MS35-1-1 Data-driven approach for the solution of the phase problem in crystallography: first insights #MS35-1-1

T. Rekis [1], A.S. Larsen [1], A.Ø. Madsen [1]
[1]University of Copenhagen - Copenhagen (Denmark)

Abstract
For a crystal structure to be solved from the diffraction data, it is necessary to obtain the complex structure factors, $\boldsymbol{F}_{\mathbf{H}}$, of the measured reflections. Their squared moduli, $|\boldsymbol{F}_{\mathbf{H}}|^2$, are proportional to the measured intensities, but the phase angles to reconstruct the complete complex numbers cannot be determined experimentally. Since the exact solution to this phase problem is not known, several methods to overcome this crucial step in crystal structure determination have been developed over time. For example, direct methods or charge flipping algorithm can be used for most organic, inorganic, and metal-organic structures. Nevertheless, these methods fail if, for example, the available data resolution is not sufficiently high. Furthermore, they cannot be used to solve protein structures intrinsically having large unit cells with hundreds of atoms in the asymmetric units.
In our study, we have attempted to solve the phase problem using a neural network. We focused on small organic molecule and metal-organic structures in the most common space group $P2_1/c$. Millions of fictive structures containing metal atoms and/or molecular fragments were generated to train the neural network. Validation set consisted of over a thousand structures retrieved from the Cambridge Structural Database for which the structure factor amplitudes, $|\boldsymbol{F}_{\mathbf{H}}|$, were generated at several resolution limits and fed into the trained network to output phases. The phases could be retrieved with a striking accuracy leading to correct structure solutions for over 99% of the validation set entries. Furthermore, the phase accuracy was also high if the resolution limit was chosen to be low, i.e. d_{min} = 1.9 Å. Several dozen of experimentally measured diffraction data sets were also used for validation. Our results hint that deep learning could be used to obtain electron density maps of structures for which only a limited resolution data can be obtained and which are problematic to solve using currently available methods.