

A Computational Method to Identify RNA Binding Sites in Proteins

Presented by

Jeff Sander
Iowa State University
Rocky 2006

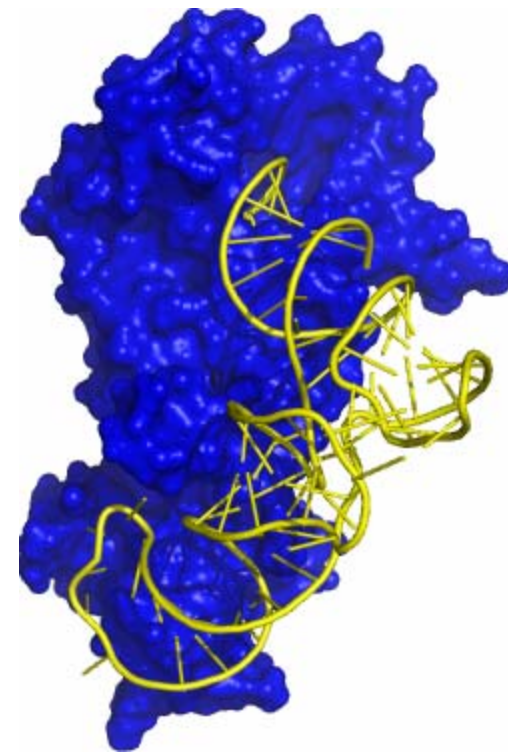
Biological Motivation

- Protein / RNA complexes are
Which amino acids are
directly responsible for
binding RNA?
 - mRNA, tRNA, rRNA
 - miRNA, siRNA, RNAi
 - RNA viruses

Terribilini et al (2006) RNA

Wang & Brown (2006) NAR

Jeong & Miyano (2006) Tras Sys Biol



AMINOACYL TRANSFER RNA SYNTHETASE

Sequence Based Predictions

- **Dataset**
 - Protein RNA complexes from the Protein Data Bank
 - Less than 30% identity and 3.5Å or better resolution
 - 147 Protein RNA complexes
 - 14% interacting residues / 86% non-interacting
- **Classifier** - Naïve Bayes

A R V H N T R Q Q G A T L A F L T L R Q Q A S L I Q



- **Results:**
CC: 0.33 Acc: 0.86 Sp+: 0.46 Sen+: 0.36

- **Server: RNABindR**
<http://bindr.gdcb.iastate.edu/RNABindR/>

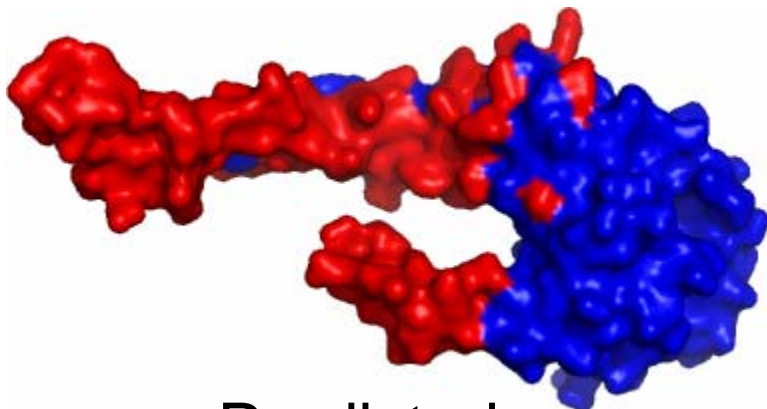
Terribilini et al (2006) RNA

Using Additional Information

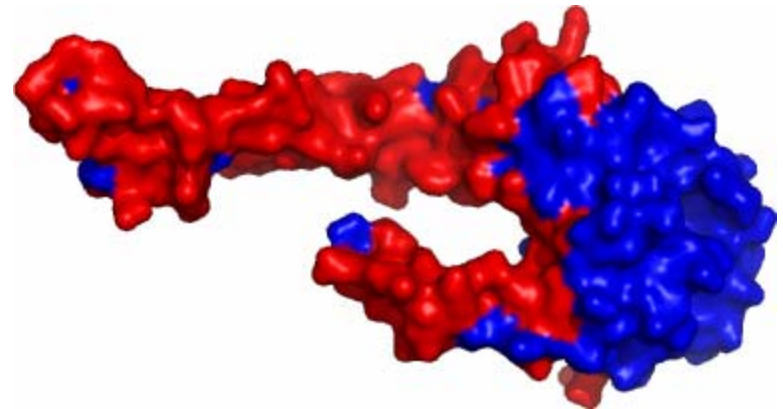
- PSSM **A R V H N T R Q Q G A T L A F**
- PSI-Blast **A R V H N T R Q Q G A T L A F**
- Str Neighbor **A R V H N T R Q Q G A T L A F**

- Combination **A R V H N T R Q Q G A T L A F**

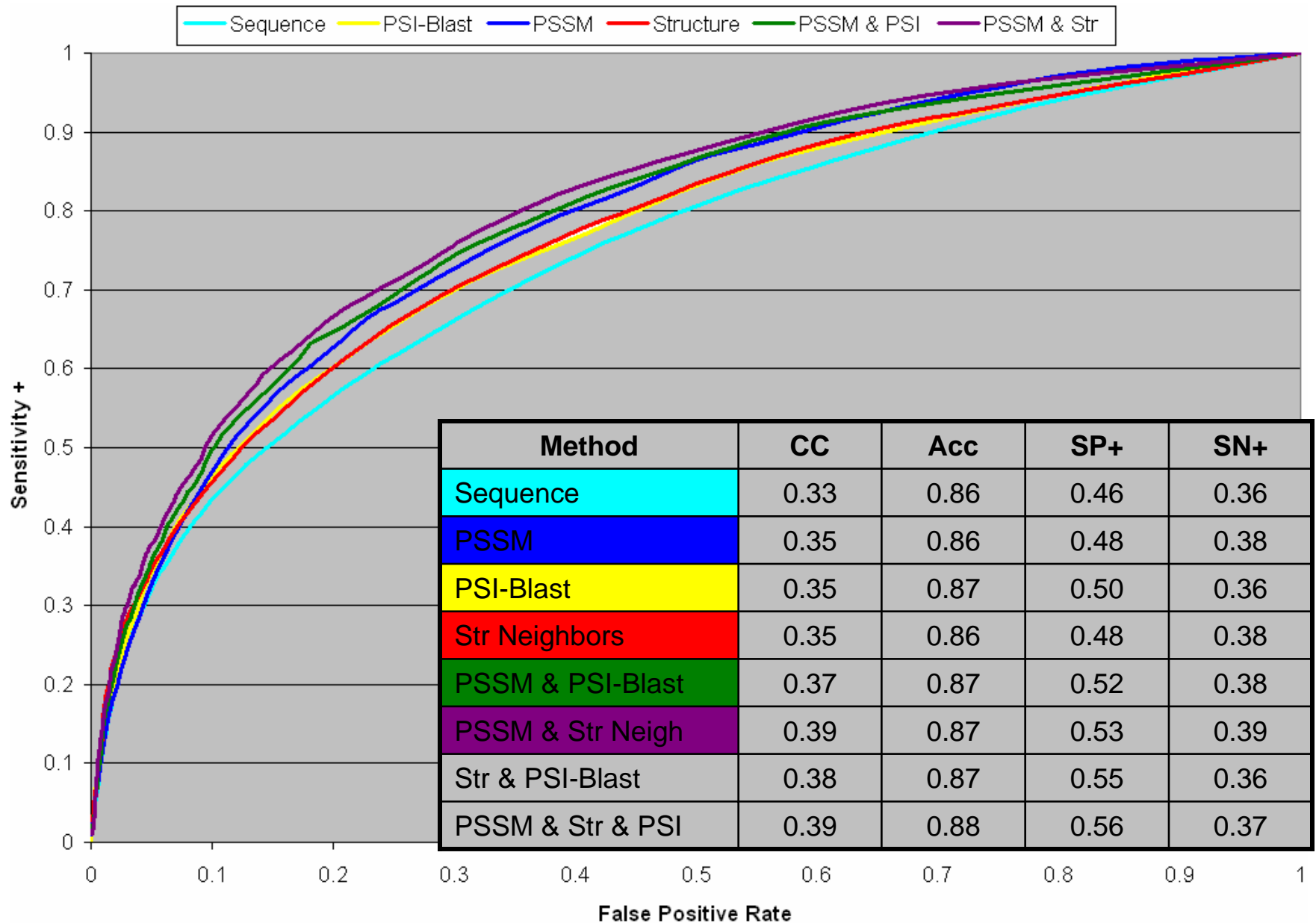
- Actual **A R V H N T R Q Q G A T L A F**



Predicted



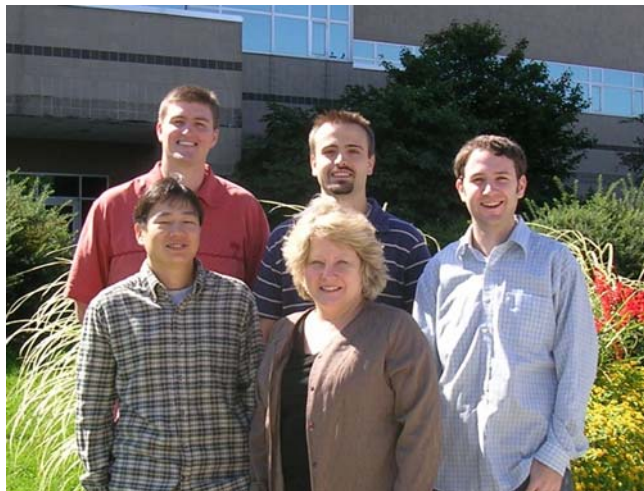
Actual



Conclusions

- Evolutionary and structural information can enhance prediction over basic sequence
- Combining classifiers can provide enhanced predictions over individual classifiers

Acknowledgements



- Michael Terribilini
- Changhui Yan
- Jae-Hyung Lee
- Drena Dobbs
- Vasant Honavar
- Robert Jernigan

Funding: USDA MGET, NIH, CIAG (ISU)