mCSM-PPI v2: predicting the effects of mutations in protein-protein binding affinity from sequence and structural features

Willy Garabini Cornelissen[1], David B. Ascher[2], Douglas E.V. Pires[1],

1 Instituto René Rachou, Fundação Oswaldo Cruz
2 Department of Biochemistry and Molecular Biology, University of Melbourne, Melbourne

Abstract

Recent studies have been showing that a large proportion of disease mutations, including inherited genetic diseases as well as cancer-related mutations, occur at protein-protein interfaces (PPIs). In fact, edgatic mutations (those affecting interactions) seem to be overrepresented in diseases. Considering their importance in biology and for public health, the ability of predicting the effects of missense mutations on PPIs from protein sequence and structural evidences has become an important step towards understanding their role in diseases as well as an important tool for protein engineering. Different computational approaches have been proposed in recent years for this purpose, although much less attention has been devoted to mutations in protein-protein interactions and how they interfere with the binding affinity of the complex.

Here we present mCSM-PPI v2, an optimised machine learning-based method that relies on sequence and structural information to quantitatively predict the impact of single-point missense mutations in the affinity of protein-protein complexes. The improved method encodes distance patterns between atoms in a feature vector aiming to capture both the geometry and physicochemical properties of protein residue environments as well as distance patterns between residues on a sequence level. These signatures are then used as evidence for algorithms to learn from mutations with available experimental thermodynamic data.

mCSM-PPI v2 has been retrained, tested and optimised on low redundancy data sets derived from the SKEMPI database. Our method obtained a Pearson's correlation coefficient of up to 0.80, considerably more accurate than alternative approaches (BeAtMuSiC = 0.47; Li et al. = 0.58; Moal et al. = 0.73;), showing that using sequence and structural information in combination was advantageous. mCSM-PPI v2 was also able to identify hot-spots on PPIs via computational alanine scanning as well as optimising peptide affinity via in silico saturation mutagenesis.

We believe mCSM-PPI v2 will be scalable, robust, quantitative approach for analysing large data sets of mutations in PPIs that may aid in optimizing protein-protein binding and modulation as well as understanding deleterious mutations and their relation with diseases.

Funding: Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte