MONDAY

COLLOQUIUM

Mapping the Chemical Space with Data-Driven Computational Chemistry

Raghunathan Ramakrishnan (TIFR, Hyderabad)

13 Oct 2025 (Monday) | 16:00 Hrs (Tea / Coffee 15:45 Hrs) | Venue: TIFRH Auditorium

Chemical space is the set of all possible molecules that can, in principle, exist according to the laws of chemistry and physics. Like outer space, it is effectively infinite because the number of ways atoms can combine, arrange, and bond grows exponentially with molecular size. The challenge of navigating this vast landscape lies not in imagination but in tractability. Although the number of conceivable molecules is enormous, progress has long been limited by data scarcity. The advent of QM9, a consistent and comprising more than 130,000 organic molecules, transformed this landscape by shifting exploration from enumeration to statistical learning and hypothesis generation. Building on such datasets, approaches such as machine-learned potentials, active learning, and generative modelling now accelerate molecular discovery across many research domains. These datasets have also enabled the testing of unconventional hypotheses, such as molecules that appear to violate Hund's rule, offering new directions for the design of next-generation optoelectronic materials.

When QM9 was first conceived in 2014, its goal was not to answer specific questions but to enable entirely new ones. Since then, it has fuelled unforeseen developments across scientific and creative disciplines, from identifying molecules with extremely long carbon-carbon single bonds and graph-based fuel design to benchmarking data imbalance in drug discovery and mapping molecules to music. These diverse applications illustrate how a well-curated computational dataset can transcend its original purpose, evolving from a benchmark into a catalyst of discovery that continues to reshape the connection between chemistry, data, and imagination.

tifr HYDERABAD | TATA INSTITUTE OF FUNDAMENTAL RESEARCH