Tata Institute of Fundamental Research
Survey No. 36/P, Gopanpally Village, Serilingampally, Ranga Reddy Dist., Hyderabad – 500 046

Students' Annual Seminar

Unbiased learning of protein's structural representation via unsupervised random forest

Mohammad Sahil

Describing protein's conformational dynamics by single or few collective variables prevents comprehensive understanding of its function. Hence, high dimensional representation of protein conformation like contact metrics are routinely being employed to study functions like allostery and building models like MSMs. But many times this led to introduction of noise in the input data, thereby not yielding any functional insights. Through our recent and ongoing work, I shall discuss how we achieved shortlisting of high dimensional protein representation via supervised and unsupervised versions of Random Forest algorithm, using multiple protein systems like T4L, MopR, P450, a-syn etc. By successfully implementing unsupervised Random Forest (URF) in python, this pipeline can be used on MD generated trajectories to study higher order protein functions like allostery and building better MSM models.

Friday, Jan 12th 2024

14:30 Hrs (Tea / Coffee 14.15 Hrs)

Seminar Hall, TIFR-H