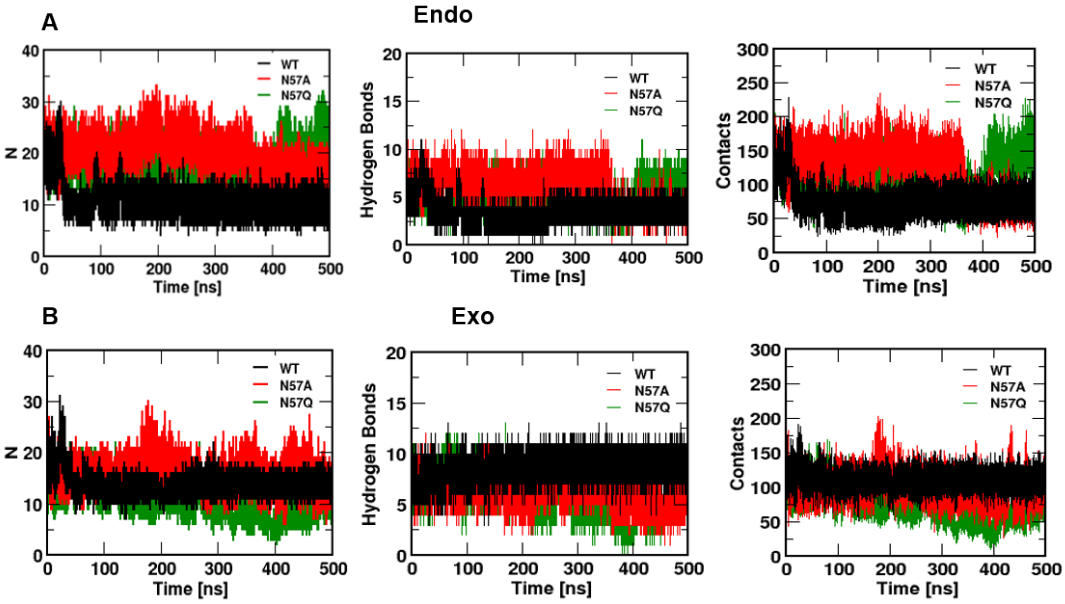

Figure ESI-7: Water dynamics around 4 Å of BLR endo and exo moeities. A) Time evolution number of water molecules around 4 Å of BLR *endo* moiety (see Figure 1 of manuscript for nomenclature), H-bonds, and contacts to the water molecules in each system. B) Time evolution number of water molecules in 4 Å.

4. Bilirubine geometry and surrounding analysis

The distance between H-bond donor (HBD) and H-bond acceptor (HBA) residues and BLR atoms, within a distance ≤ 3.5 Å has been considered for the further hydrogen bond analysis. To do so, we took 6 residues which are having distance ≤ 3.5 Å and that are also reported in UNIPROT[2] as binding residues; we then calculated the time evolution of the H-bonds.

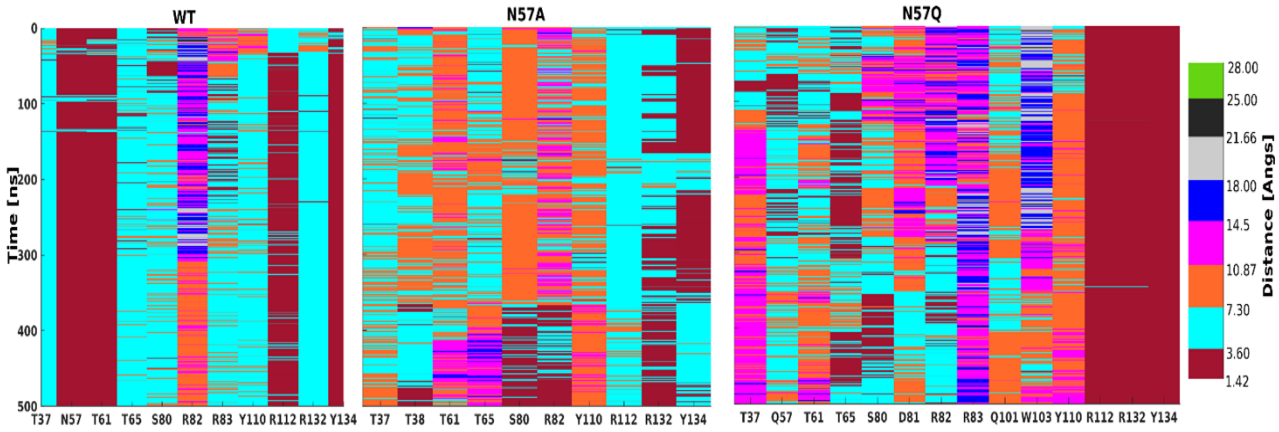

Figure ESI-8: Time evolution distance between the HBD and HBA atoms of the protein residues with BLR in each system.

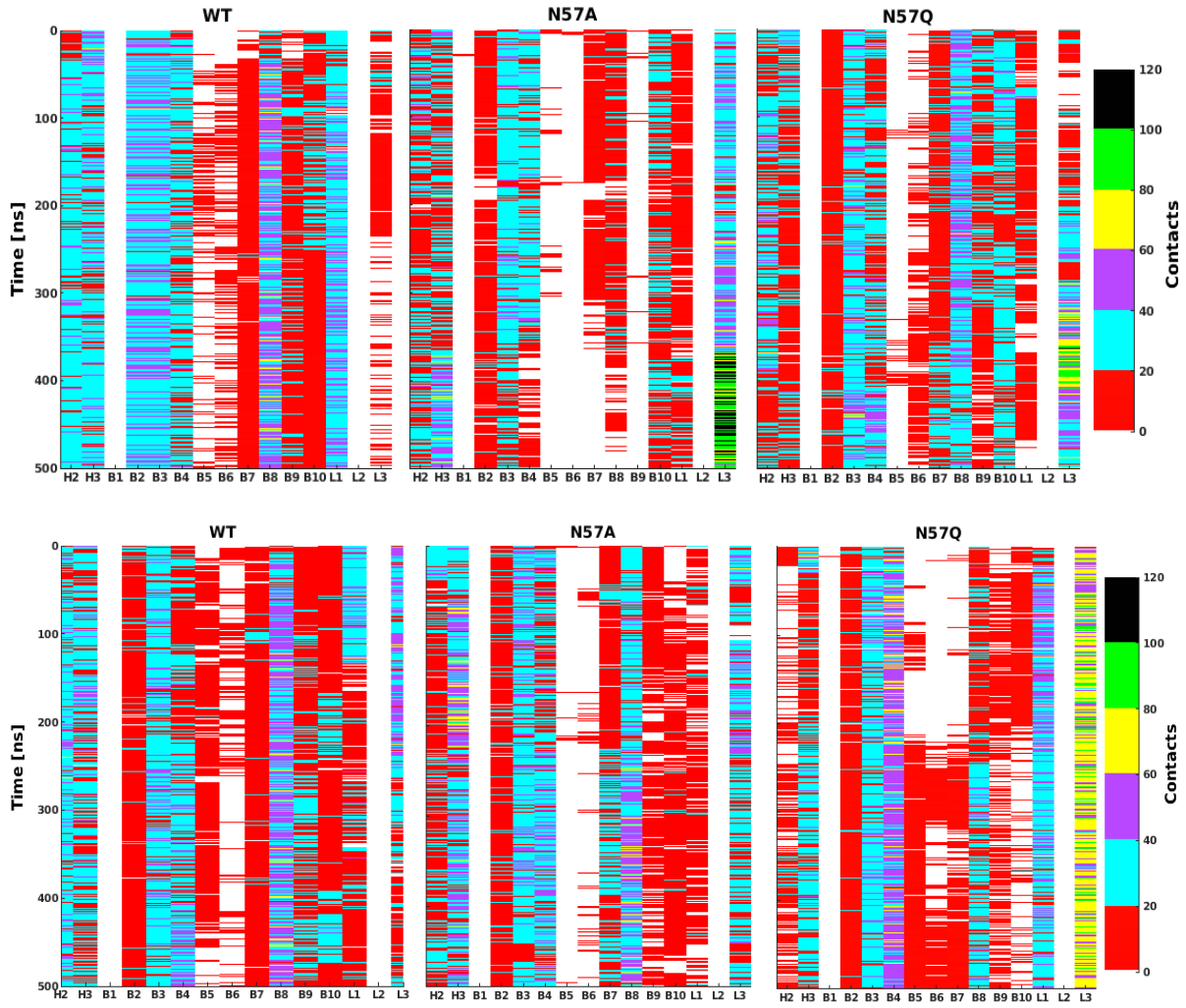

Figure ESI-9: Time evolution of the number of protein domain contacts with BLR in each system in the first (Top) and second simulation (Bottom).

MM/PBSA Binding energies:

Binding free energy calculations between UnaG and BLR binding were performed using g_mmpbsa program[3]. Molecular mechanics Poisson-Boltzman surface area (MM-PBSA), a method to estimate the interaction free energies of biomolecular interactions. g_mmpbsa calculates the three energy components, a) Calculations of potential energy b) calculations of polar solvation energy, c) calculation of non-polar solvation energy, and at the end calculates the average binding energy from those three calculated energetic terms using a python scrip provided in the g_mmpbsa tutorial. The binding energy between BLR and protein was computed every 1ns structure.

To probe the binding energetics of UnaG and BLR, the MM/PBSA energy calculations were estimated between protein and BLR. WT shows stable binding energy (average value: -32.50 ± 6.23 kcal/mol) while mutants N57A, in average -36.11 ±10.7 kcal/mol and N57Q in average -35.70 ±9.0 kcal/mol (see Figure ESI-10). Next, we decomposed the energetics by into a residue analysis focusing on the ones within 4 Å of BLR (23 residues). Interestingly we observed that N57 has contributed to – 0.5 kcal/mol in WT, while for A57 to –0.1 kcal/mol and for Q57 to 0.6 kcal/mol. Considering T61 and T65 which are also contributing the contribution is -1.2 kcal/mol in WT while less in mutants. Highlighting some hydrophobic residues as L21, L41, I55, L63, V67, and V121, we notice that they also contributes in the binding energy in WT while in a lesser extend in mutants. The RMSD values of 23 binding residues clearly shown that WT binding residues were more stable compare to mutants (Figure ESI).

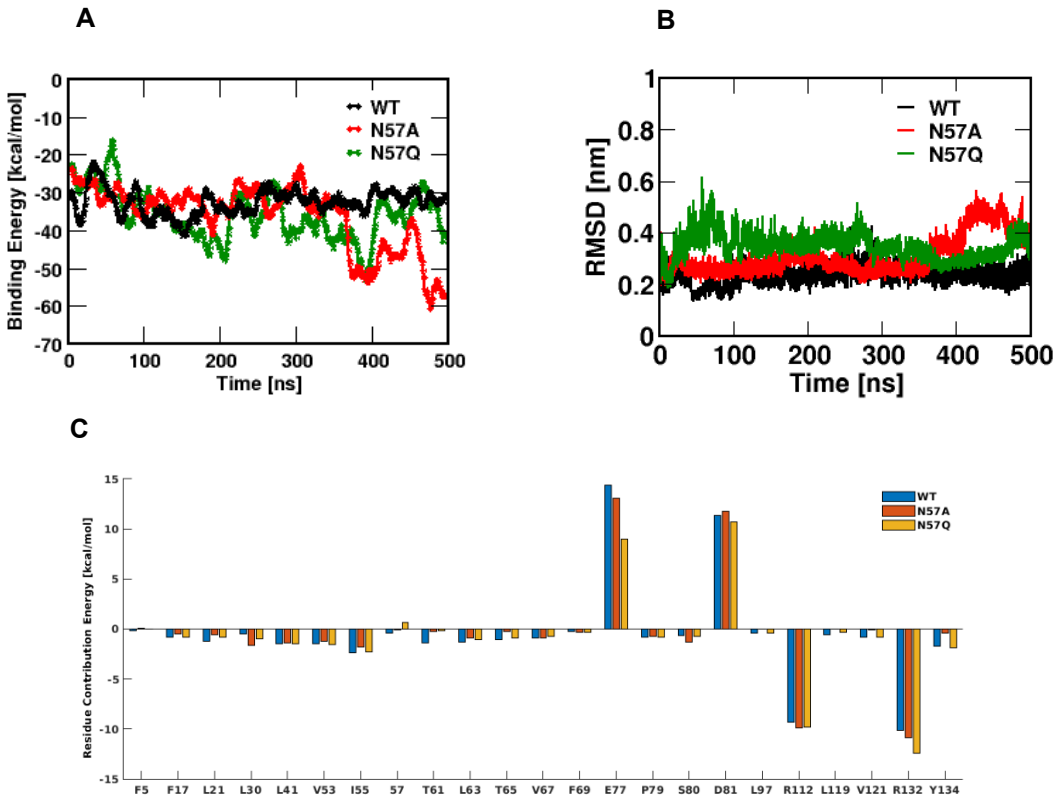

Figure ESI-10 A) Binding energy calculation of each system with respect to time. B) Evolution of RMSD values of residues being around 4 Å of BLR. C) Energy contribution of each protein residues to the binding energy for residues around 4 Å of BLR.

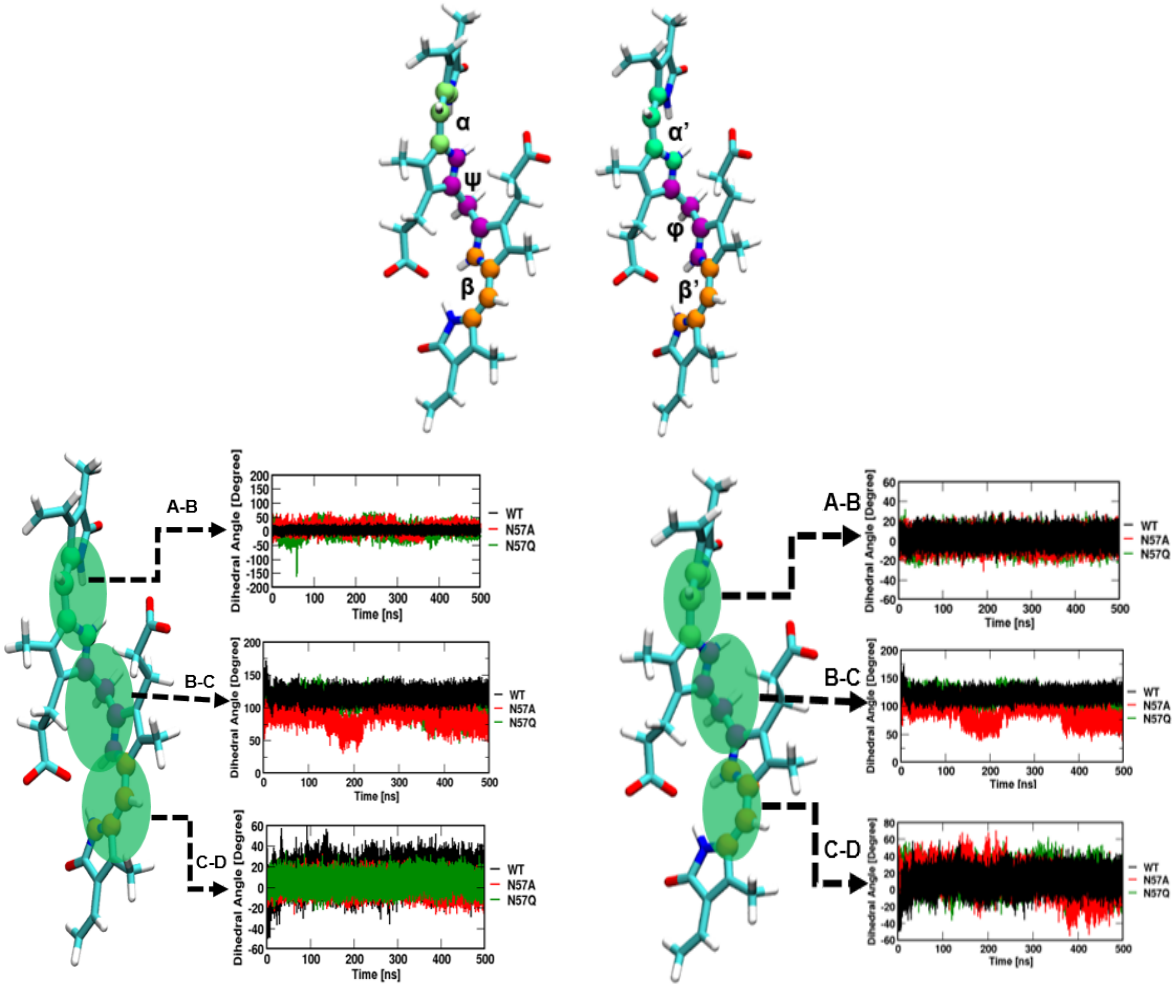

Figure ESI-11: Time evolution of the six dihedral angles for different BLR ring of each system. Similar results are obtained for the duplicate.

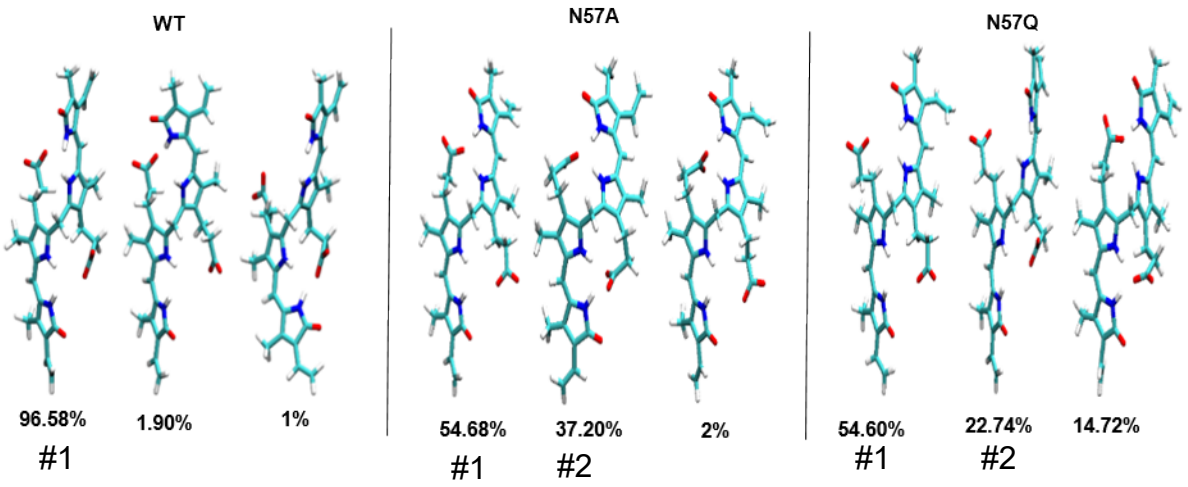

Figure ESI-12: Molecular structure of BLR along the dynamics for each system were clustered to obtain to major conformational states. The similarity RMSD cutoff was used as 0.1 nm. One major cluster of WT-BLR (96.58%), was found throughout the simulations.

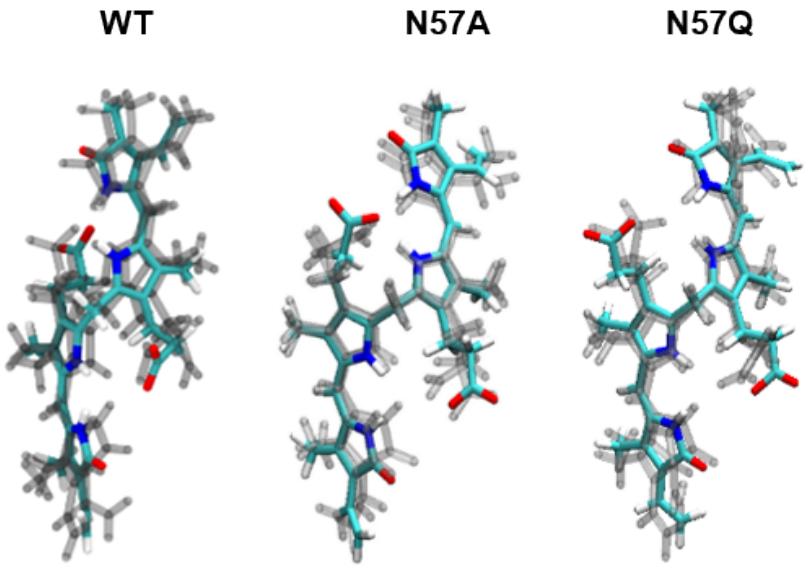

Figure ESI-13: Superimposed conformations of BLR in WT , N57A, and N57Q across the major and minor clusters revealing that compound WT-BLR remains in a distinct rigid state. In contrast, compounds N57A and N57Q present clear variations in their molecular arrangements. Conformations are dumped at several time points (shown in gray), and the major clauter (in cpk color) are shown to represent the conformational variability of the BLR in the UnaG bound state. The protein and water are not shown for clarity.

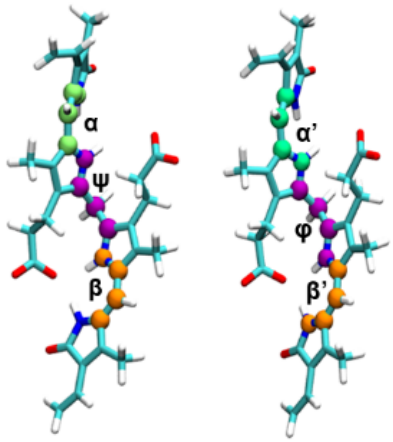

Table ESI-3: Average value of BLR dihedrals (see nomenclature in Figure ESI-11) along the simulation (Avg) and within each cluster having a population higher than 20 % for each investigated system.

		WT		N57A			N57Q		
	Dihedral angles	Avg	Cluster #1	Avg	Cluster #1	Cluster #2	Avg	Cluster #1	Cluster #2
A-B rings	α	3.0	6.81	-0.23	-3.11	6.80	-0.11	16.38	1.16
	α′	3.0	5	11.13	2.58	-1.23	1.2	-15.49	36.93
B-C rings	ψ	119.80	116.88	88.24	94.41	78.26	113.3	102.32	124.09
	φ	119.80	125.25	87.06	91.26	77.97	105.9	106.41	125.01
C-D rings	β	11.28	13.6	12.0	16.86	7.47	-3.16	-0.53	10.76
	β′	11.28	-2.88	1.49	1.16	1.13	4.8	4.54	-5.50

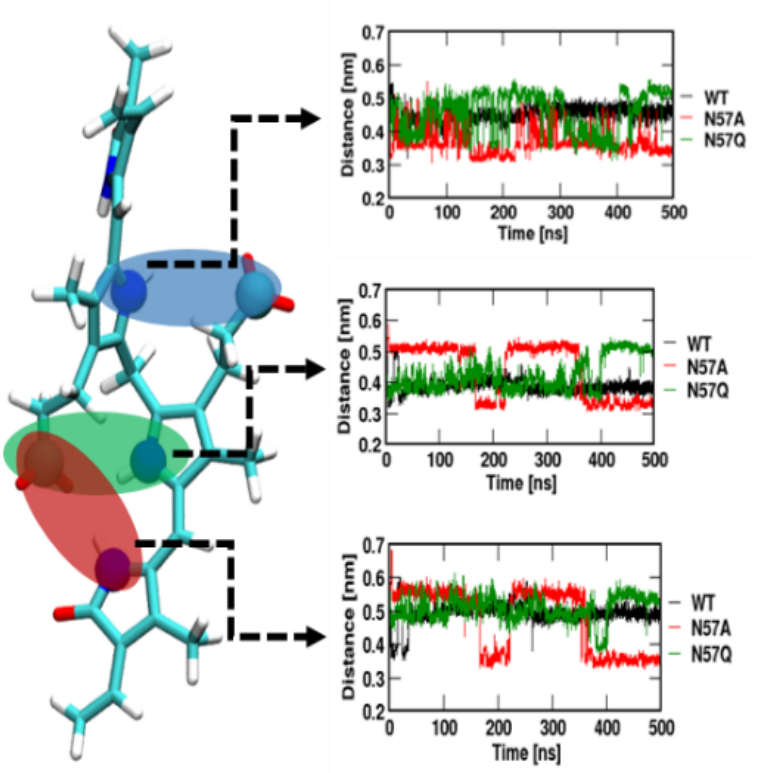

Figure ESI-14: Intramolecular interactions: distance between the carboxylate and NH functional group as obtained from the dynamics of each system.

5. ONIOM QM/MM calculations

In ONIOM calculations, the 4 Å residues around BLR are PHE5, PHE17, LEU21, LEU30, LEU41, VAL53, ILE55, ASN/GLN/ALA57, THR61, LEU63, THR65, VAL67, PHE69, GLU77, PRO79, SER80 ASP81, LEU97, ARG112, LEU119, VAL121, ARG132, TYR134

Table ESI-4 detailed information of two layers ONIOM scheme (no water molecules are considered herein).

X-ray	λ_{max} (nm)	*f*	Molecular Orbitals	MD clusters	λ_{max} (nm)	*f*	Molecular Orbitals
WT	459	1.42	154 -> 156 -0.29846 154 -> 157 0.22502 **155 -> 156 0.57472** 155 -> 157 0.14034	**WT (96%)**	455	1.58	154 -> 156 0.11846 154 -> 157 -0.30885 **155 -> 156 0.58124** 155 -> 157 0.20508
N57A	451	1.46	154 -> 157 -0.26706 **155 -> 156 0.61313** 155 -> 157 0.19809	**N57A (54%)**	441	1.56	154 -> 156 -0.11301 154 -> 157 0.33050 **155 -> 156 0.56957** 155 -> 157 0.19002
				N57A (37%)	443	1.31	154 -> 156 0.30108 154 -> 157 0.29085 **155 -> 156 0.50926** 155 -> 157 -0.21376
N57Q	452	1.65	154 -> 156 -0.24020 154 -> 157 -0.22726 **155 -> 156 0.53963** 155 -> 157 -0.29712	**N57Q (54%)**	452	1.55	154 -> 156 0.16775 154 -> 157 0.27509 **155 -> 156 0.59744** 155 -> 157 -0.16684
				N57Q (23 %)	460	1.64	154 -> 157 -0.30802 **155 -> 156 0.62252**

Table ESI-5 BLR dihedral angles after optimization with the ONIOM scheme onto representative snapshots from MD simulations.

System		No water molecules	Water molecules		No water molecules	Water molecules
#1 WT (96%)	**α**	-4.48	2.43			
	α'	9.32	7.95			
	Φ	113.01	122.74			
	Ψ	103.36	117.81			
	β	10.7	21.17			
	β'	-1.0	9.47			
#1 N57A (54%)	**α**	2.84	-1.89	#2 N57A (37%)	**α**	0.65	-1.81
	α'	-7.19	-5.30		**α'**	17.31	26.72

	Φ	103.54	105.85		**Φ**	72.39	68.17
	Ψ	88.95	70.38		**Ψ**	69.47	75.29
	β	9.22	37.40		**β**	17.18	27.37
	β'	2.56	0.21		**β'**	-4.8	2.93
#1 N57Q (54%)	**α**	2.56	-6.61	#2 N57Q (22%)	**α**	8.99	4.81
	α'	-2.24	-0.65		**α'**	26.32	24.14
	Φ	102.97	91.22		**Φ**	118.30	121.40
	Ψ	101.42	104.35		**Ψ**	115.93	111.81
	β	6.62	11.35		**β**	8.68	20.49
	β'	-3.25	5.71		**β'**	2.17	9.12

Table ESI-6 detailed information of two layers ONIOM scheme.

X-ray structure							**MD representative clusters**						
	λ_{max} (nm)	*f*	ψ φ	λ_{max} (nm)	*f*	ψ φ		λ_{max} (nm)	*f*	ψ φ	λ_{max} (nm)	*f*	ψ φ
	Without water			With water				Without water			With water		
WT	459	1.42	124.8 100.5	456	1.68	119.0 110.5	**WT (96%)**	455	1.58	113.6 103.3	461	1.90	122.7 117.8
N57A	451	1.46	123.7 108.9	445	1.13	103.5 93.5	**N57A (54%)**	441	1.56	103.5 88.9	427	1.91	105.85 70.38
							N57A (37%)	443	1.31	72.4 69.5	444	1.44	68.2 75.3
N57Q	452	1.65	104.8 96.4	453	1.30	131.70 104.97	**N57Q (54%)**	452	1.55	103.0 101.4	443	1.50	91.2 104.4
							N57Q (22 %)	460	1.64	118.3 115.9	461	1.39	121.4 111.8

References

1 B. J. Grant, L. Skjaerven and X.-Q. Yao, *Protein Sci*, 2021, **30**, 20–30.

2 A. Kumagai, R. Ando, H. Miyatake, P. Greimel, T. Kobayashi, Y. Hirabayashi, T. Shimogori and A. Miyawaki, *Cell*, 2013, **153**, 1602–1611.

3 R. Kumari, R. Kumar, Open Source Drug Discovery Consortium and A. Lynn, *J Chem Inf Model*, 2014, **54**, 1951–1962.