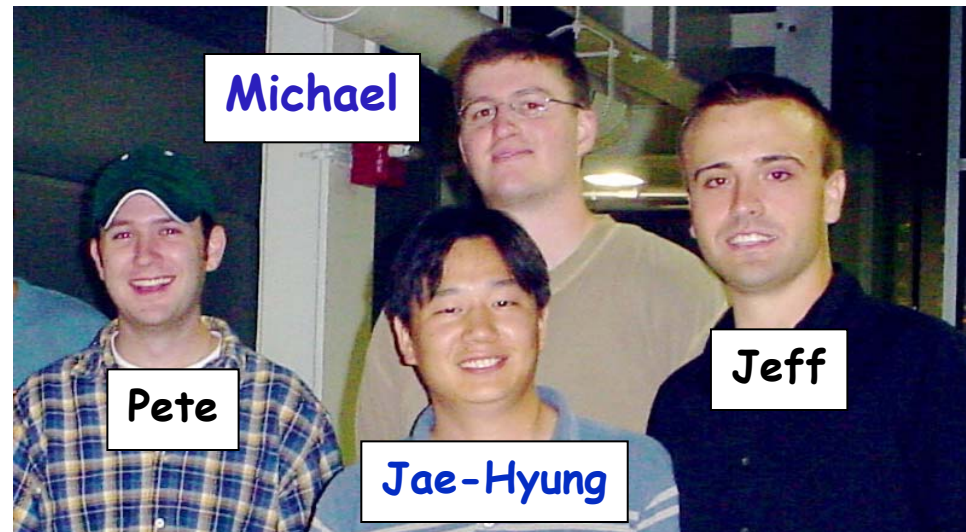


## *Computational Identification of Binding Sites in Proteins*

**Drena Dobbs, BCB at ISU**

with my graduate students:

**M Terribilini, J-H Lee,  
J Sander, P Zaback**



& our collaborators:

**C Yan, F Wu & V Honavar**

**W Sparks & S Carpenter**

**Y Ihm, C Wang & K-M Ho**

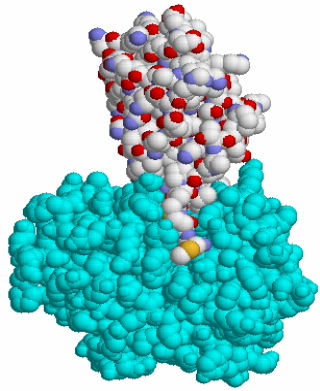
**T Sen & R Jernigan**



**Changhui**

QuickTime™ and a  
TIFF (LZW) decompressor  
are needed to see this picture.

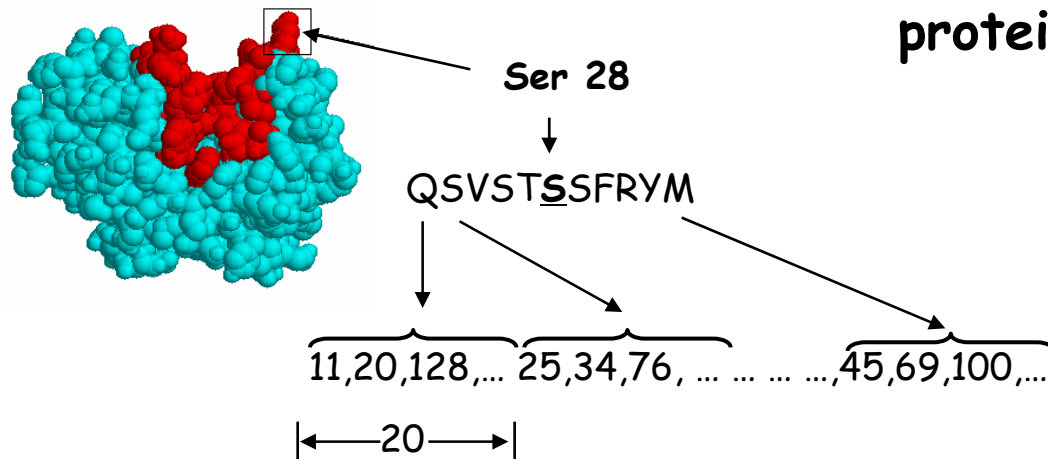
**Vasant**



**PROBLEM:** Given the sequence of a protein (& possibly its structure), predict which amino acids participate in protein-protein or protein-nucleic acid interactions

**APPROACH:** Generate datasets of known complexes from PDB to train & test machine learning algorithms (Naïve Bayes, SVM, etc.)

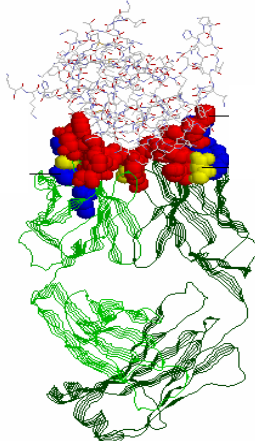
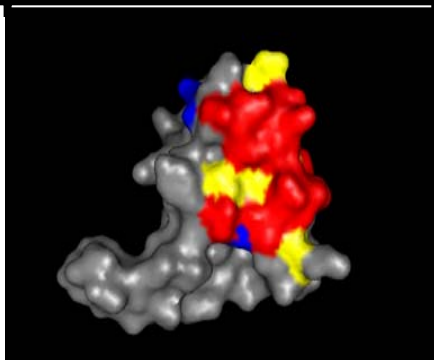
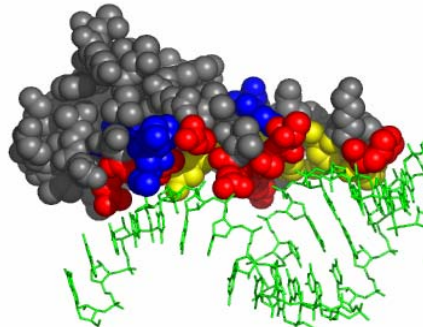
**An example:** use SVM to predict protein-protein interface residues



- **Input:** 11-residue window centered on target aa
- **Representation:** each aa = 20-element vector (derived from HSSP alignment)

**RESULTS:** Average classification performance in predicting interface residues (leave-one-out cross-validation)

Using only protein *sequence* as input:

Complex	Protein-Protein	Protein-DNA	Protein-RNA
Classifier	2-stage classifier SVM + Naïve Bayes	Naïve Bayes	Naïve Bayes
Accuracy			
Specificity			
Sensitivity			
Correlation coefficient			

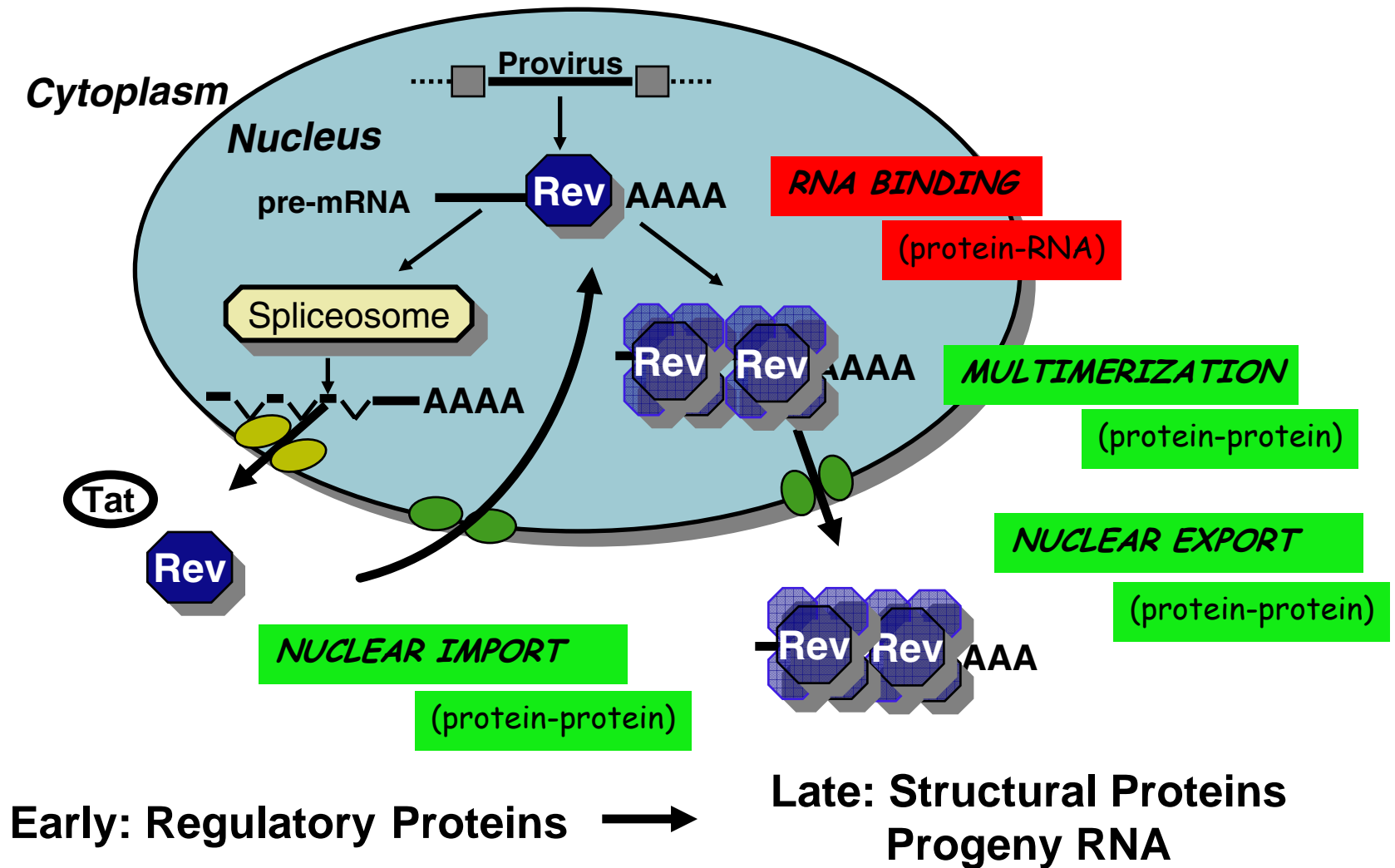
**TRUE +**  
**FALSE +**  
**FALSE -**  
**TRUE -**

*Ab FabN10*  
 Acc = 87%  
 CC = 0.65

*λ Repressor*  
 Acc = 88%  
 CC = 0.66

*dsRNA Binding Protein*  
 Acc = 85.5%  
 CC = 0.59

# Macromolecular interactions mediated by the Rev protein in lentiviruses (HIV & EIAV)

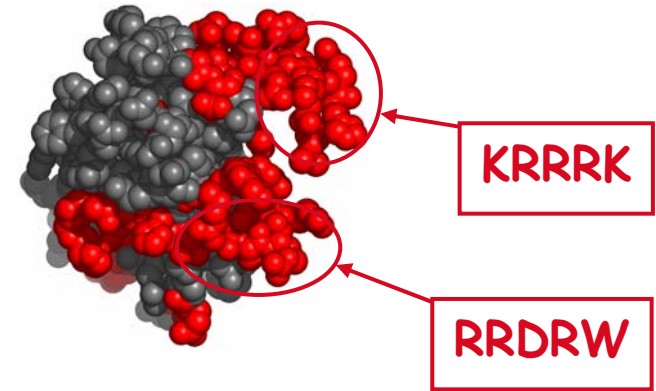


**PREDICTED:**

*Structure* →

*Protein binding residues* +

*RNA binding residues* +



41 51 61 71 81 91

GP L E S D Q W C R V L R Q S L P E E K I S Q T C I A R R H L G P G P T Q H T P S **RRDRW** I R E Q I L Q A E V L Q **ERLE** W R I

+++++++ ++  
 ++++++

**VALIDATED:**

*Protein binding residues* [Green box]

*RNA binding residues* [Red box]

131 141 151 161

Q R G D F S A W G D Y Q Q A Q E R R W G E Q S S P R V L R P G D S **KRRRK** H L

+++++++ ++ +++ ++++++

+ ++++++

