

Bioinformatics Approach to Identify Physicochemical and Structural Properties Associated with Successful Cell-free Protein Synthesis

Authors : Alexander A. Tokmakov

Abstract : Cell-free protein synthesis is widely used to synthesize recombinant proteins. It allows genome-scale expression of various polypeptides under strictly controlled uniform conditions. However, only a minor fraction of all proteins can be successfully expressed in the systems of protein synthesis that are currently used. The factors determining expression success are poorly understood. At present, the vast volume of data is accumulated in cell-free expression databases. It makes possible comprehensive bioinformatics analysis and identification of multiple features associated with successful cell-free expression. Here, we describe an approach aimed at identification of multiple physicochemical and structural properties of amino acid sequences associated with protein solubility and aggregation and highlight major correlations obtained using this approach. The developed method includes: categorical assessment of the protein expression data, calculation and prediction of multiple properties of expressed amino acid sequences, correlation of the individual properties with the expression scores, and evaluation of statistical significance of the observed correlations. Using this approach, we revealed a number of statistically significant correlations between calculated and predicted features of protein sequences and their amenability to cell-free expression. It was found that some of the features, such as protein pI, hydrophobicity, presence of signal sequences, etc., are mostly related to protein solubility, whereas the others, such as protein length, number of disulfide bonds, content of secondary structure, etc., affect mainly the expression propensity. We also demonstrated that amenability of polypeptide sequences to cell-free expression correlates with the presence of multiple sites of post-translational modifications. The correlations revealed in this study provide a plethora of important insights into protein folding and rationalization of protein production. The developed bioinformatics approach can be of practical use for predicting expression success and optimizing cell-free protein synthesis.

Keywords : bioinformatics analysis, cell-free protein synthesis, expression success, optimization, recombinant proteins

Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development

Conference Location : Chicago, United States

Conference Dates : December 12-13, 2020