Electronic Supplementary Material (ESI) for Faraday Discussions.

Supporting Information

Optical Backbone-Sidechain Charge Transfer Transitions in Proteins Sensitive to Secondary Structure and Modifications

*Imon Mandal†, Sanjoy Paul†, Ravindra Venkatramani†**

† Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India.

* Corresponding Author

CONTENTS

Page Number

Figure S1: Inter-residue sidechain contact maps for MD and NMR solution structures of ubiquitin

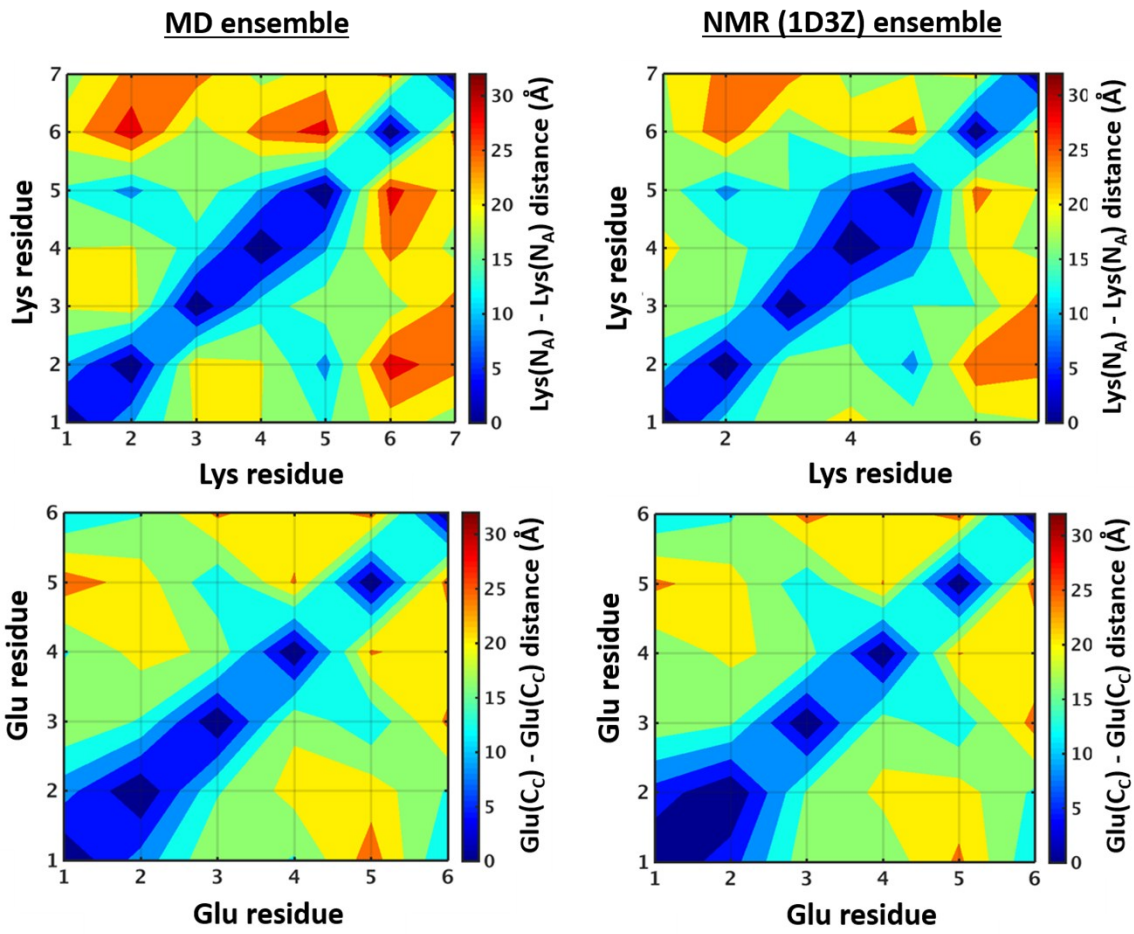

Figure S1: Contact maps representing average distances between amino nitrogen (N_A) atoms for all pairs of Lys residues (top panels) and carboxylate carbon (C_c) atoms for all pairs of Glu residues (bottom panels) in ubiquitin. The panels on the left show contact maps for our MD simulation ensemble. The panels on the left show contact maps for an NMR derived ensemble (PDB id: 1D3Z).

Figure S2: Variation of geometric and spectral parameters over the sampled MD conformations of Lys

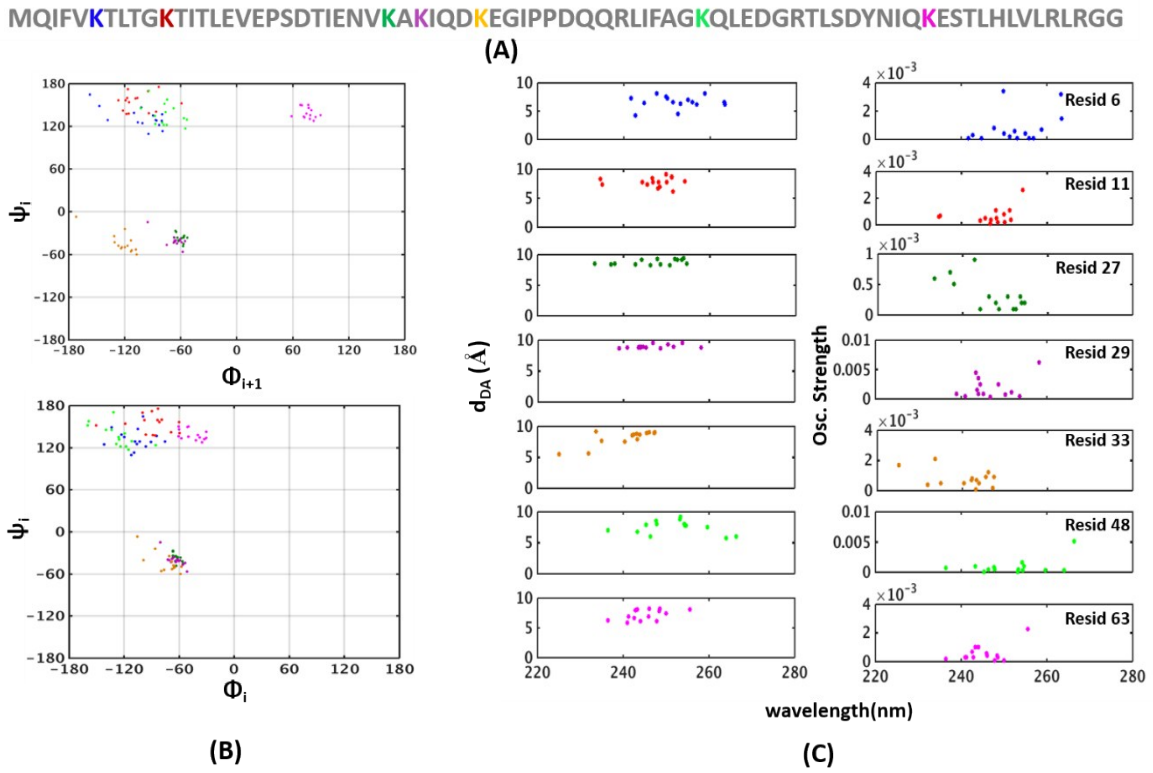

Figure S2: For our spectra calculations, we sampled 14-15 conformations along the 1 μs MD trajectory (protein snapshots captured every 70 ns) for all 7 possible instances of Lys residues in the protein sequence leading to a total of 100 monomer amino acid conformations. Panel (A) shows the positions of the 7 Lys residues in the ubiquitin sequence with different colors assigned to each instance. Panels (B) show the variation of Ramachandran angles (Φ,Ψ) over the 100 Lys conformations. Ramachandran angle data for each instance of Lys in the ubiquitin sequence is shown using colors which are assigned in panel (A). The top panel in (B) shows Ramachandran angles (Φ_{i+1},Ψ_i) defined for the extended backbone monomer Lys models in Figure 2 . The bottom panel shows corresponding Ramachandran angles (Φ_i,Ψ_i) for Lys residues in ubiquitin. Panel (C) shows the variation of the distance between the extended backbone and the sidechain amino group (d_{DA} defined in Figure 2C) and oscillator strengths for the lowest energy transitions as a function of the lowest energy excitation wavelength. Data for each instance of Lys in the ubiquitin sequence is plotted in different panels using colors which are assigned in panel (A).

Figure S3: Variation of geometric and spectral parameters over the sampled MD conformations of Arg

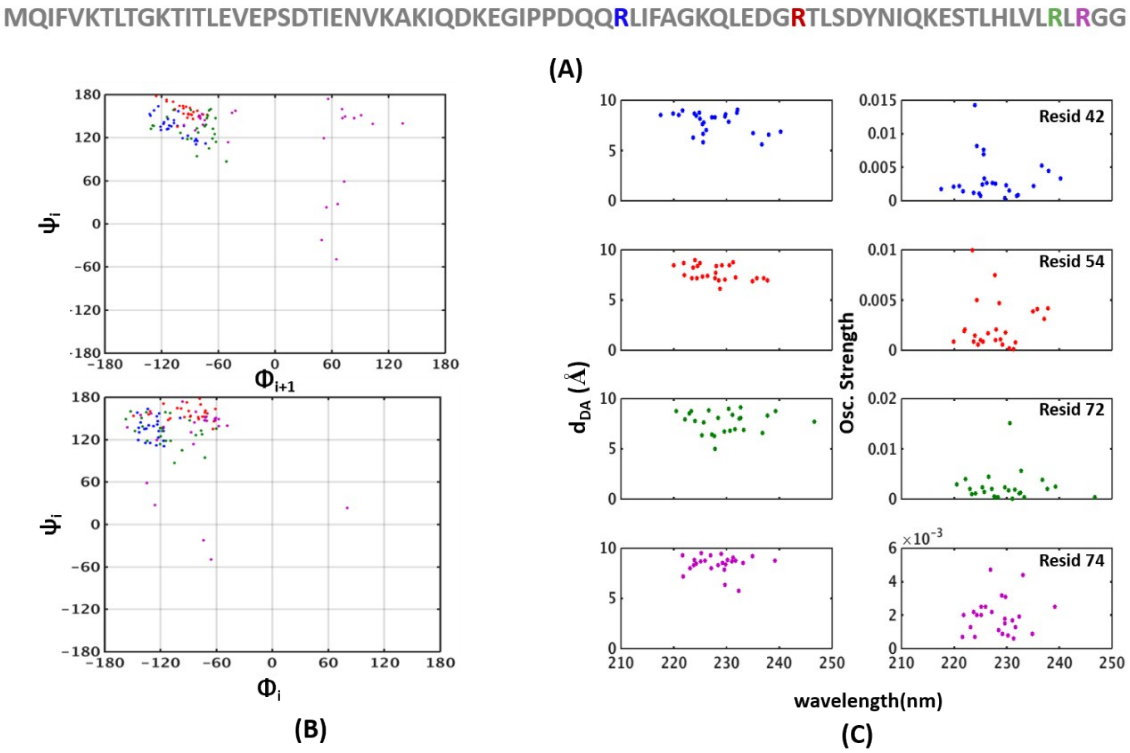

Figure S3: For our spectra calculations, we sampled 25 conformations along the 1 μs MD trajectory (protein snapshots captured every 40 ns) for all 4 possible instances of Arg residues in the protein sequence leading to a total of 100 monomer amino acid conformations. Panel (A) shows the positions of the 4 Arg residues in the ubiquitin sequence with different colors assigned to each instance. Panels (B) show the variation of Ramachandran angles (Φ,Ψ) over the 100 Arg conformations. Ramachandran angle data for each instance of Arg in the ubiquitin sequence is shown using colors which are assigned in panel (A). The top panel in (B) shows Ramachandran angles (Φ_{i+1},Ψ_i) defined for the extended backbone monomer Arg models in Figure 2 . The bottom panel shows corresponding Ramachandran angles (Φ_i,Ψ_i) for Arg residues in ubiquitin. Panel (C) shows the variation of the distance between the extended backbone and the sidechain guanidine group (d_{DA} defined in Figure 2C) and oscillator strengths for the lowest energy transitions as a function of the lowest energy excitation wavelength. Data for each instance of Arg in the ubiquitin sequence is plotted in different panels using colors which are assigned in panel (A).

Figure S4: Variation of geometric and spectral parameters over the sampled MD conformations of Glu

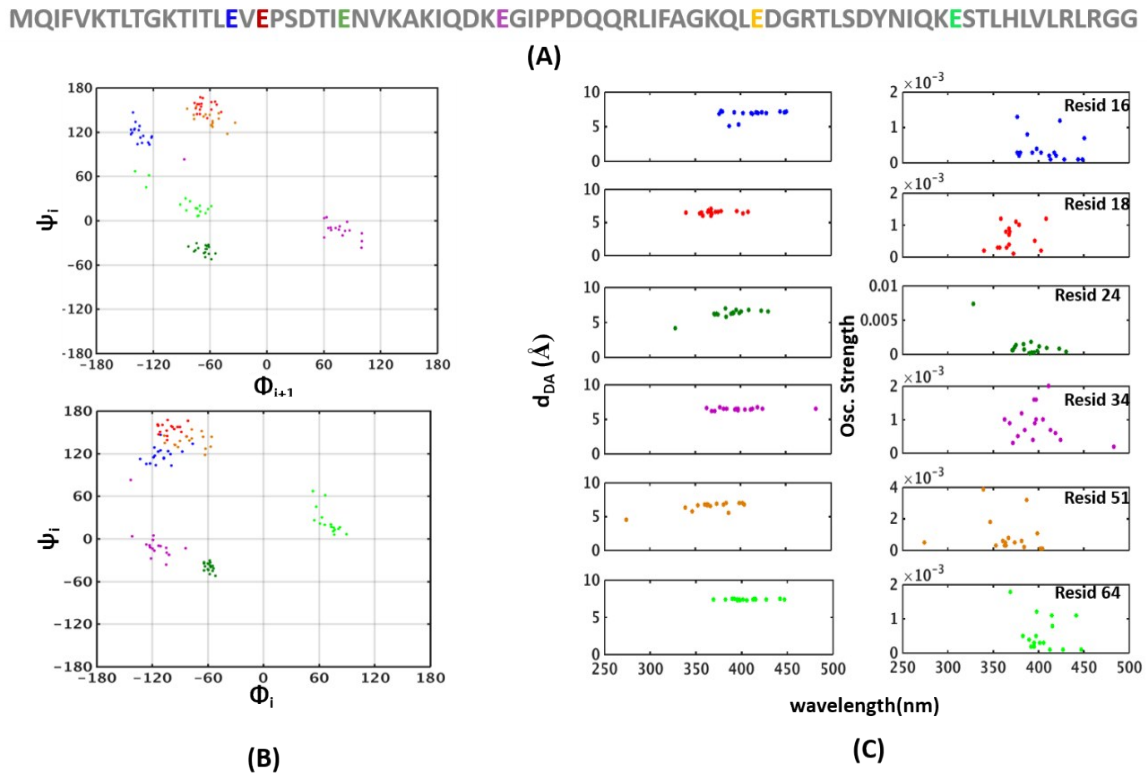

Figure S4: For our spectra calculations, we sampled 16-17 conformations along the 1 μs MD trajectory (protein snapshots captured every 60 ns) for all 6 possible instances of Glu residues in the protein sequence leading to a total of 100 monomer amino acid conformations. Panel (A) shows the positions of the 6 Glu residues in the ubiquitin sequence with different colors assigned to each instance. Panels (B) show the variation of Ramachandran angles (Φ,Ψ) over the 100 Glu conformations. Ramachandran angle data for each instance of Glu in the ubiquitin sequence is shown using colors which are assigned in panel (A). The top panel in (B) shows Ramachandran angles (Φ_{i+1},Ψ_i) defined for the extended backbone monomer Glu models in Figure 2 . The bottom panel shows corresponding Ramachandran angles (Φ_i,Ψ_i) for Glu residues in ubiquitin. Panel (C) shows the variation of the distance between the extended backbone and the sidechain carboxylate group (d_{DA} defined in Figure 2C) and oscillator strengths for the lowest energy transitions as a function of the lowest energy excitation wavelength. Data for each instance of Glu in the ubiquitin sequence is plotted in different panels using colors which are assigned in panel (A).

Figure S5: Variation of geometric and spectral parameters over the sampled MD conformations of Asp

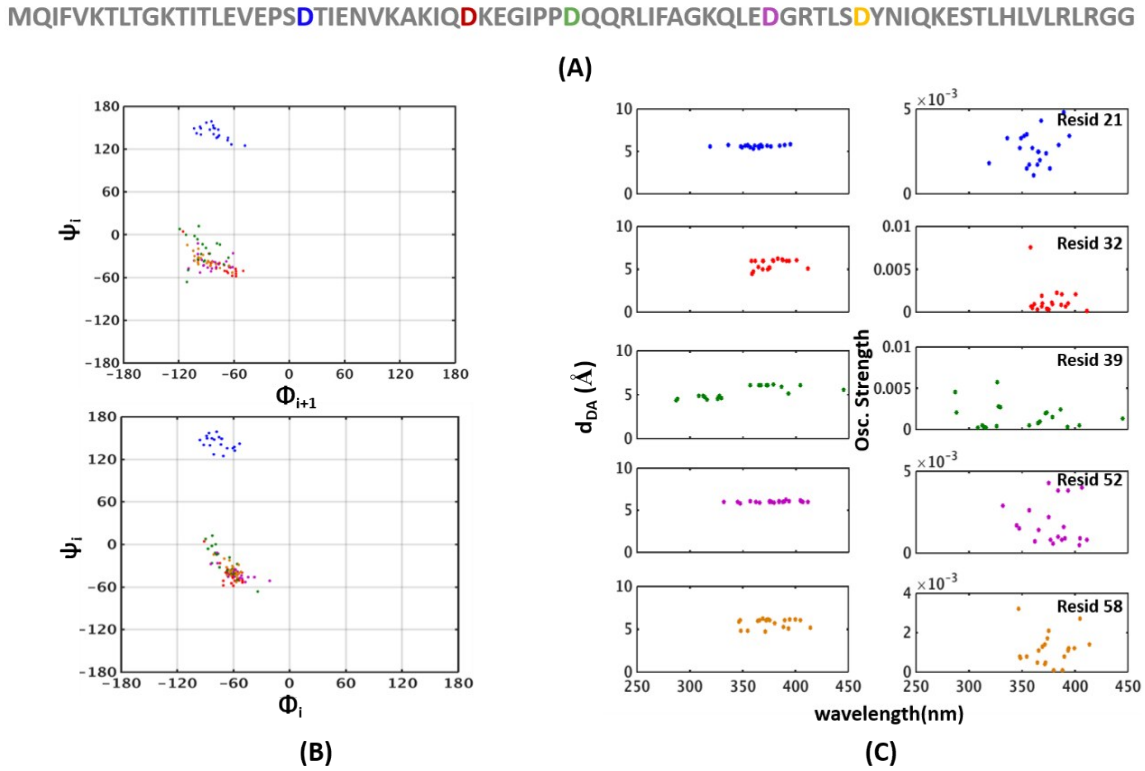

Figure S5: For our spectra calculations, we sampled 20 conformations along the 1 μs MD trajectory (protein snapshots captured every 50 ns) for all 5 possible instances of Asp residues in the protein sequence leading to a total of 100 monomer amino acid conformations. Panel (A) shows the positions of the 5 Asp residues in the ubiquitin sequence with different colors assigned to each instance. Panels (B) show the variation of Ramachandran angles (Φ,Ψ) over the 100 Asp conformations. Ramachandran angle data for each instance of Asp in the ubiquitin sequence is shown using colors which are assigned in panel (A). The top panel in (B) shows Ramachandran angles (Φ_{i+1},Ψ_i) defined for the extended backbone monomer Asp models in Figure 2 . The bottom panel shows corresponding Ramachandran angles (Φ_i,Ψ_i) for Asp residues in ubiquitin. Panel (C) shows the variation of the distance between the extended backbone and the sidechain carboxylate group (d_{DA} defined in Figure 2C) and oscillator strengths for the lowest energy transitions as a function of the lowest energy excitation wavelength. Data for each instance of Asp in the ubiquitin sequence is plotted in different panels using colors which are assigned in panel (A).

Figure S6: Control calculations: Capping scheme used for TDDFT calculations of monomer units

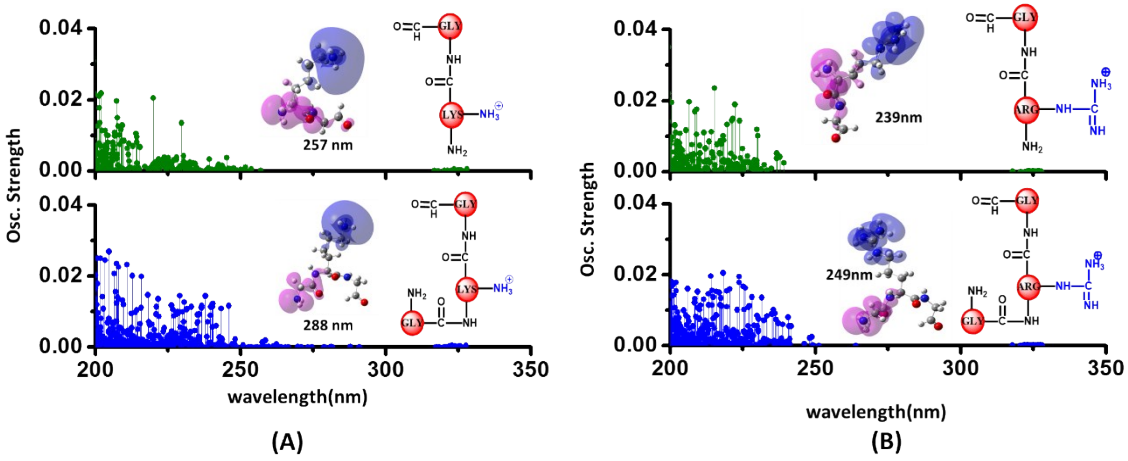

Figure S6: Comparison of spectra from a 30 conformations for Lys (left panels) and Arg (right panels) with two different backbone capping strategies which create an extended backbone model mimicking a polypeptide chain: (top panel) model used in the present study, (bottom panel) extended backbone with Gly units appended symmetrically. Difference density plots for the red-most transitions are shown for each panel. For both Lys and Arg, the lowest energy transitions are still PBS-CT transitions in the top and bottom panels. The spectral range for both Lys and Arg chromophores is extended by a few 10s of nm to the red. However relative trends of oscillator strengths and spectral range between the cationic species and nature of the lowest energy transitions for these chromophores obtained by the capping scheme used in our study are not significantly changed by further extending the backbone. In our previous study (reference 14 of the main manuscript we carried out control calculations on Glu to confirm that extending the backbone did not alter either the spectral range or nature of the transitions for anionic species. The control studies presented here and in reference 14 show that the conclusions on the spectral features of the PBS-CT transitions in the present study will not be affected by extending the backbone in our fragment models.

Figure S7: Decomposition of Lowest Energy PBS-CT Transitions for Glu, Asp, pSer, pThr, and pTyr

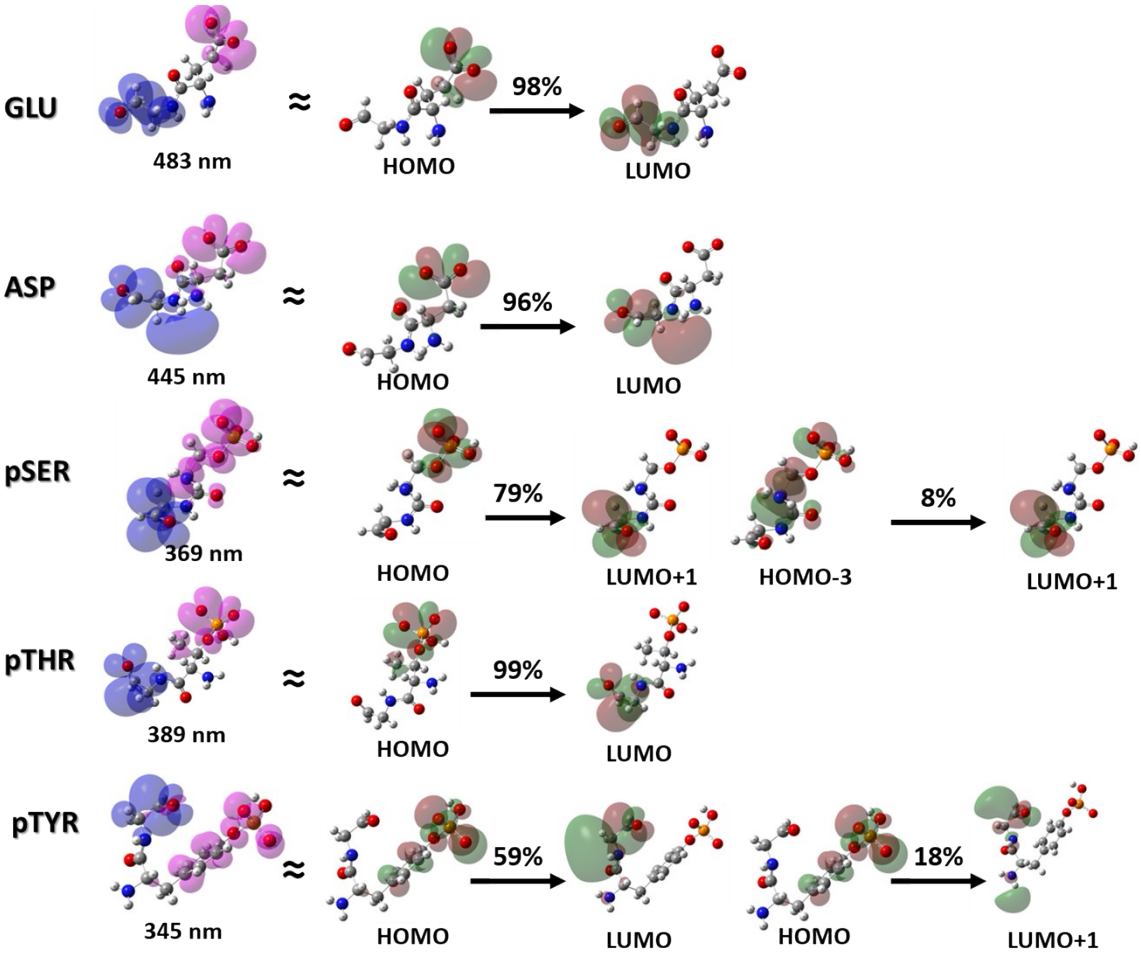

Figure S7: Dominant molecular orbital (MO) pair contributions for the transitions in anionic species depicted in Figure 4 (main manuscript) through difference density plots.

Figure S8: Decomposition of Lowest Energy PBS-CT Transitions for Lys, Arg, Hsp, and His

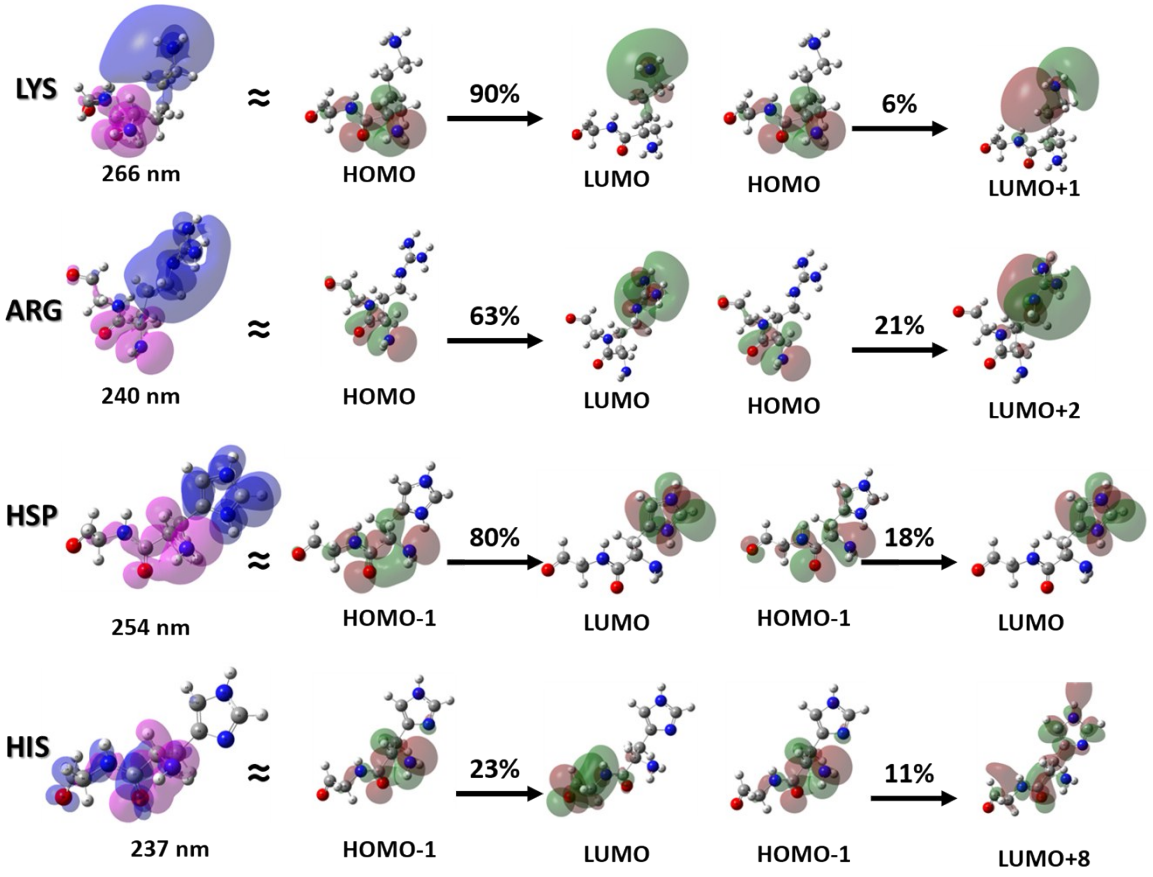

Figure S8: Dominant molecular orbital (MO) pair contributions for the transitions in cationic species depicted in Figure 6 (main manuscript) through difference density plots.

Figure S9: Correlation of excitation wavelength of PBS-CT transitions in charged amino acids with ground state energy gap and $1/D_{CT}$

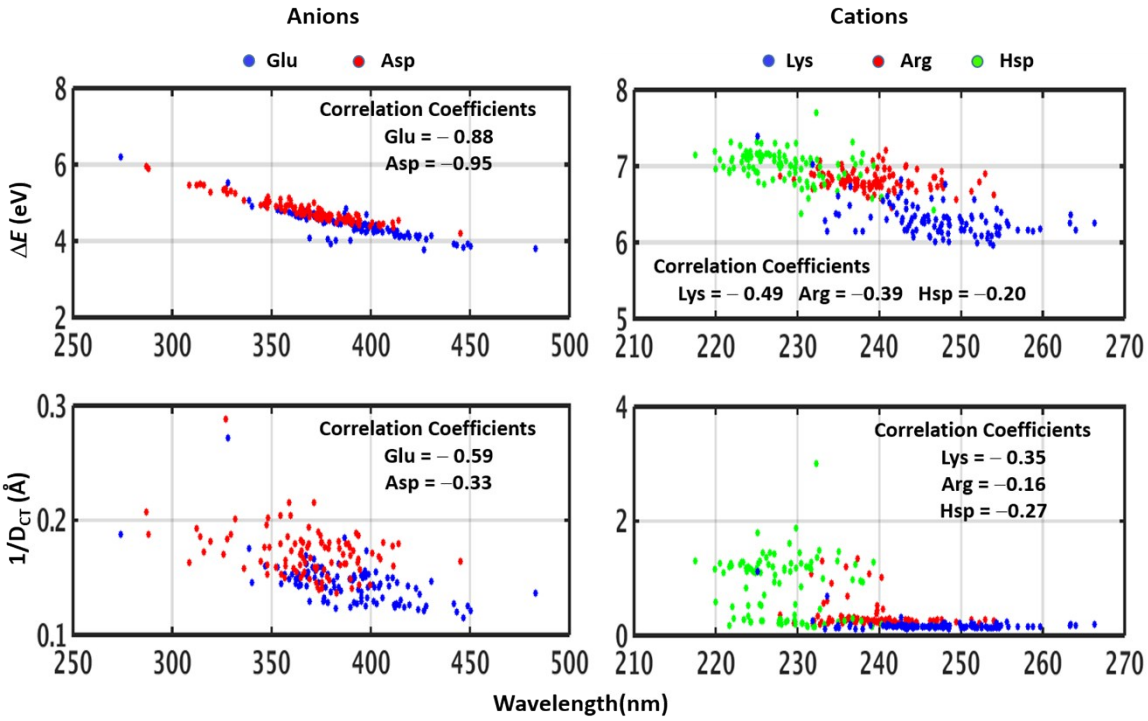

Figure S9: Correlation plots of the HOMO-LUMO energy gap (top panels) and inverse of the hole-electron separation distance (bottom panels) vs excitation wavelength for the lowest energy transitions extracted from 100 conformations of anionic (left panels) and cationic (right panels) amino acids