Latent Structures-based Modeling of Mutated Protein-Protein Interaction Networks

Abel Folch-Fortuny\textsuperscript{1}, Gabriel Bosque\textsuperscript{2}, Jesús Picó\textsuperscript{2}, Alberto Ferrer\textsuperscript{1}, and Santiago F. Elena\textsuperscript{3,4}

\textsuperscript{1} Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Valencia, Spain
\texttt{abfolfor@upv.es}
\textsuperscript{2} Institut Universitari d’Automàtica i Informàtica Industrial, Universitat Politècnica de València, Valencia, Spain
\texttt{gabbosch@upv.es}
\textsuperscript{3} Instituto de Biología Molecular y Celular de las Plantas, CSIC-Universitat Politècnica de València, Valencia, Spain
\textsuperscript{4} The Santa Fe Institute, Santa Fe, USA

**Keywords:** Quantitative Structure-Activity Relationship, mutation, Protein-Protein Interaction Network, fitness, Partial Least Squares regression, potyvirus

A new methodology is presented here with the aim of relating mutations, Protein-Protein Interaction Networks (PPINs) and organismal fitness, following the Quantitative Structure Activity Relationship (QSAR) philosophy. A latent structures-based method, called Partial Least Squares regression (PLS), is applied to identify few functional modules of the PPIN and relate them to significant changes in the fitness. The Tobacco etch potyvirus (TEV) is used as case study.

*Potyvirus* is the largest genus of *Potyviridae* family, containing more than 180 different members. The PPIN of *potyvirus*, representing the map of physical contacts between its proteins, is built as a first step of the study, based on a literature review \cite{1}. The integration of different experimental studies yields a PPIN with 11 proteins and 24 interactions. Additionally, data from mutations performed on TEV are used in combination with the PPIN \cite{3}. These data consists of a set of single and double mutations and the organismal fitness generated. In this way, it is possible to discern whether the mutation changes the fitness, compared to the wild-type, or not.

The mathematical modeling of the three sources of data is described graphically in Figure 1. Each mutation is performed at a gene level, and affects a single protein in the PPIN. The change in this protein affects its interactions with the neighbor proteins in the PPIN. Finally, the mutation makes the organismal fitness increase or decrease.

QSAR modeling has been applied in the literature to relate chemical, physical and topological features of molecules to its biological activity \cite{2}. Latent structure-based methods, such as PLS regression \cite{4}, are widely used in QSAR
modeling. In this study, the predictors are the mutations performed on TEV and the changes in the interactions of the PPIN, the predicted variable is the organismal fitness, and the latent constructs, based on the existing correlations among variables, represent functional modules (clusters) in the PPIN.

The results show that there exist four functional modules on the PPIN. One is activated via mutation PC6 on protein P1, affects proteins VPg and CI, and makes the fitness increase (see Figure 1). There are two other modules which make the fitness increase: 6K2-NLaPro and NLaPro-CP-Nlb. The module containing HCPro and NLaPro makes the fitness decrease. However, this is produced when mutation PC22 is applied on HCPro, but when PC19 is applied on the same protein the fitness increases, producing no effect on the PPIN.

The data fusion presented here allows to detect and model the influence between mutations, PPIN and fitness, and also the direct effect of mutations on the fitness, without affecting the PPIN. This deeper knowledge of potyvirus and its PPIN permits to wisely perform mutations in order to obtain stronger/weaker mutants, depending on the aim of the experiment.

Fig. 1. Example of mutations-PPIN-fitness modeling. PC6 mutation affects P1 and its adjacent interactions with VPg and CI. Finally, the organismal fitness increases.

References