

## PhAI: Neural network for structure determination

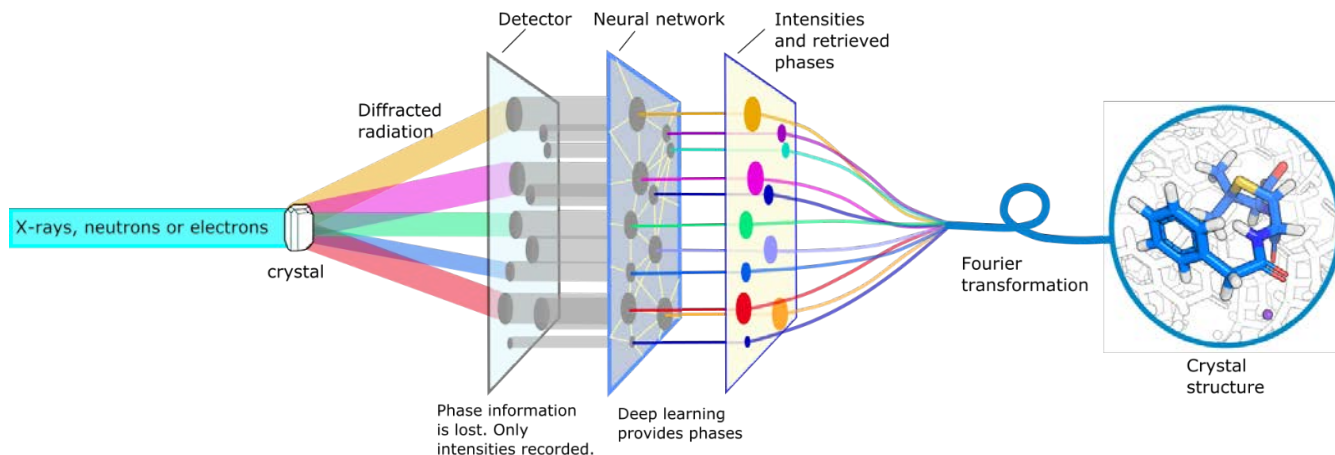
Anders S. Larsen<sup>1</sup>, Toms Rekis<sup>2</sup>, Philipp Hans<sup>1</sup> and Anders Ø. Madsen<sup>1</sup>

<sup>1</sup>Department of Pharmacy, University of Copenhagen, <sup>2</sup>Goethe University Frankfurt.

a.madsen@sund.ku.dk

For a crystal structure to be solved from diffraction data, it is necessary to obtain the complex structure factors,  $F_H$ , of the measured reflections. Experimentally, only their squared moduli,  $|F_H|^2$ , which are proportional to the measured intensities, can be determined, not the phase angles needed to reconstruct the corresponding complex numbers. Since no exact solution to this phase problem exists, several methods to overcome this problem in crystal structure determination have been developed over time. For example, direct methods or the 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, the completeness of the data is low, or the number of atoms in the asymmetric unit is very high - as is the case for macromolecular crystals and some framework structures.

Here we demonstrate how the phase problem can be solved using a neural network, PhAI, which has been trained on millions of fictive structures containing metal atoms and/or molecular fragments.



PhAI was subsequently validated on thousands of structures retrieved from the Cambridge Structural Database for which the structure factor amplitudes,  $|F_H|$ , 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. The phase retrieval accuracy was also high even if the resolution limit was low, i.e.  $d_{\min} = 2 \text{ \AA}$ . Additionally, several dozens of experimentally measured diffraction data sets were used for validation. Our results indicate that deep learning can be used to obtain electron density maps for structures where data could only be obtained at very limited resolution and are more than challenging to solve by state-of-the-art methods.

Our published model (Larsen *et al.*, 2024) focused on structures in the centrosymmetric space group P21/c with small unit cells (cell dimensions less than  $10 \text{ \AA}$ ). In this contribution, we discuss the performance and architecture of an improved neural network that works in the general case (space group P1), i.e. without symmetry restrictions, and on larger unit cells. We also show our progress in implementing the PhAI deep learning for structure solution approach into the Olex2 program.

Larsen, A. S., Rekis, T. & Madsen, A. Ø. (2024). *Science* **385**, 522–528.