CATH Functional Families (FunFams) – insights into the impacts of genetic variations

Prof. Christine Orengo

UCL, Structural and Molecular Biology, Div. of Biosciences, London, UK

We have used structure comparisons, sequences analyses and manual curation to identify ~5000 protein domain superfamilies. These evolutionary superfamilies have been expanded with sequences from UniProt, predicted to belong to them. Our CATH superfamily resource currently classifies 150 million protein domains, representing ~70% or all domains currently identified in nature. Recent work has resulted in new algorithms sub-classifiying these superfamilies into functional families (FunFams) and this has been found to improve the prediction of structures and functions for uncharacterised proteins. FunFams have been top ranked in blind independent assessments of molecular function prediction (CAFA, 2020).

We have used our FunFams to explore the impacts of splicing on protein function and show that certain types of splicing, homologous mutually exclusive exon (MXE) events, are likely to be having a functional impact. We have also used the FunFams to identify residues likely to be affecting the binding of the SARS-CoV-2 Spike protein to the human ACE2 receptor, an event involved in infection by the virus. This has allowed us to understand whether genetic variations between human and animal ACE receptors will stabilise or destabilise the complex and therefore impact on infection. Our analyses allowed us to explore host range of the virus and identify animals in contact with humans that are likely to be susceptible to infection, and could therefore potentially become reservoirs of the disease.