

Speaker: Christine Orengo, University College London

Date: Wednesday, February 16, 2022

Time: 11:30 AM to 1:00 PM

Host: Bonnie Berger

Title: Protein structure and family data informs functional mechanisms and the risk of disease

Abstract: Classifying proteins into evolutionary families is important for identifying conserved sequence and structure features that are key to the functional mechanisms of these proteins. Our in-house CATH classification currently classifies ~450,000 protein structures and nearly 150 million protein domain sequences into ~5500 evolutionary families. The recent success in protein structure prediction by DeepMind's AlphaFold2 (AF2) method and the release of hundreds of thousands of AF2 models, have changed the scientific landscape by massively extending the structural data available for these protein evolutionary families. We have developed computational strategies to bring this extensive new 3D data into CATH families and are examining how it will expand our understanding of structure – function relationships and our ability to detect functional sites. Functional site predictions can be enhanced by combining structural features and evolutionary conservation patterns and some examples will be given of the application of CATH functional site data to understand protein splice events, the risk of Covid infection and the development of more aggressive lung cancer following whole genome duplication.