Dependency and Co-dependency: Predicting amino acid interactions within and between proteins using co-variation

Elisabeth R. M. Tillier,
University Health Network and Department of Medical Biophysics,
University of Toronto

This talk will describe how the detection of co-variation of amino acids sites in a multiple sequence alignment can be used to measure phylogenetic signal. This is useful for the prediction of functional interactions between amino acids and proteins.

Dependency:

Motivation:

Multiple sequence alignments of homologous proteins are useful for inferring their phylogenetic history and to reveal functionally important regions in the proteins. Functional constraints may lead to co-variation of two or more amino acids in the sequence, such that a substitution at one site is accompanied by compensatory substitutions at another site. It is not sufficient to find the statistical correlations between sites in the alignment because these may be the result of several undetermined causes. In particular, phylogenetic clustering will lead to many strong correlations.

Result:

A procedure is developed to detect statistical correlations stemming from functional interaction by removing the strong phylogenetic signal that leads to the correlations of each site with many others in the sequence.

Availability:

The program and supplementary figures tables are available from the site

`http://www.uhnres.utoronto.ca/tillier/Depend/dependence.html`

Codependency:

Motivation:

Approaches for the determination of interacting partners from different protein families (such as ligands and their receptors) have made use of the property that interacting proteins follow similar patterns and relative rates of evolution. Interacting protein partners can then be predicted from the similarity of their phylogenetic trees or evolutionary distances matrices.

Results:

Here we present a novel method for the determination of interacting partners by maximizing co-evolutionary signals. The order of sequences of the multiple sequence alignments from two protein families is determined in such a manner as to maximize the similarity of substitution patterns at amino acid sites in the two alignments and thus phylogenetic congruency. This is achieved by maximizing the total number of interdependencies of amino acids sites between the alignments. Once thus ordered, the corresponding sequences in the two alignments then indicate the predicted interacting partners. We demonstrate the efficacy of the approach with computer simulations and in analyses of several protein families.

Availability:

A program implementing our method, Codep will be freely available to academic users from our website:

`http://www.uhnresearch.ca/labs/tillier/`

after publication.