“AI brings structural insights on protein fold and function evolution”.

Dr Christine Orengo . University College of London

Abstract: Our group is interested in protein evolution and has developed a classification of protein domain structures (CATH) to understand mechanisms by which novel protein functions emerge. Currently the experimental structural data is rather limited (only half a million domain structures in the PDB compared to more than 200 million protein sequences in UniProt). However, novel advanced machine learning tools that exploit AI and Deep Learning have revolutionized the field of structural biology. DeepMind’s AlphaFold algorithm can predict protein structures with quality comparable to experimental data in many cases.  In parallel, other AI based approaches are significantly improving our ability to detect domains within proteins and their evolutionary relationships. We recently combined these approaches with other tools to mine the vast predicted protein structure (214 million protein structures) provided by AlphaFold and we discover novel protein families and vast numbers of new relatives for some CATH superfamilies. We also discovered some beautiful novel protein folding arrangements.  Our strategies can give insights on functions of previously uncharacterized pathogen proteins. We also used AlphaFold models to examine evolutionary relationships in a protein family targeted with antibiotics.