Invited Lecture

AI and Structural Biology: The Future of Biomolecular Structure Determination

Andrea Thorn[1]

[1]*Institut für Nanostruktur und Festkörperphysik, Universität Hamburg, Luruper Chaussee 149 (Bldg. 610 - HARBOR), Germany*
andrea.thorn@uni-hamburg.de

Structural biology is key to understanding basic processes of life and a major driver for the development of new therapies. However, our structures are merely models which interpret and explain the observed experimental data according to *a priori* knowledge. Consequently, our structures are only as good as our limited understanding of the underlying principles.

Machine learning methods, in particular convolutional neural networks, have been applied to a variety of problems in crystallography, automation and high-throughput the foremost among them [1]. However, they suffer from the same limitations as more manual structure determinations. This is easily illustrated by AlphaFold2, which is trained on sequence information and PDB-deposited molecular models, but not the underlying experimental data, which differ on average from the models by 24%. Hence, prior problems add to the limited acceptance of these "black boxes" by the community and/or the “black box” feel of many software.

Whether we will be able to exploit this potential fully will depend on the manner in which we use machine learning: training data must be well-formulated, methods need to utilize appropriate architectures, and outputs must be critically assessed, which may even require explaining artificial intelligence decisions. In this talk, an overview of current applications of machine learning in structural biology will be given, including examples from our own work, how experimentalist may use fold prediction methods, and how AI could change crystallography in the future.

[1] Thorn, A.* (2022). *Curr. Opin. Struct. Biol.* **74**, https://doi.org/10.1016/j.sbi.2022.102368.

[2] Edich, M., Briggs, D.C., Kippes, O., Gao, Y., Thorn, A.* (2022). *Faraday Discussions* **240**, 184−195, https://doi.org/10.1039/D2FD00072E.

This work was supported by the German Federal Ministry of Education and Research [grant number 05K19WWA and 05K22GU5], Deutsche Forschungsgemeinschaft [project TH2135/2-1