BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	Structural Bioinformatics
Title	SOLart: a structure-based method to predict protein solubility and aggregation
All Authors	Qingzhen Hou (1), Jean Marc Kwasigroch (1), Marianne Rooman (1), Fabrizio Pucci (1,2)

Affiliation

(1) Computational Biology and Bioinformatics, Université Libre de Bruxelles, 1050 Brussels, Belgium;
(2) John von Neumann Institute for Computing, Julich Supercomputer Centre, Forschungszentrum Julich, 52428 Julich, Germany

Motivation

The solubility of a protein is a fundamental property that is often decisive for its proper function. Lack of solubility is frequently a major bottleneck in high-throughput structural genomic studies and in industrial applications requiring high-concentration production of recombinant proteins. Moreover, the formation of protein aggregates such as amyloid fibrils constitutes a pathological characteristic of a wide variety of diseases.
Since the experimental characterization of protein solubility is time-consuming and expensive, there is a significant need for computational approaches that predict protein solubility properties and design more soluble variants.

Methods

We have recently introduced solubility-dependent statistical potentials that are able to unravel the role of amino acid interactions in promoting or decreasing protein solubility. Extending their construction, we developed a series of new solubility-dependent potentials that depend not only on interresidue distances but also on backbone torsion angles and solvent accessibility. We integrated them, together with other structure- and sequence-based features, into a random forest model that we trained on a set of E.coli proteins with experimentally determined structure and solubility values. We obtained in this way the SOLart protein solubility predictor, whose most informative features turned out to be differences in folding free energy computed from statistical potentials derived from either soluble or aggregation-prone proteins.

Results

SOLart's performances are good: it has a Pearson correlation coefficient between experimental and predicted solubility values of 0.7 both in cross-validation on the learning dataset and on an independent dataset of S. Cerevisiae proteins. It outperforms state-of-art solubility predictors, which are generally based on sequence alone. SOLart can be used with high-resolution, but also with low-resolution structures. Indeed, we applied it to a set of modeled structures and observed only a limited drop in performances. SOLart is available through a user-friendly webinterface, which is easy to use by non-experts scientist who are interested in the analysis and modulation of protein solubility and aggregation. The SOLart webserver is freely available at: http://babylone.ulb.ac.be/SOLART/ .

Info

-

Figure

-

Availability	https://doi.org/10.1101/600734

Corresponding Author

Name, Surname	Qingzhen, Hou
Email	qinghou@ulb.ac.be
Submitted on	26.04.2019