An information-theoretic classification of amino acids for the optimization of interfaces descriptions in protein-protein docking

Christophe Jardin[1], Arno Stefani[2], Martin Eberhardt[1], Johannes Huber[2], Heinrich Sticht[1]

[1]Institute for Biochemistry and [2]Institute for Information Transmission, Friedrich-Alexander University Erlangen-Nuremberg, Germany

One major challenge in protein-protein docking is the definition of suitable criteria for a scoring function that allows the identification of a good docking solution among many false arrangements.

In a previous work we demonstrated that the concept of mutual information (MI) from information theory can be used to investigate structural features of protein-protein interfaces and to assess their information content in protein docking. These MI-values can be converted into a scoring function [1]. However, this first "proof-of-concept" also revealed aspects that had to be improved to result in a robust and widely applicable approach.

Originally, structural features were derived after the grouping of the 20 aminoacids into a four-letter alphabet [1]. The use of such reduced alphabets has the advantage of an improved statistics; however, a four-letter alphabet might be too small to adequately represent the properties of individual amino acids.

Therefore, the main goal pursued in the present work was the generation of more sophisticated aminoacid alphabets (or structural descriptors) that allow a more accurate description of the biological features that govern protein-protein recognition.

Optimized aminoacid alphabets were generated starting from known ones using a clustering approach and a MI-based assessment of their information content. The alphabets resulting from this approach were tested for their performance in scoring docking solutions in a five-fold cross-validation and demonstrated a significantly improved performance. Furthermore, a web service for scoring protein-protein docking solutions based on our approach has been developed and will be soon made available to the scientific community.

[1] Othersen et al. *J. Mol. Model.*, **2012**, 18(4): 1285-1297.