# Target selection for structural genomics

Lukasz Kurgan
Computer Science, Virginia Commonwealth University

**Abstract**

The rapidly expanding protein universe, defined as a set of all proteins of all organisms, creates a significant challenge to understand and functionally annotate these bigdata. Perhaps surprisingly, we know very little about the protein universe. Structural genomics (SG) is an international effort to solve structures of important biological macro-molecules, focusing on mapping structures of the protein universe. Knowledge of protein structures is necessary to understand their biochemical and cellular functions, to decipher whether and how they interact with other molecules, including drugs. 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, the 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 short runtime. Utilizing fDETECT, we will answer the question whether the structures of protein across the entire protein universe can be determined with the help of the X-ray crystallography and computational modelling?