West-Life

Milestone M24

Work Package 7, Milestone 1: Ranking in Protein CCD

This milestone concerns the provision of specifications for new ranking services in ProteinCCD. Specifically we wish to add solubility and crystallizability scoring and ranking algorithms from available servers, to analyze the user-designed constructs. This design leads to deliverable 7.3 in month 24.

Work Package 7, Task 1 Joint Research

This Joint Research Activity is aimed at exploring new ways to use existing or close to existing services so that broader user communities will be reached.

Background

The Protein Crystallographic Construct Design [1] ProteinCCD (https://xtal.nki.nl/ccd/) software aims to increase the efficiency in the design of several truncation constructs of a protein under investigation – for more details see the report in Deliverable 7.2.

State of the art

The functionality of ProteinCCD is now extended: we implemented it in a new computational platform allowing a more interactive and efficient interface to the user, and we are providing new analysis options, including: parallel processing of server requests, more efficient interface for construct design, more cloning methods, an extended collection of existing vectors, local execution of some algorithms for improving response time, new servers for meta-analysis, easy bookkeeping, and better data security.

Milestone objective

To aid the users in making optimal decision within the framework of ProteinCCD, we wish to add solubility and crystallizability scoring and ranking algorithms from available servers

Specifications

We have split the analysis and ranking in three tiers.

Tier 1: Use algorithms from the BioPython library to calculate the molecular weight, isoelectric point, and absorption coefficience for all designed constructs.

Tier 2: For providing solubility information we will evaluate and choose from the following severs:

Milestone M24

- http://biotech.ou.edu This model was created using logistic regression of 32 possible parameters. In addition, the protein database used to create this model was increased to 212 proteins. Results from the model were 94% accurate when compared to lab results. The server is available at the University of Oklahoma, School of Chemical, Biological, and Materials Engineering, Recombinant Protein Solubility Prediction
- http://mips.helmholtz-muenchen.de/proso/proso.seam PROSO and PROSO II try to answer the following question: "*Which of my cloned proteins have the best/worst chances to be soluble upon heterologous expression?*" The prediction is based on a classifier exploiting subtle differences between soluble proteins from TargetDB and PDB and notoriously insoluble proteins from TargetDB and literature mining.

Tier 3: For providing crystallisability information we will evaluate and choose from the following servers:

- http://mips.helmholtz-muenchen.de/secret/secret.seam SECRET tries to answer the questions: "*What is the chance that my soluble protein will crystallize?*" and"*Which of my soluble proteins have the best/worst chances to crystallize?*" The prediction is based on a classifier trained with the dataset SMALL.
- http://biomine-ws.ece.ualberta.ca/CRYSTALP2.html CRYSTALP2 is a a kernel-based method that predicts the propensity of a given protein sequence to produce diffraction-quality crystals. This method utilizes the composition and collocation of amino acids, isoelectric point, and hydrophobicity, as estimated from the primary sequence, to generate predictions.

Technical considerations

The framework designed and implemented for Deliverable 7.2 will be used for all extensions.