Correlated Sequence-Signatures as Markers of Protein-Protein Interaction.

Sprinzak, E. and Margalit, H.
Department of Molecular Genetics and Biotechnology, Hadassah Medical School, Hebrew University of Jerusalem

As protein-protein interaction is intrinsic to most cellular processes, the ability to predict which proteins in the cell interact can aid significantly in identifying the function of newly discovered proteins, and in understanding the molecular networks they participate in. An appealing approach would be to predict the interacting partners by characteristic sequence motifs that typify the proteins that are involved in the interaction. Valuable insight towards this end can be gained by mining databases of experimentally determined interacting proteins. Conventionally, single protein sequences have been clustered into families by distinct sequence signatures. Here we propose a novel approach for clustering different *pairs* of interacting proteins by *combinations* of their sequence signatures. To identify such informative signature combinations, a database of interacting proteins is required, as well as a scheme for characterizing protein sequences by their signatures. In the current study we demonstrate the potential of this approach on a comprehensive database of experimentally determined pairs of interacting proteins in the yeast S. *cerevisiae*. The proteins are characterized by sequence signatures, as defined by the InterPro classification. A statistical analysis is performed on all possible combinations of two sequence signatures, identifying combinations of sequence signatures that are over-represented in the database of pairs of interacting proteins. It is proposed that such correlated sequence signatures can be used as markers for predicting unknown protein-protein interactions in the cell. Such an approach reduces significantly the search in the interaction space, and enables directed experimental interaction screens.

References:

Sprinzak E, Margalit H. (2001), J Mol Biol. 311(4):681-92.