INVITED SPEAKER

Artificial intelligence to solve the X-ray crystallography phase problem : a case study report

I. BARBARIN-BOCAHU

Laboratoire de Biologie Structurale De La Cellule (BIOC), CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, F-91128 Palaiseau, France (France)

The determination of three dimensional structure of macromolecules is one of the actual challenge in biology towards the ultimate objective of understanding their function. So far, X-ray crystallography is the most popular method to solve structure, but this technique relies on the generation of diffracting crystals. Once a correct data set has been obtained, the calculation of electron density maps requires to solve the so-called « phase problem ». Different approaches can be used towards this goal. The most frequently used is molecular replacement, which relies on the use of the structure of a protein sharing strong structural similarity with the studied protein. Hence, its success rate is directly correlated with the quality of the models used for the molecular replacement trials. In the absence of models of sufficient quality, it is necessary to use alternative approaches such as multiple isomorphous replacement (MIR) using heavy atoms derivatives, single or multiple wavelength anomalous diffusion (SAD or MAD) from crystals of SeMet-substituted proteins. These latter approaches are not always successful and can be time consuming. Hence, the availability of models as accurate as possible is definitely critical.

Very recently, a breakthrough step has been made in the field of protein structure prediction thanks to the use of machine learning approaches as implemented in the AlphaFold2 or RoseTTAFold. structure prediction programs. I will describe how these recent improvements helped me to solve the crystal structure of a protein involved in the nonsense-mediated mRNA decay pathway (NMD), an mRNA quality control pathway dedicated to the elimination of eukaryotic mRNAs harboring premature stop codons. Indeed, although I obtained crystals of this protein diffracting up to 2.45Å resolution, I struggled for almost 2 years trying to solve its structure. Different approaches were tried unsuccessfully. I have then used both AlphaFold2 and RoseTTAFold programs to generate new structural models for our protein of interest and I will show that these models led to rapid determination of the structure while models generated using few other programs did not.