Disulfide Connectivity Prediction using Generalized Recursive Neural Networks and Evolutionary Information - (session: Structural Genomics)

Alessandro Vullo and Paolo Frasconi

Dipartimento di Sistemi e Informatica, Università di Firenze

We focus on the prediction of disulfide bridges in proteins starting from their amino acid sequence and from the kowledge of the disulfide bonding state of each cysteine. The location of disulfide bridges is a structural feature that conveys important information about the protein main chain conformation and it can therefore help towards the solution of the folding problem. Existing approaches based on weighted graph matching algorithms do not take advantage of evolutionary information. Recursive neural networks (RNN), on the other hand, can handle in a natural way complex data structures such as graphs whose vertices are labeled by real vectors, allowing us to incorporate multiple alignment profiles in the graphical representation of disulfide connectivity patterns.
The core of the method is the use of machine learning tools to rank alternative disulfide connectivity patterns. We develop an ad-hoc RNN architecture for scoring labeled undirected graphs that represent connectivity patterns. We report experimental results on a set of cysteine-rich non-homologous sequences. We found that using multiple alignment profiles allows us to obtain a significant improvement of prediction accuracy, clearly demonstrating the important role played by evolutionary information.