"Characterization of crystallization propensity of protein chains for Structural Genomics"

Structural Genomics (SG) is an international effort that aims at solving three-dimensional shapes of important biological macro-molecules with primary focus on proteins. One of the main bottlenecks in SG is the ability to produce diffraction quality crystals for the X-ray crystallography based protein structure determination. SG pipelines allow for certain flexibility in target selection which motivates development of in-silico methods for sequence-based prediction/assessment of the protein crystallization propensity. We introduce, contrast and empirically compare modern sequence-based predictors of crystallization propensity including SECRET, OB-Score, CRYSTALP, ParCrys, XtalPred, and CRYSTALP2. We show that although their success rate is only at about 70%, they provide useful and complimentary predictions and some of them are already utilized within the SG pipelines. We also discuss our new ensemble-based predictor, MetaPPCP, which allows for predictions at over 80% success rate. Our findings support the claim that the crystallization propensity can be accurately predicted directly from the protein chain