Poster I-53

A computational method to identify amino acid residues involved in protein-RNA interactions

Authors:
Michael Terribilini (*Iowa State Univeristy*)
Jeffry Sander (*Iowa State Univeristy*)
Byron Olson (*Iowa State Univeristy*)
Jae-Hyung Lee (*Iowa State Univeristy*)
Robert Jernigan (*Iowa State Univeristy*)
Vasant Honavar (*Iowa State Univeristy*)
Drena Dobbs (*Iowa State Univeristy*)

Short Abstract: Protein-RNA interactions are vital for many biological processes. This work is the first reported attempt to combine sequence and three-dimensional structure information for prediction of RNA-binding residues in proteins. It should be valuable both for investigations of specific RNA-binding proteins and for large scale efforts.

Long Abstract:
Protein-RNA interactions are vitally important in a wide range of biological processes, including regulation of gene expression, protein synthesis, and replication and assembly of many viruses. The ability to reliably predict which residues of a protein directly contribute to RNA binding would significantly enhance our understanding of how proteins recognize RNA and potentially generate new strategies for clinical intervention in both genetic and infectious diseases.

We have developed a machine learning approach for predicting which amino acids of an RNA-binding protein are involved in protein-RNA interactions, using either amino acid sequence only or a combination of both sequence and structure-derived information as input. Interfaces from known protein-RNA complexes in the PDB were extracted to generate a non-redundant set of 147 RNA-binding protein chains. Using this dataset, several types of classifiers were trained to predict which residues in a given RNA binding protein are located at the protein-RNA interface. The first classifier uses only the amino acid sequence as input, whereas the second classifier uses structural information as input. We then combine the results of these two individual classifiers to obtain the final prediction. The combined classifier identifies interface residues with 87% overall accuracy, correlation coefficient of 0.37, specificity for interface residues of 57% and sensitivity for interface residues of 33%. The classifier can be calibrated to increase the specificity or sensitivity of interface residue prediction for specific functional classes of RNA-binding proteins.

To the best of our knowledge, this is the first reported attempt to exploit both amino acid sequence and three-dimensional structure for prediction of RNA-binding residues in proteins. We have applied the sequence-based classifier to several clinically important proteins for which experimental structure information has been lacking, e.g. HIV-1 Rev and the human telomerase reverse transcriptase (hTERT) protein. In both cases, the predicted RNA-binding residues agree with the biochemically mapped RNA-binding regions.