Laboratoire de Biochimie Théorique
Institut de **B**iologie **P**hysico-**C**himique
13, rue Pierre et Marie Curie
75005 PARIS

S E M I N A I R E

Ranjit P. Bahadur
Computational Structural Biology Lab. Department of Biotechnology, IIT Kharagpur, India

« Structural Analysis of Protein-RNA complexes: what we learn from the Protein Data Bank »

We have developed protein-RNA docking benchmark, which consists of bound and unbound structures of partner molecules. Besides, we have also curated a binding affinity dataset of 40 protein-RNA complexes, for which at least one unbound partner is available in the docking benchmark. Regression models were trained to predict the binding affinity from structural and physicochemical parameters of the protein-RNA interfaces. The best-fit model with the lowest maximum error is provided with three interface parameters: relative hydrophobicity, conformational change upon binding and relative hydration pattern. This model has been used for predicting the binding affinity on a test dataset, generated using mutated structures of yeast aspartyl-tRNA synthetase, for which experimentally determined ΔG values of 40 mutations are available. The predicted ΔG values highly correlate with the experimental observations. Further, we use evolutionary conservation derived from structure alignment of polypeptide sequences along with structural and physicochemical attributes of protein–RNA interfaces to probe the binding hot spots at protein–RNA recognition sites. We find that the degree of conservation varies across the RNA binding proteins; some evolve rapidly compared to others. Additionally, irrespective of the structural class of the complexes, residues at the RNA binding sites are evolutionary better conserved than those at the solvent exposed surfaces. For recognitions involving duplex RNA, residues interacting with the major groove are better conserved than those interacting with the minor groove. We identify multi-interface residues participating simultaneously in protein–protein and protein–RNA interfaces in complexes where more than one polypeptide is involved in RNA recognition, and show that they are better conserved compared to any other RNA binding residues. Finally, we develop a Random Forests model using structural and physicochemical attributes for predicting binding hot spots. The model accurately predicts 80% of the instances of experimental ΔΔG values in a particular class. The dataset provided in this study should be useful for the further development of the docking algorithms and binding affinity prediction methods. Moreover, the model developed in this study enhances our understanding on the structural basis of protein-RNA binding affinity, and provides a platform to engineer protein-RNA interfaces with desired affinity.

References:

1. A structure-based model for the accurate prediction ofprotein-RNA binding affinity. Nithin C, Mukherjee S & Bahadur RP. Nucleic Acids Research (in major revision).
2. A non-redundant protein-RNA docking benchmark version 2.0. Nithin C, Mukherjee S & Bahadur RP. Proteins 2017; 85(2):256–267.
3. Probing binding hot spots at protein-RNA recognition sites. Barik A, Nithin C, Karampudi NBR, Mukherjee S & Bahadur RP. Nucleic Acid Res.2016; 44(2):e9.

Jeudi 1er février 2018
14h30
Salle de conférence