Speaker: Germaine NEZA HOZANA (African Institute for Mathematical Sciences (AIMS Rwanda), Rwanda)

Title: "Data-Driven Discovery of the Origins of UV-Absorption in Alpha-3C Protein"

Authors: Germaine Neza Hozana (University of Trieste - ICTP), Gonzalo Diaz Miron (ICTP) and Ali Hassanali (ICTP)

Abstract: Over the last decade, there has been a growing body of experimental work showing that proteins devoid of aromatic and conjugated groups can absorb light in the near-UV beyond 300 nm and visible light. Understanding the origins of this phenomena offers the possibility of designing non-invasive spectroscopic probes for local interactions in biological systems. It was recently found that the synthetic protein α3C displays UV-vis absorption between 250-800 nm which was shown to arise from charge-transfer excitations between different types of charged amino acids. In this work, we use data-driven discovery to revisit the origins of these features using molecular dynamics and excited-state simulations. Specifically, an unsupervised learning approach beginning with encoding local environments with the smooth overlap of atomic positions (SOAP) descriptors, is employed to automatically detect relevant structural motifs. We identify three main motifs corresponding to different hydrogen-bonding patterns that are subsequently used to perform QM/MM simulations including the entire protein and solvent bath with the density-functional tight-binding (DFTB) approach. Hydrogen-bonding structures involving arginine and carboxylate groups appear to be the most prone to near-UV absorption. We show that these features are highly sensitive to the size of the QM region employed as well as to the inclusion of explicit solvation underscoring the limitations of using gas-phase cluster models.