

X-informatics: A (very) personal view of interesting endpoints to model

Hanoch Senderowitz

Department of Chemistry, Bar Ilan University, Ramat Gan, 5290002, Israel

Cheminformatics emerged in the late 20th century at the intersection of chemistry, computer science, and information theory, driven by the need to store, search, and analyze rapidly growing collections of chemical structures in order to address chemistry-related questions. Early developments focused on molecular representation, database systems, and structure–activity relationships, laying the foundation for modern computational modeling in chemistry. While cheminformatics has been most prominently applied in drug discovery, its concepts and methodologies have progressively spread into other scientific domains, giving rise to related disciplines such as agro-informatics, materials informatics, and forensic informatics. Although these fields differ in their applications, they share a common assumption: that the objects of interest can be meaningfully encoded into descriptors suitable for modeling. At the same time, both the nature of these objects and the associated modeling tasks vary substantially across disciplines, leading to distinct and domain-specific challenges.

In this talk, I offer a deliberately personal perspective on endpoints encountered across these domains, focusing less on algorithmic considerations and more on the often-overlooked question of representation. Through selected examples, I will highlight how the choice of descriptors—whether molecular fingerprints, materials features, environmental variables, or encodings of forensic evidence—fundamentally shapes what can be learned, predicted, or even asked. In many cases, the primary limitation is not model capacity, but the adequacy and faithfulness of the underlying representation, particularly when dealing with heterogeneous, noisy, or poorly defined systems.

As a final take-home message, I will emphasize the importance of close collaboration between computational and experimental disciplines across all cheminformatics-related fields, both to better interpret observed trends and to fully realize the potential of data-driven approaches.