Invited Lecture

Advancing protein complex analysis with AI and electron diffraction

S.V. Matinyan[1*], P. Filipcik[1], E. van Genderen[2], J.P. Abrahams[1,2*]

[1] *Biozentrum, University of Basel, Switzerland*
[2] *Laboratory of Nanoscale Biology, Paul Scherrer Institute, Switzerland*
jp.abrahams@unibas.ch, senik.matinyan@unibas.ch

In our research group, we focus on improving the analysis of large protein complexes using electron diffraction. Our aim is to achieve higher resolution for flexible proteins, surpassing the limitations of current methods.

One major challenge we face is the vast amount of data generated by our equipment, which is crucial for structural analysis. This makes automated data triage essential to filter out useful data from noise. A novel, ultrafast lossless data compression algorithm has been developed to solve this issue, achieving 10 Gbit/s compression in single-threaded, in-memory processes, with equal or better compression efficiency than conventional data compression techniques [1].

Our scanning process uses a continuous beam and fast scan, with the probability of a protein hit being dependent on beam size and sample distribution. We developed a set of machine learning tools to automatically select diffraction frames originating from protein positions [2]. These advancements address the need for efficient data management in electron diffraction studies.

We are further developing novel algorithms based on deep learning to interpret electron diffraction data and extract valuable structural information. We used conditional generative adversarial networks (GANs) to phase high-resolution diffraction data using low-resolution images [3]. This approach is currently under further development to tackle potential experimental shifts.

Although we are still in the early stages of directly determining protein structures through electron diffraction, we are confident in its potential to achieve better signal-to-noise ratios than conventional methods.

[1] Matinyan S, Abrahams JP. terse/prolix (trpx) – a new algorithm for fast and lossless compression and decompression of diffraction and cryo-em data. Acta Crystallogr A Found Adv. 2023;79(6):536-541.
[2] Matinyan S, Demir B, Filipcik P, Abrahams JP, van Genderen E. Machine learning for classifying narrow-beam electron diffraction data. Acta Crystallogr A Found Adv. 2023;79(4):360-368.
[3] Matinyan S, Filipcik P, van Genderen E, Abrahams JP. DiffraGAN: a conditional generative adversarial network for phasing single molecule diffraction data to atomic resolution. Front Mol Biosci. 2024;11:1386963.