MATERIALS SCIENCE AND TECHNOLOGY

AMMIN - Data Driven Approaches for Complex Molecular Systems and Materials

Funded By	Politecnico di TORINO [P.iva/CF:00518460019]
Supervisor	PAVAN GIOVANNI MARIA - giovanni.pavan@polito.it
Contact	
Context of the research activity	This research aims to develop and apply data-driven frameworks useful to characterize complex molecular systems and materials by integrating machine learning and information-theory approaches to extract from multiscale molecular simulations information relevant for understanding their behavior. The project aims to propose data-driven approaches to uncover the key features that govern emergent behaviors in different types of self-assembling systems/materials and discover models to predict them.
	Many molecular systems, both natural and synthetic, display a high degree of complexity, which arises not from the intricacy of their components, but from the dynamic and collective behavior that emerges through their interactions. These molecular components can recognize each other, assemble and disassemble, and adapt in response to external stimuli. Through continuous interactions with one another and with the surrounding environment, they give rise to highly dynamic architectures such as supramolecular fibers, micelles, and host–guest assemblies. Understanding the structural determinants and dynamic responses of such systems is essential to advance our ability to design responsive, functional materials. This PhD project aims to develop and apply data-driven strategies to characterize and understand the behavior of complex molecular systems. The core idea is to combine high-resolution molecular dynamics simulations, both atomistic and coarse-grained, with advanced machine learning and advanced statistical methods to extract interpretable insights from simulation datasets. These systems often exhibit nontrivial dynamics, including time-dependent exchange of components, collective motions, and emergent behavior across scales. Traditional analysis methods are insufficient to capture such complexity, requiring the integration of new tools from the data science domain. A central focus of the project is on quantifying the information content embedded in structural and dynamical features in different types of self-organizing materials/systems, and on measuring information gain as a function of time, resolution, or descriptor. This approach will allow us to determine which aspects of a system’s behavior carry the most chemically or

Objectives	physically relevant insight, and how these depend on molecular architecture and environmental conditions. Methods such as dimensionality reduction, unsupervised clustering, and time-series analysis will be employed to identify key descriptors and develop predictive models of dynamic behavior. While the primary data source will be molecular simulations, the methodologies developed may be extended, in principle, to the analysis of experimental datasets, such as time-resolved microscopy or video recordings of the trajectory of self-organizing systems of different scales. These complementary data streams can be integrated to refine models and validate computational predictions, further enhancing the robustness and relevance of the analysis. Examples of systems to be investigated include supramolecular fibers, micelles, interlocked molecules, such as rotaxanes, and molecular shuttles, which are dynamic architectures central to the development of next-generation molecular machines. The project will explore how molecular design and external stimuli influence assembly mechanisms, structural variability, and dynamic adaptation. Multiscale simulations, including enhanced sampling methods such as metadynamics, will generate the required data, while machine learning methods will be used to extract patterns and identify the most informative features governing behavior. The PhD candidate will join an interdisciplinary environment with strong international collaborations. The project lies at the interface between chemistry, physics, and data science, and offers a unique opportunity to develop expertise in computational modeling, molecular simulation, and information-theoretic approaches to complex systems. The long-term objective is to contribute to a general framework for decoding and steering the dynamic behavior of complex molecular systems through the lens of data and information.

Skills and competencies for the development of the activity	• Experience in molecular modeling and molecular simulation techniques (molecular dynamics, etc.) • Familiarity in the use of molecular modeling and molecular dynamics software (e.g., GROMACS, LAMMPS, etc.) • Strong motivation for scientific research • Solid background in chemistry, physics, or materials science • Good scientific programming skills (e.g., python) • Experience in working in the Linux environment • Prior experience in machine learning approaches • Good written and spoken English