Towards the in silico reconstruction of protein interaction networks: identification of different binding sites in the interaction of proteins with other proteins, small molecules, DNA and RNA

Résumé du projet de recherche (Langue 1)

Protein interactions are essential to all biological processes and they represent increasingly important therapeutic targets. We recently developed a new method for accurately predicting protein-protein interfaces, understanding their properties, origins and binding to multiple partners (Laine & Carbone, PLoS Comput. Biol. 2015, under last revision). Contrary to machine learning approaches, our method combines in a rational and very straightforward way three sequence- and structure-based descriptors of protein residues: evolutionary conservation, physico-chemical properties and local geometry. The implemented strategy yields very precise predictions for a wide range of protein-protein interfaces and discriminates them from small-molecule binding sites. Beyond its predictive power, the approach permits to dissect interaction surfaces and unravel their complexity. We show how the analysis of the predicted patches can foster new strategies for PPIs modulation and interaction surface redesign. The approach is implemented in JET2, an automated tool for sequence-based protein interface prediction that is based on the Joint Evolutionary Trees (JET) method (Engelen … Carbone, PLoS Comput. Biol. 2009). JET2 will be soon freely available at www.lcqb.upmc.fr/JET2. In parallel, we have developed new strategies for predicting protein interfaces involved in protein-DNA interactions (master thesis of Flavia Corsi, 2015) and obtained very encouraging results on this type of protein interactions too. These interaction surfaces are expected to satisfy characteristics different from protein-protein interfaces that had to be unravelled for prediction. During the master thesis, we analyzed the evolutionary conservation, physico-chemical and geometrical properties of protein-DNA interfaces. We observed that not only the physico-chemical properties (already known result) but also the geometrical patterns holding for protein-protein interactions are not anymore true for DNA-protein interactions. Then, by approaching the question as in JET2, we defined a few new rational heuristics leading to accurate protein-DNA interface identifications (not yet published). Encouraged by these two successful results, in this thesis, we propose to develop further the analysis started with the master thesis of Flavia Corsi and converge to an optimal model of protein-DNA interaction. A number of strategies to be investigated are already well determined and their impact on our understanding of transcriptional regulation is expected. For this task, we shall be helped by R. Lavery, who is an expert in the molecular modeling of protein-DNA interactions. But with this thesis we wish to go further and investigate, in a first phase, also: i. RNA-protein interaction and ii. better characterize binding sites involved in small molecules-protein interactions, already partially addressed when analyzing protein-protein interactions (Laine & Carbone, 2015). This will be done by exploiting the three descriptors that demonstrated being successful, as described above. In a second phase, we wish to include a more complex component into the prediction. Namely, we want to check the hypothesis that evolutionary signals combined to local geometry and physico-chemical properties could guide the identification of hidden (cryptic?) protein interaction sites. We will integrate a new level of complexity in our analysis, which is the dynamical behaviour of the proteins under study. We wish to test whether signals extracted from protein sequence analysis coupled with the exploration of protein conformational dynamics can predict alternative functional states of the protein and reveal potential sites displayed in these alternative states. We also intend to characterize the internal dynamics of these sites, how they appear, change or are disrupted by conformational change, to help guide their targeting by small molecule compounds. This step will require the development of a sophisticated algorithm modeling the possibilities of conformational changes and the exploration of the sites. This algorithm should provide a fast scanning of the conformations and a fast reconstruction of potential binding sites. The aim being to realize these analyses at large scale and, possibly, on a runtime mode.

Résumé du projet de recherche (Langue 2)
The main purpose of the thesis is the identification of the different origins of interaction for proteins in complexes. This is an important question in computational biology that will lead, one day, to the in silico reconstruction of protein interaction networks, involving protein partners such as other proteins but also DNA, RNA and small molecules. The understanding of protein interaction sites, their overlapping and their origin, will provide fundamental knowledge on the criteria to follow to model protein interactions in a crowded environment of a virtual cell. The presence of protein conformational changes during interactions is more and more supported by experimental evidence, in particular for protein-DNA interactions (where the protein generally has to wrap itself around DNA to stabilize the interaction), while computational approaches have not been developed at large scale yet. In particular, one would like to simulate and evaluate runtime potential interactions. We believe that we start having sufficient knowledge of protein interactions to approach this question. Notice that an answer to this question is fundamental in view of the identification of partners in the cell. We have already demonstrated (Lopes…..Carbone, PLoS Comput. Biol., 2013) that knowing about accurate protein interfaces highly helps to discriminate which proteins are partners of a given one. The characterization of the dynamical evolution of binding sites during the lifetime of a protein, in addition to their identification, also has the potential to open new avenues for drug discovery and development.

Informations complémentaires (Langue 1)

The question is fundamental in computational biology and many international laboratories and researches in biology will be directly interested in our finding..

Informations complémentaires (Langue 2)

This thesis has been funded by the LABEX CALSIMLAB.