Protein Crystallography with Speed and Finesse

Wladek Minor

University of Virginia

Abstract

X-ray crystallography is one of the most detailed “microscopes” available today for examining macromolecular structures. However, structures are only simplified models of target proteins and/or nucleic acids, and should be seen as a framework for generating hypotheses to be explored. Numerous biochemical and biophysical experiments, including new diffraction experiments, can and should be performed to verify or falsify these hypotheses. Processing of structural information, particularly when combined with functional, experimental, and sequential data in the context of pathways or interaction networks with other bio-macromolecules and/or bioactive chemical compounds, increasingly requires the use of Big Data paradigms for effective data management, as well as for checking data integrity and accuracy. This is easy to say but extremely difficult to implement, as development of these tools takes time, effort, and most of all, creativity of the leaders and developers of these projects. This is more expensive than just the purchase of supercomputers with petabyte storage. A combination of advancements in high-quality data harvesting, validation, mining, and data management tools would make it possible to convert high-throughput pipelines into high-output pipelines in target-based drug discovery and academic biomedical research.