Structural Genomics and protein universe

Lukasz Kurgan
Computer Science, Virginia Commonwealth University

Abstract
Structural Genomics (SG) is an international effort that aims at solving three-dimensional structures of important biological macro-molecules, focusing primarily on proteins. Knowledge of protein structures is necessary to understand their biochemical and cellular functions, to decipher how they interact with other molecules, and to perform rational drug design. One of the main SG goals is to map structures of the entire protein universe, defined as a set of all proteins of all organisms, and this mission is far from complete.

One of the main bottlenecks in SG is the ability to produce diffraction quality crystals for the X-ray crystallography, the main method for the determination of protein structures. SG pipelines allow for certain flexibility in target selection which motivates development of computational methods for the prediction/assessment of the protein crystallization propensity.

We will overview the currently available predictors of crystallization propensity, focusing on our newest method, fDETECT, that provides good predictive performance coupled with short runtime. Utilizing fDETECT, we answer the question whether and how the three-dimensional structures of all protein families can be determined with the help of the X-ray crystallography? This is based on a first-of-its-kind analysis of crystallization propensity for a current snapshot of the protein universe consisting of all proteins encoded in 1,953 fully sequenced genomes across eukaryotes, bacteria, archaea, and viruses. We will wrap up with a selection of take-home messages.