Computational methods for targets selection and characterization for structural genomics

Lukasz Kurgan
Computer Science, Virginia Commonwealth University

Abstract
Structural and functional annotation of the rapidly expanding protein universe (defined as a set of all known proteins) is a significant challenge. Perhaps surprisingly we know very little about the protein universe. Structural genomics (SG) is a large-scale international effort that aims to solve structures of important biological macro-molecules, primarily focusing on mapping structures for the protein universe. One of the main bottlenecks in SG is a very low success rate of the production of diffraction quality crystals for the X-ray crystallography, by far the most dominant method for the determination of protein structures. However, the SG pipelines allow for some flexibility in the selection of protein targets. This motivates development of computational methods for the prediction and assessment of the crystallization propensity of proteins, with the underlying goal of finding easier to crystallize and functionally equivalent protein targets. We will overview the currently available computational predictors of crystallization propensity. We will focus on one of our newest methods, fDETECT, which provides relatively strong predictive performance coupled with a short runtime. Utilizing fDETECT and other tools, we will answer the questions of how many protein structures across the protein universe can be determined with the help of the X-ray crystallography and computational modelling, and whether the putative propensity for crystallization can be used to accurately predict resolution of protein crystals.