

# The predictions of RNA binding probabilities in nucleic acids binding proteins

**E. Westhof**

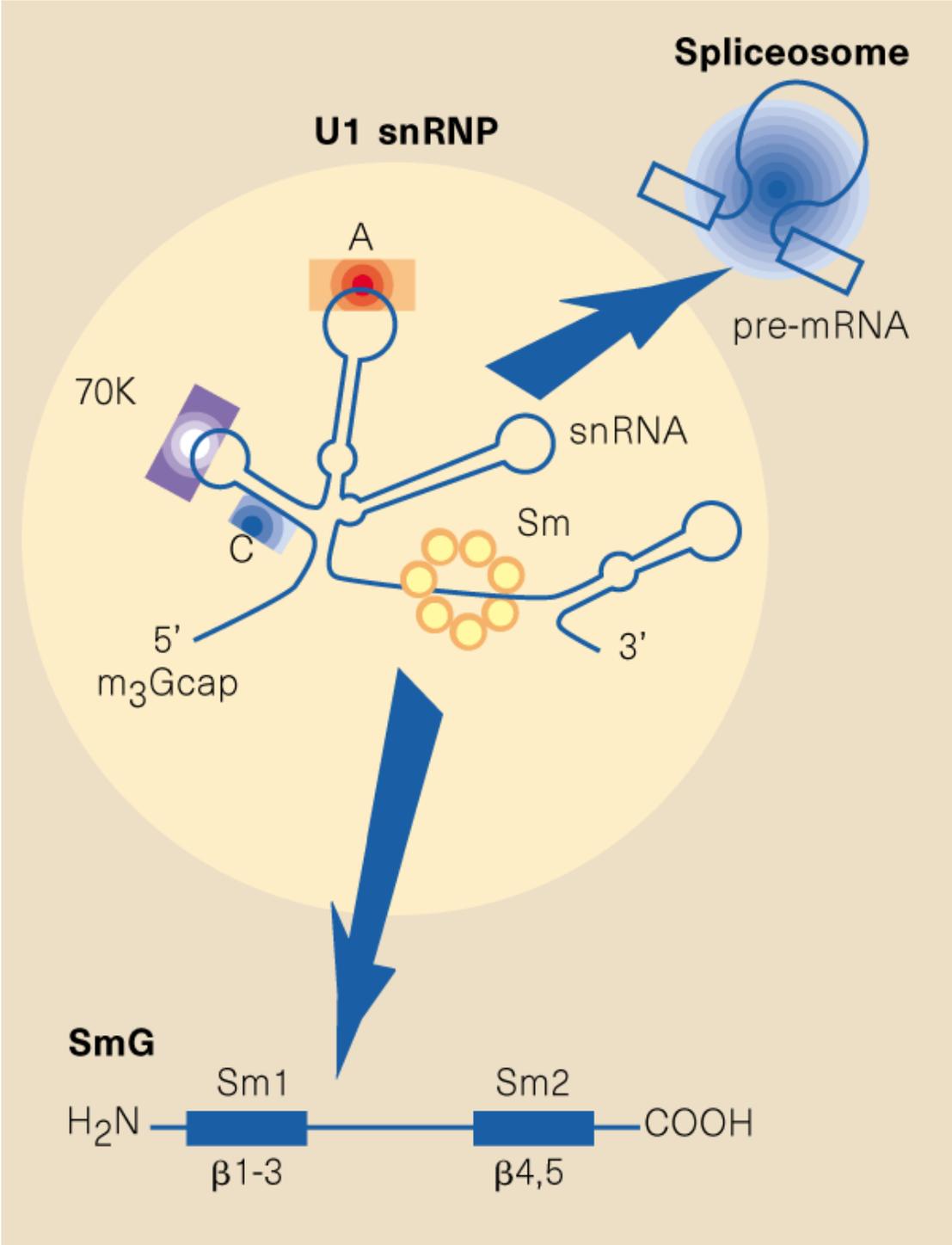
Architecture et Réactivité de l'ARN  
Institut de biologie moléculaire et cellulaire du CNRS

Université de Strasbourg

**2015**



Various proteins are recruited around RNA molecules



## Representative RNA–protein complexes.

The panels represent RNA–protein complexes between

(a) 5S rRNA and ribosomal proteins L5, L18, L21E, and L30 [PDB ID: 1JJ2]

(b) Gln tRNA and its synthetase [PDB ID: 1QTQ]

(c) domain IV of 4.5S RNA and SRP protein [PDB ID: 1HQ1]

(d) ribosomal protein L1 and its mRNA fragment [PDB ID: 2VPL]

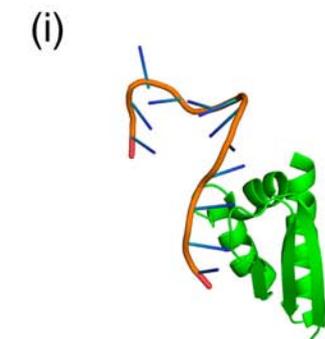
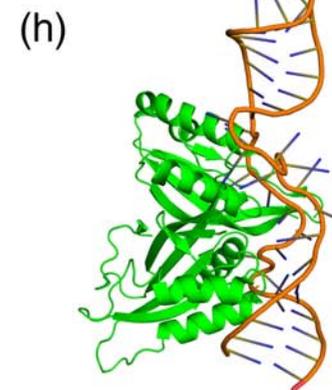
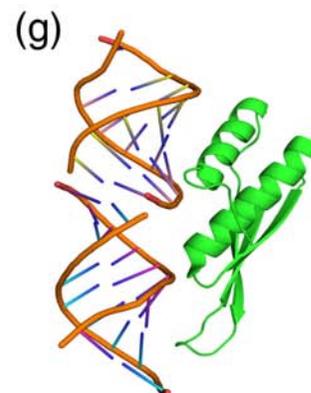
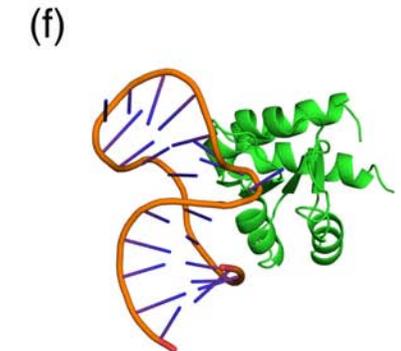
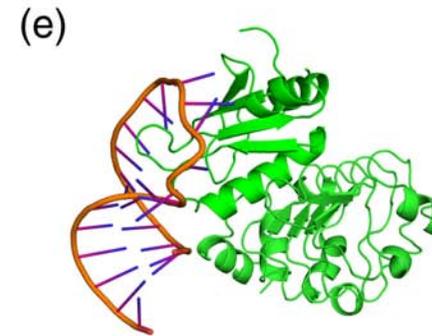
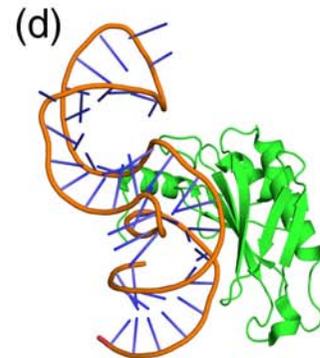
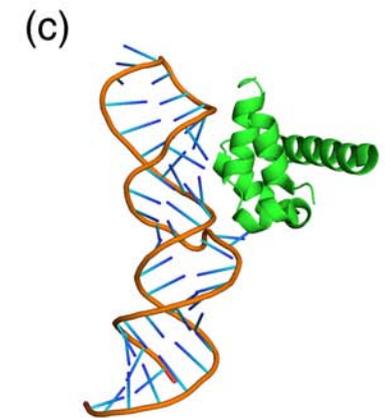
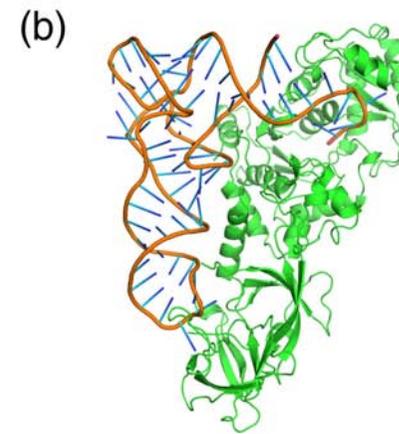
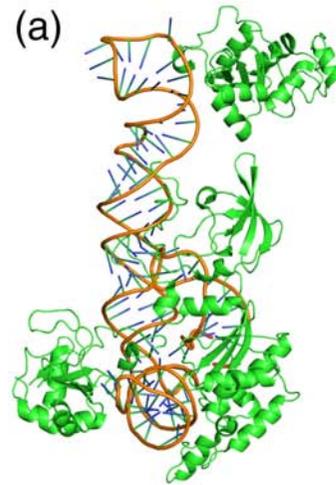
(e) hairpin loop IV of U2 snRNA and U2B"/U2A' protein complex [PDB ID: 1A9N]

(f) box C/D snoRNA and L7Ae protein [PDB ID: 1RLG]

(g) fragment of pre-miRNA and RISC-loading complex subunit [PDB ID: 3ADL]

(h) P3 domain of RNase P/MRP RNA and POP6 and POP7 protein subunits [PDB ID: 3IAB]

(i) C-rich fragment of telomeric RNA and polyC binding protein [PDB ID: 2PY9]



Aditi Gupta , Michael Gribskov

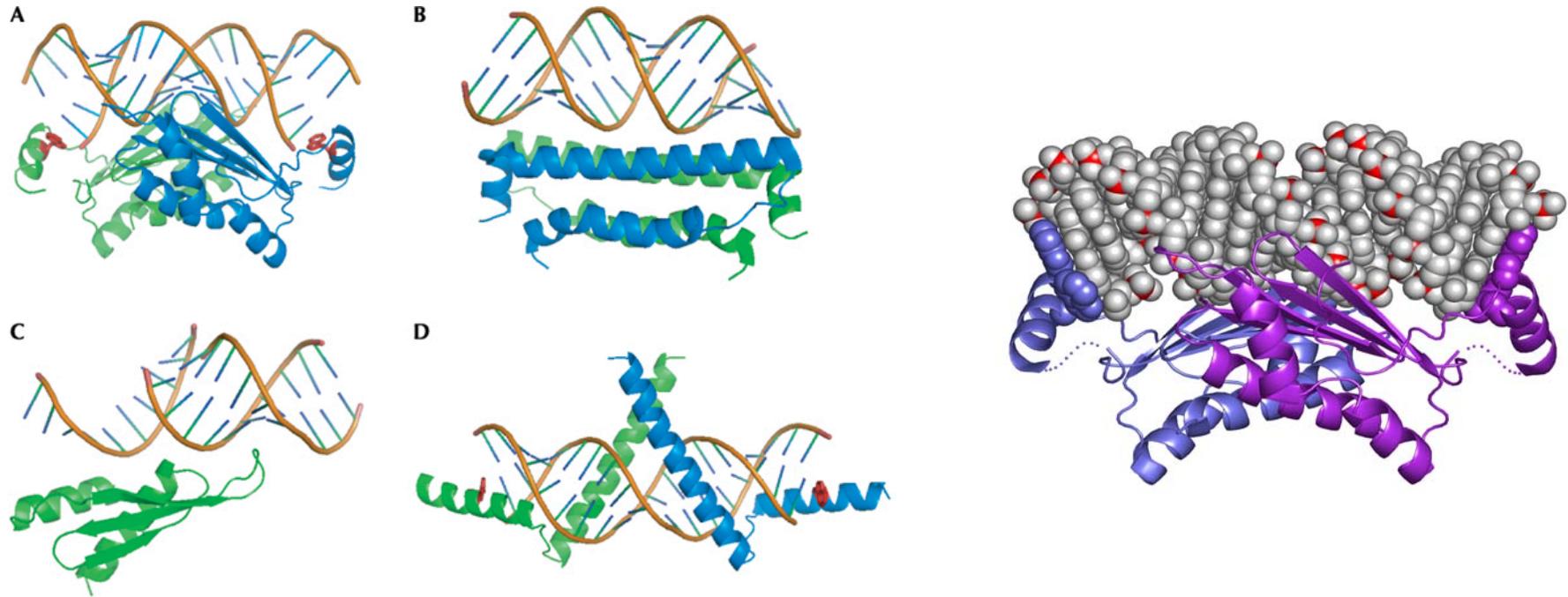
### The Role of RNA Sequence and Structure in RNA/Protein Interactions

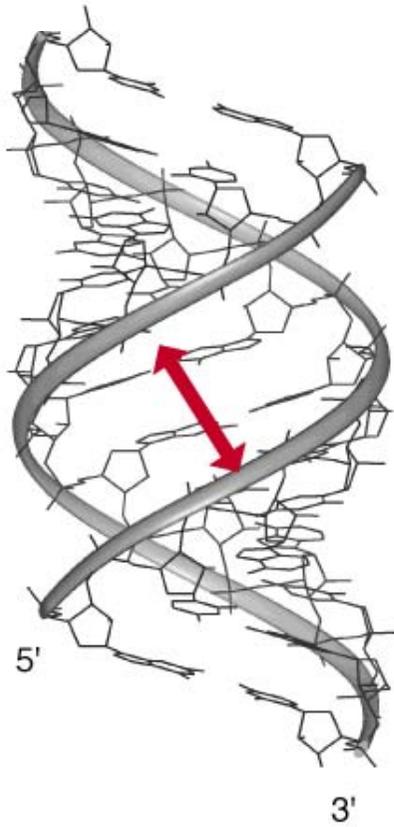
Journal of Molecular Biology 409, 4 2011 574 - 587

<http://dx.doi.org/10.1016/j.jmb.2011.04.007>

# RNA double helix recognition

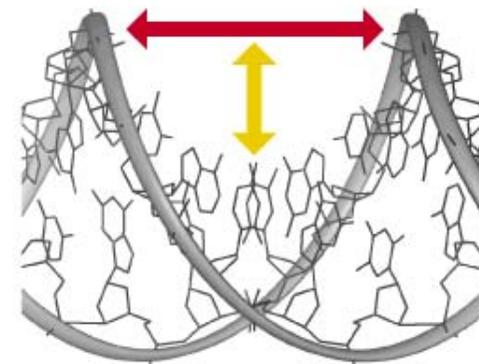
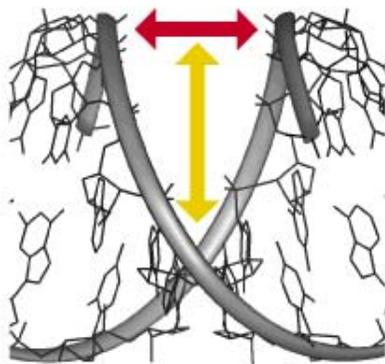
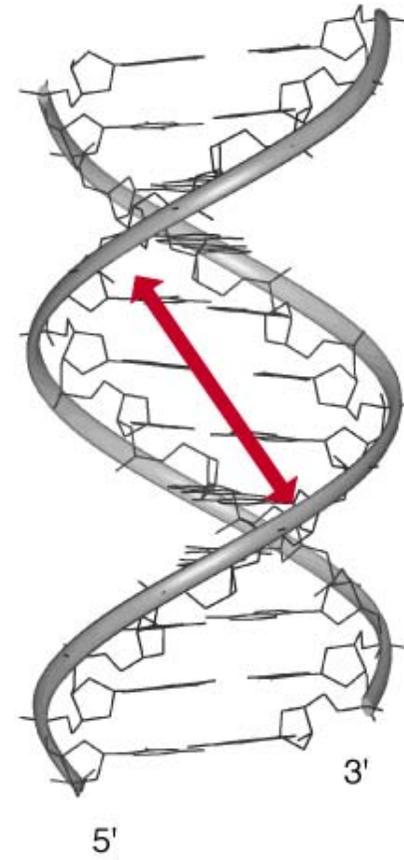
Proteins structurally very different  
recognize dsRNAs





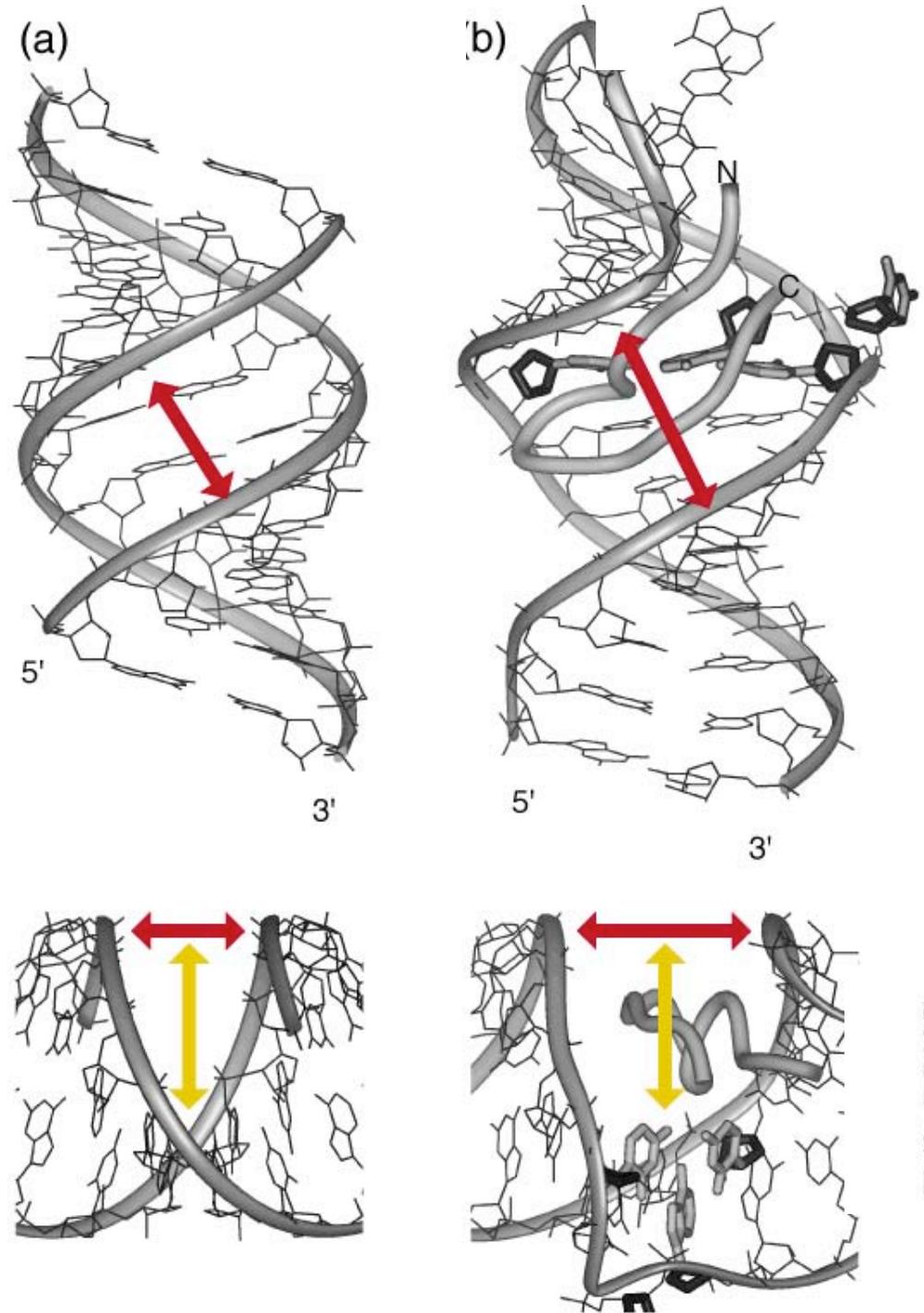
A-DNA  
RNA

B-DNA



# In RBPs

Key roles of non-Watson-Crick pairs and bulges for enlarging the deep major groove and for diversity in H-bonding sites



# RNA versus protein modules

No correlation between them

RNA Module

Protein Module

Double helix

Many folds

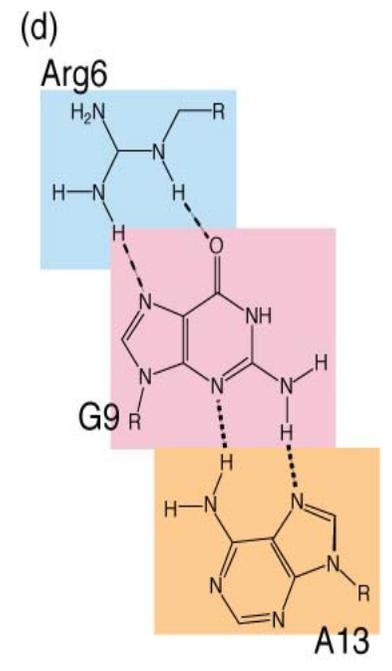
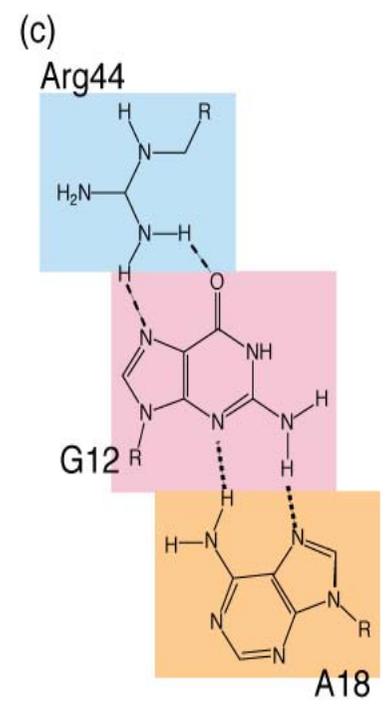
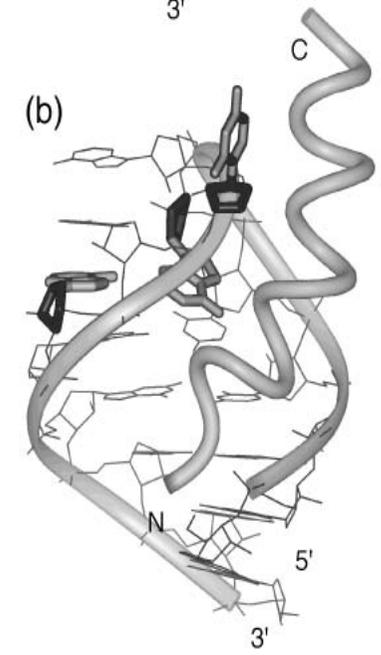
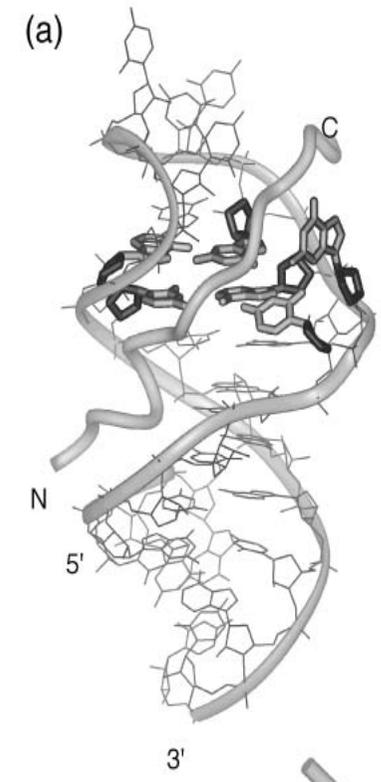
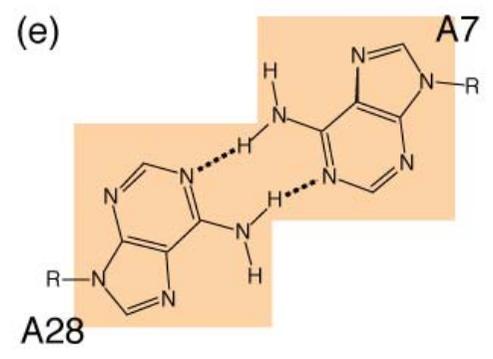
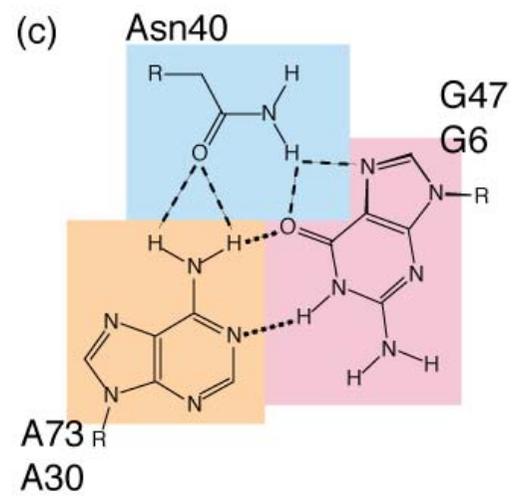
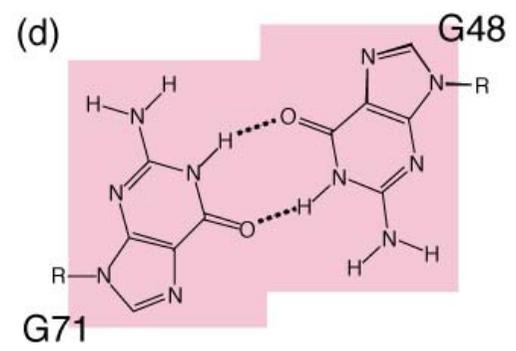
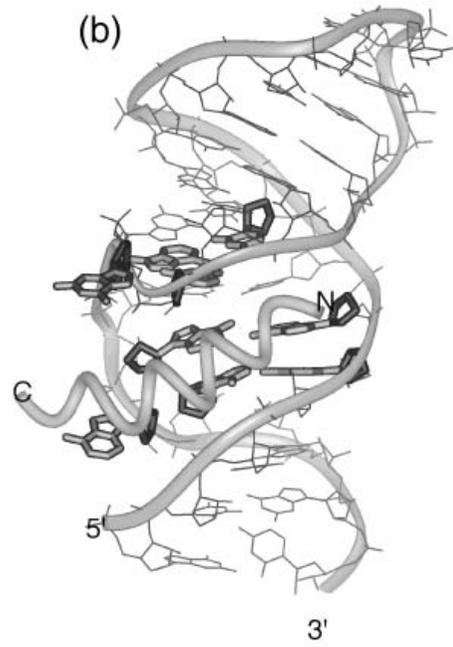
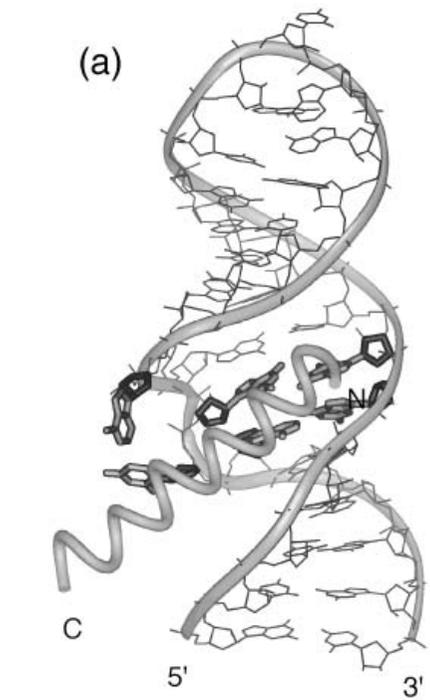
Many elements

RRM

Loop E, K-turn

Many residues





**RBscore**

**Prediction of nucleic  
acid binding probability**  
Zhichao Miao

<http://ahsoka.u-strasbg.fr/rbscore/>  
*Nucleic Acids Research, in press 2015.*

The score is directly derived from **physico-chemical** and **evolutionary** features and integrates a residue **neighboring network** approach.

The score integrates, via weight factors, the **sequence and structure** information of all known RNP crystal structures

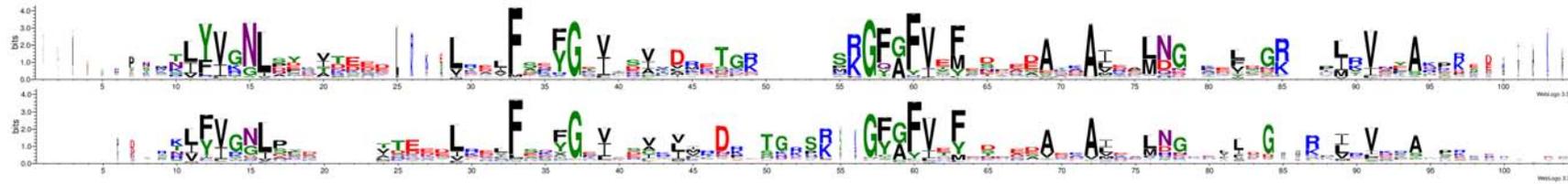
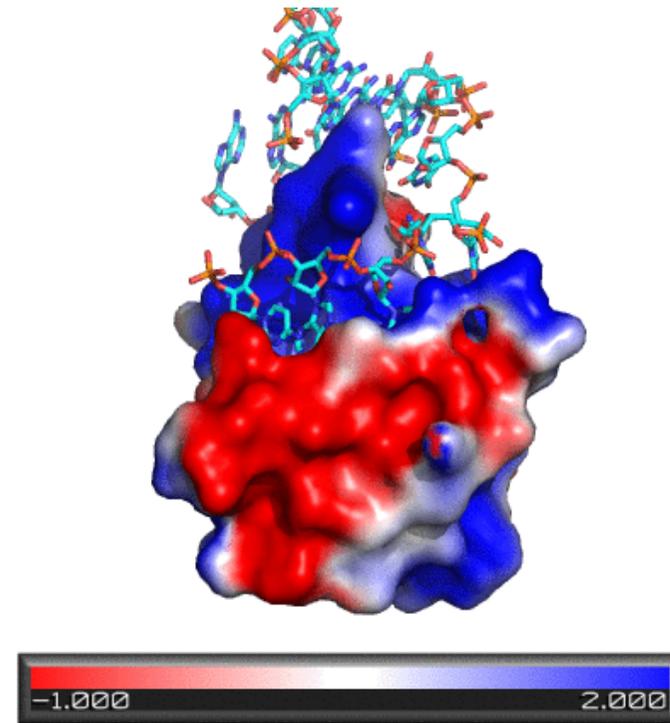
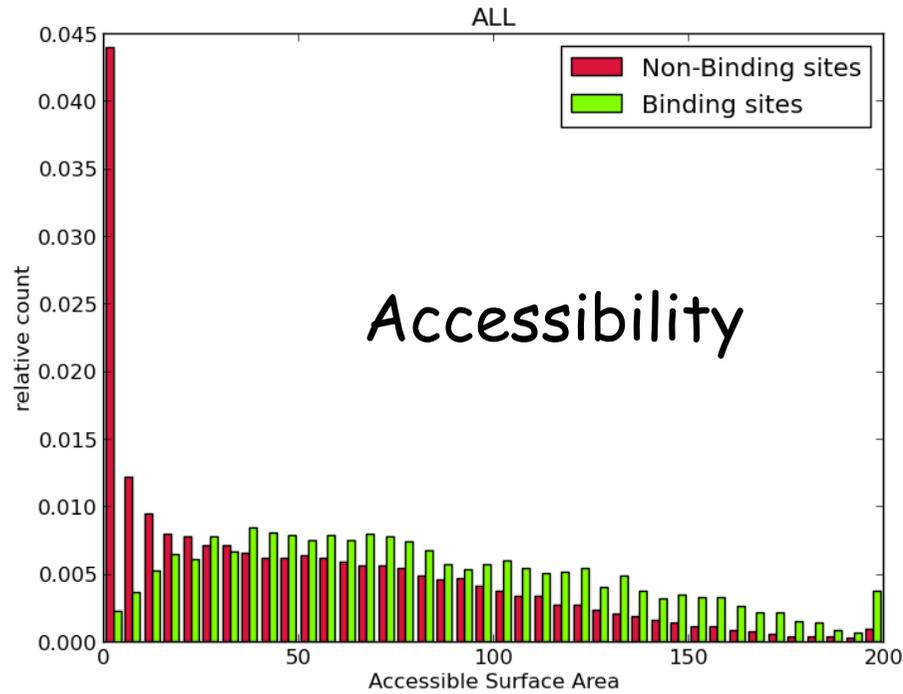
Advantages :

- direct links with the feature values;
- no difference between RNA and DNA binding proteins;
- no memory of binding motifs;
- identification of cryptic binding sites and « moonlighting » proteins;
- continuous score (>> energy funnel).

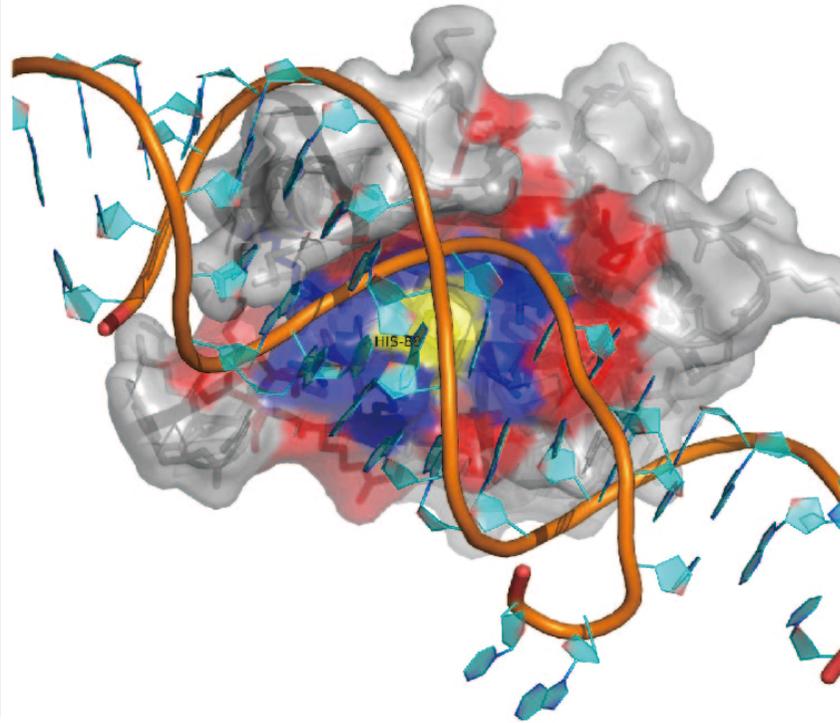
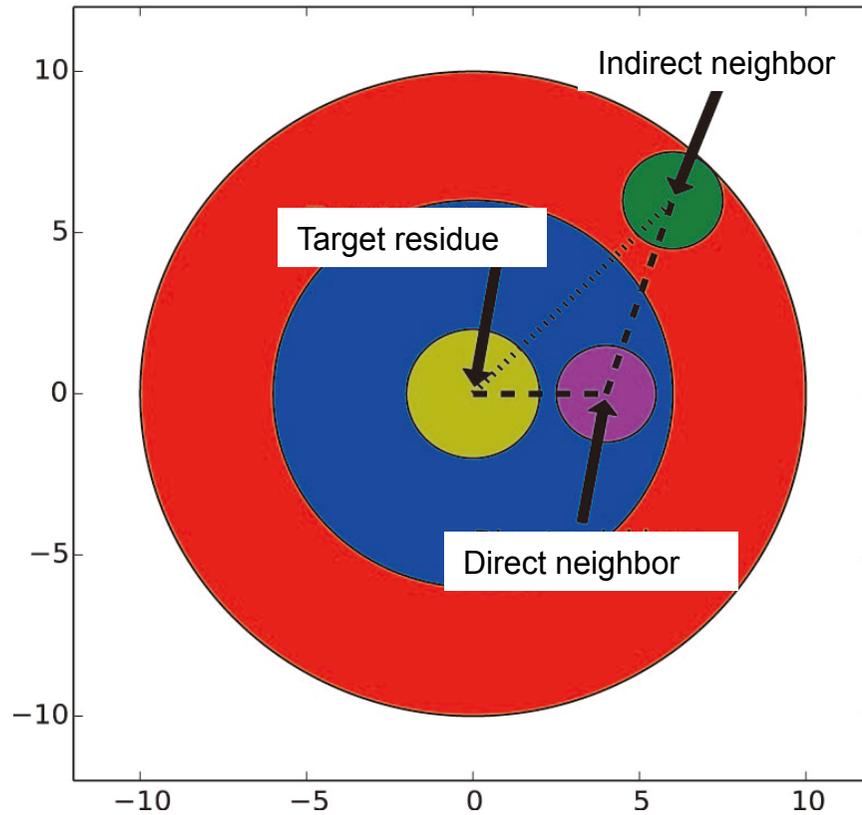
INPUT: either the coordinates of a crystal structure of the protein (full score) or the sequence of the protein (SVM)

- RBP structures were obtained from NPIDB database (Jan 2014) with resolution better than 3.5 Å and R factor  $<0.3$  as criteria. Sequence identity  $<25\%$  and TMscore  $<0.7$  were used as thresholds to remove redundancy:
- 130 protein chains were collected as a **training set**,
- 117 protein chains were taken as an **independent test set**.

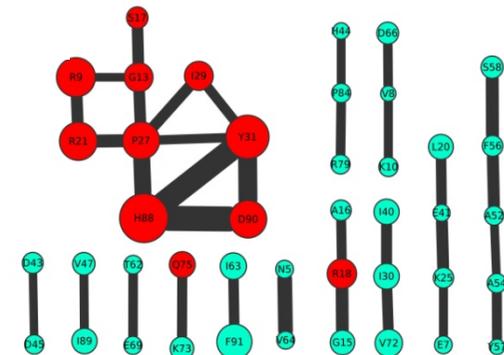
# Features of RNA binding residues



# Neighboring Network integration of the features



The RNA binding likelihood of a residue is determined by its environment including direct and indirect neighboring relationship.

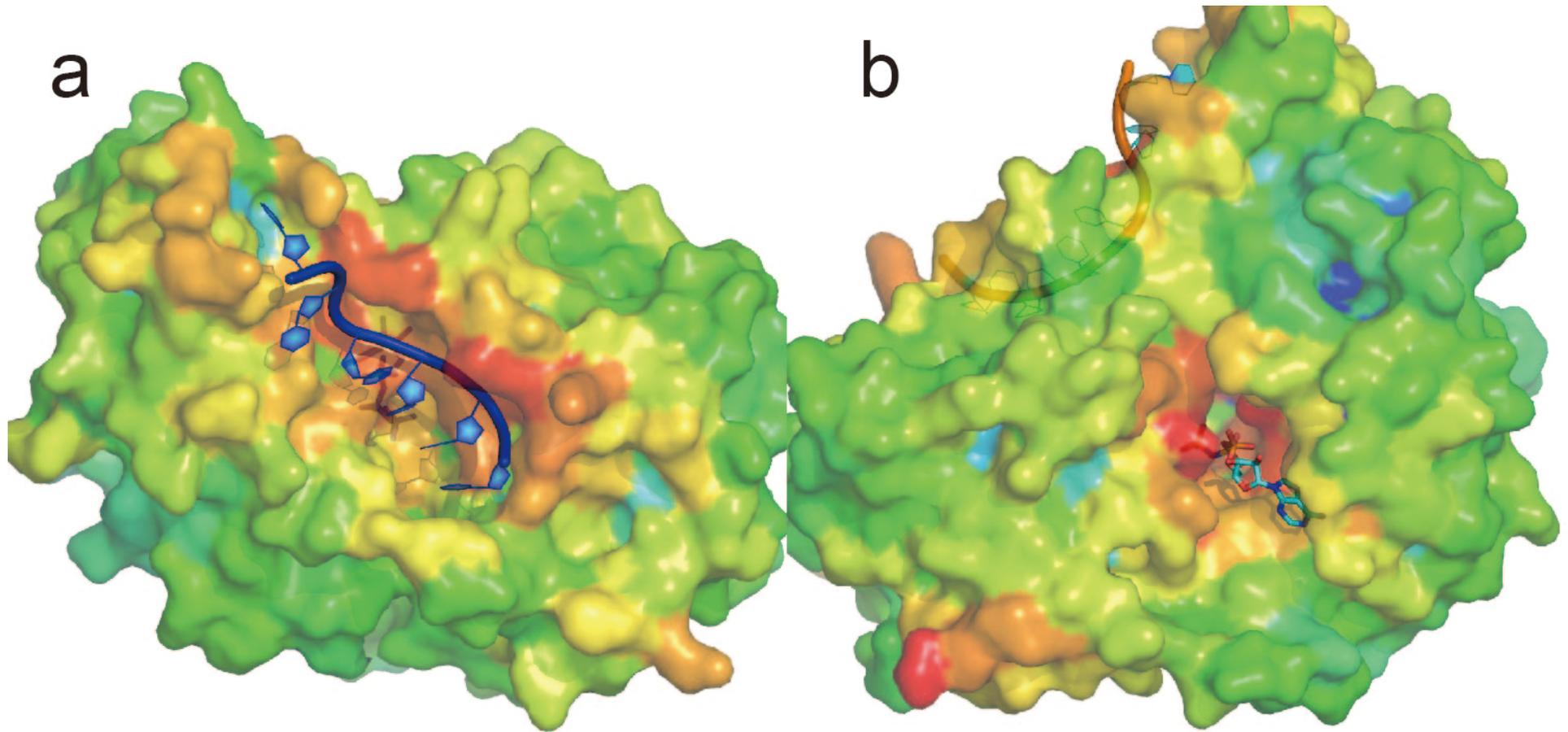


$$E_{\text{feat}} = w_{aa} \times \text{ASA} + \sum_{10} w_i \times \text{count}_{\text{ELEC}}(i) + w_{\text{CE}} \times \text{CE} + w_{\text{width}} \times \text{width} + C_{aa}$$

$$E_{\text{pred}} = E_{\text{feat}} + \frac{\sum_{\text{direct neighbors}} u_{aa} \times f_{\text{neighbors type}} \times E_{\text{feat}}}{N_{\text{direct neighbors}}} + \frac{\sum_{\text{indirect neighbors}} v_{aa} \times g_{\text{neighbors type}} \times E_{\text{feat}}}{N_{\text{indirect neighbors}}}$$

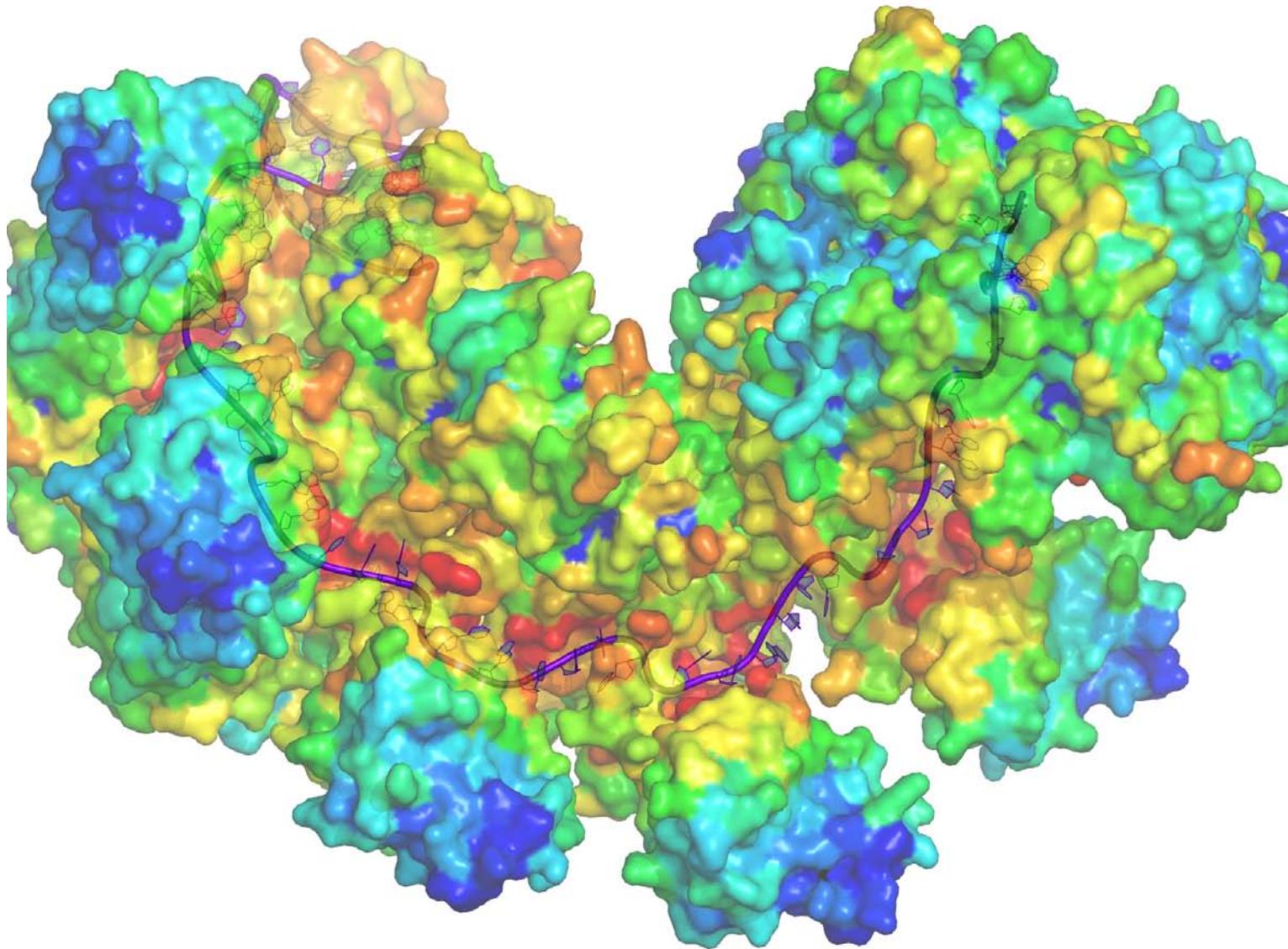
The final score includes 104 parameters (52 for feature score, 52 for neighboring network) in total and is a linear combination of all these features.





RBscore mapping on ATP-dependent RNA helicase  
DDX19 (PDB id 3G0H):

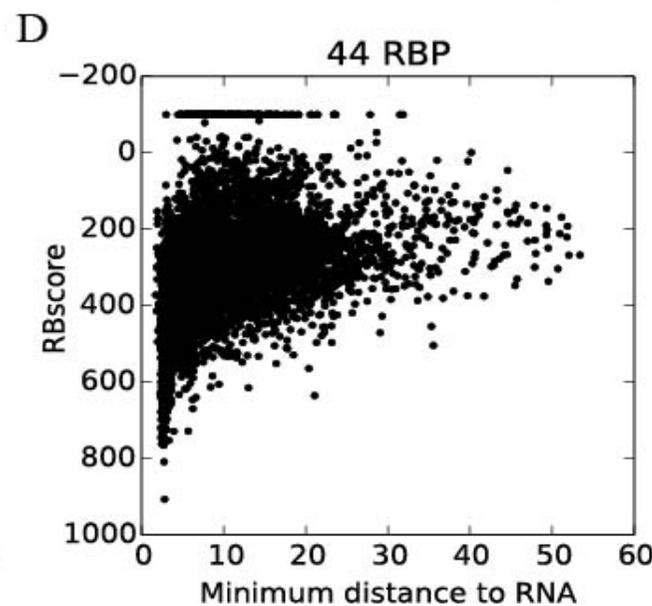
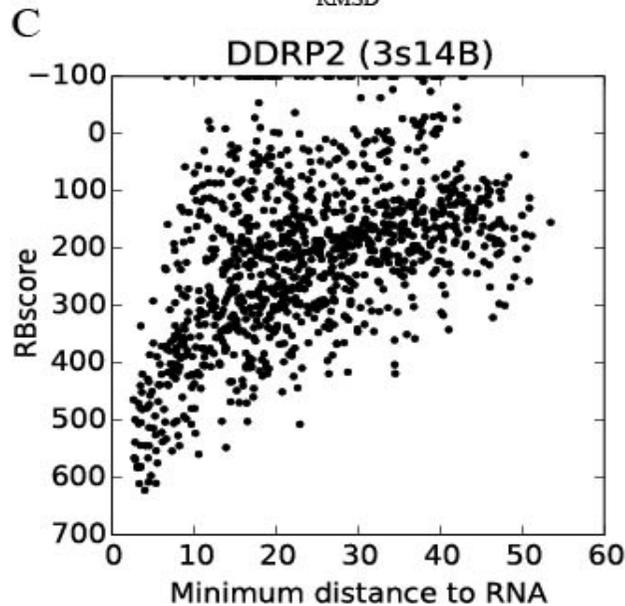
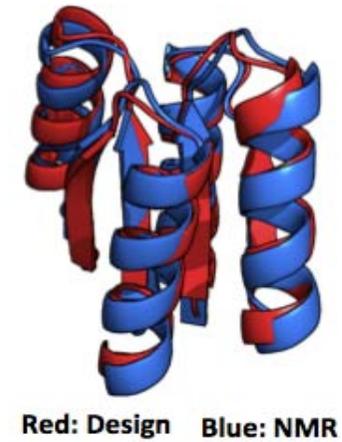
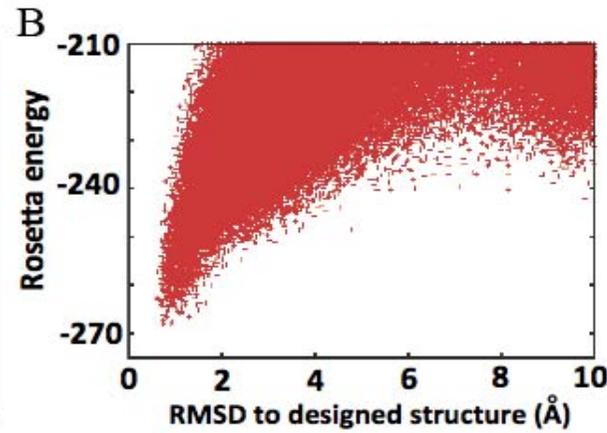
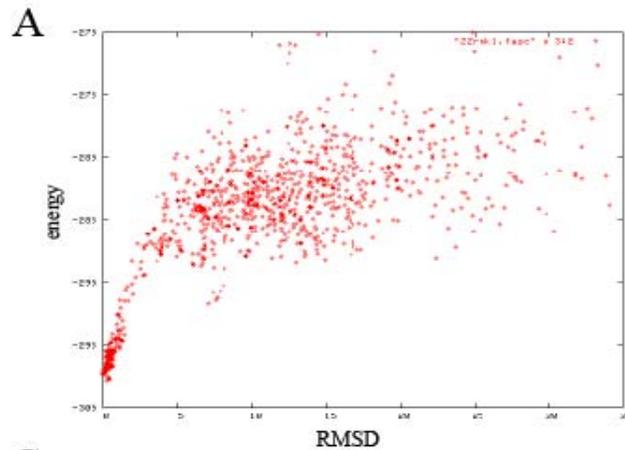
- a) The RNA binding region,
- b) the ATP-analogue (ANP) binding region.

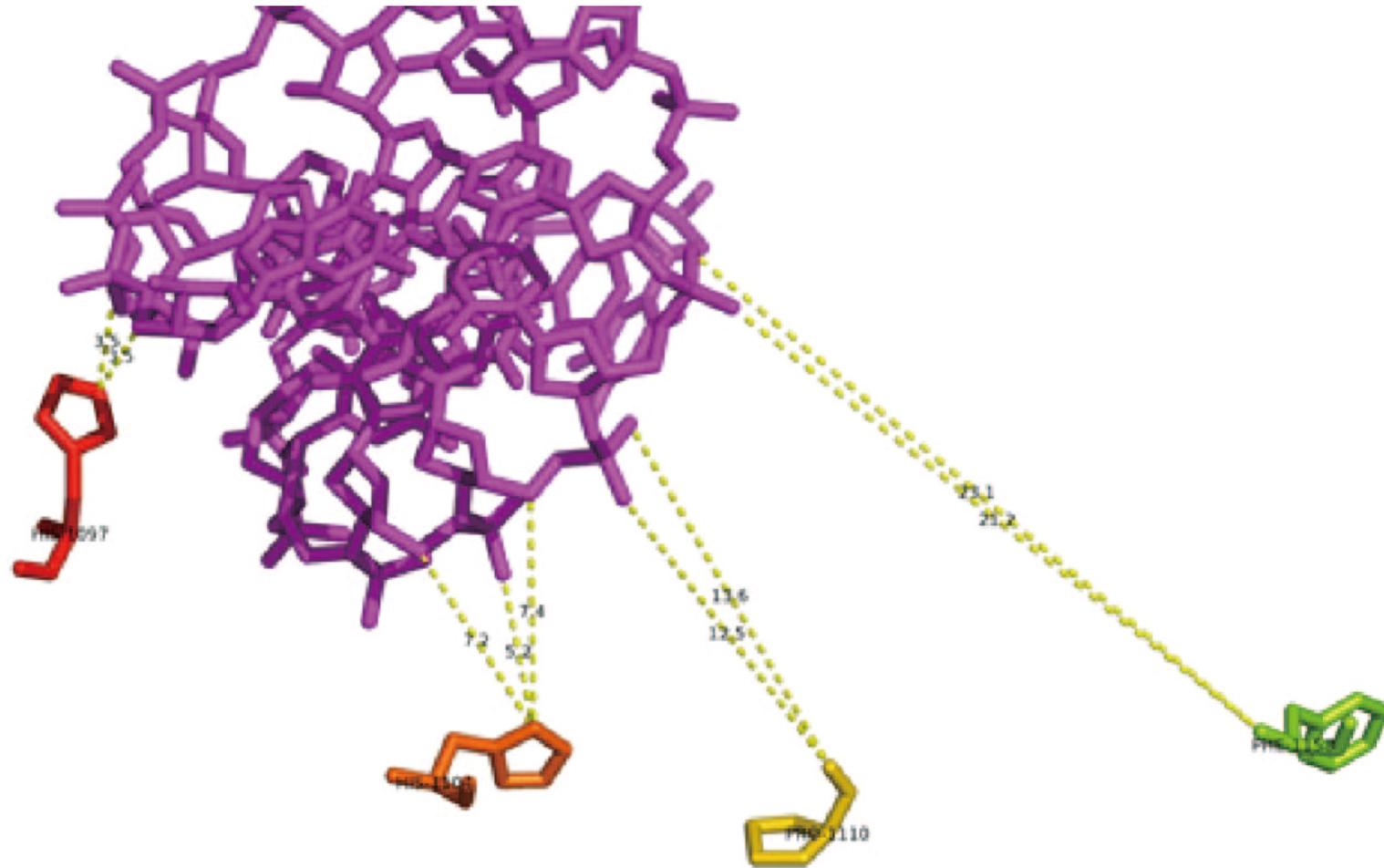


RBscore of RNA-guided Cascade complex mapped on structure with rainbow color. As each protein chain in the complex was predicted separately, the final map of RBscore still corresponds to exact binding region of RNA.

# RNA binding energy funnels

RNA binding energy funnel is similar to protein-RNA docking (A) and protein folding (B) energy funnels

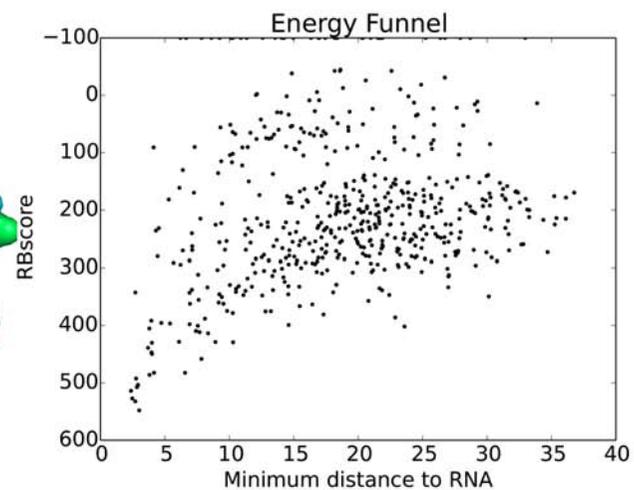
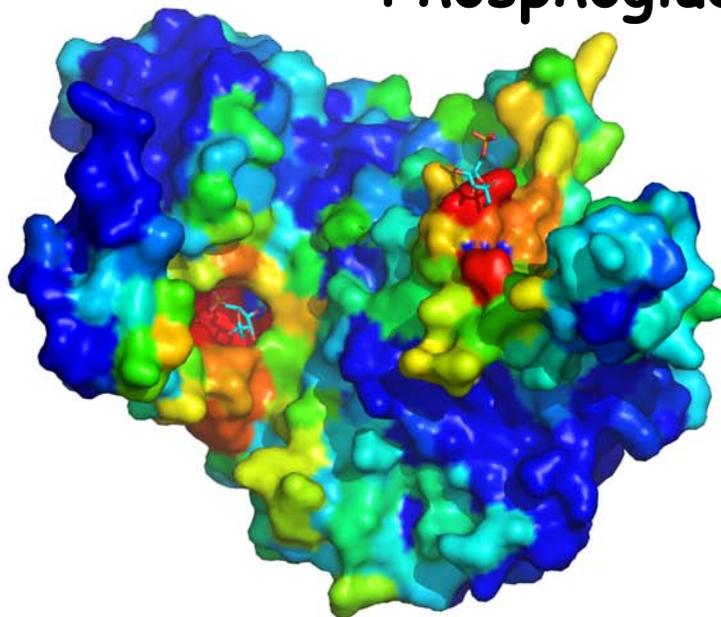




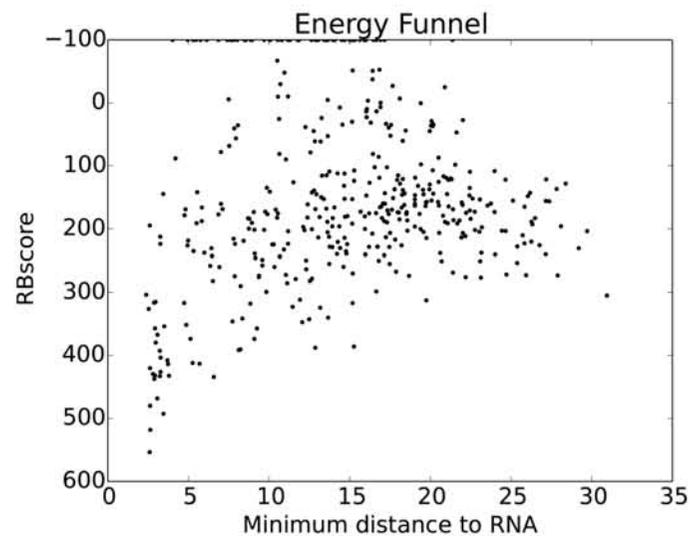
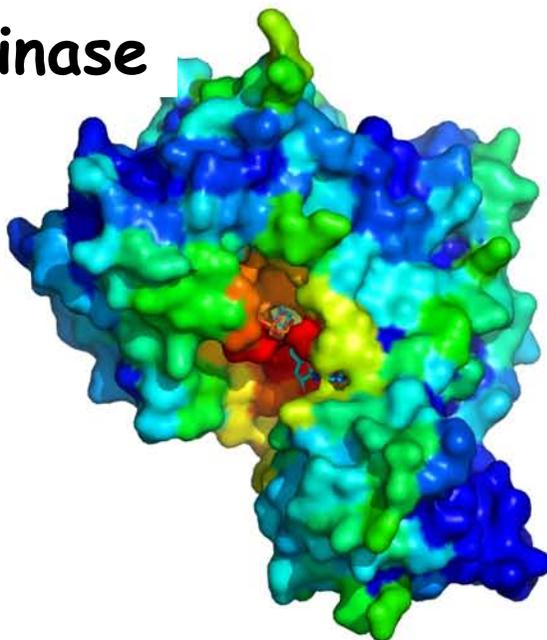
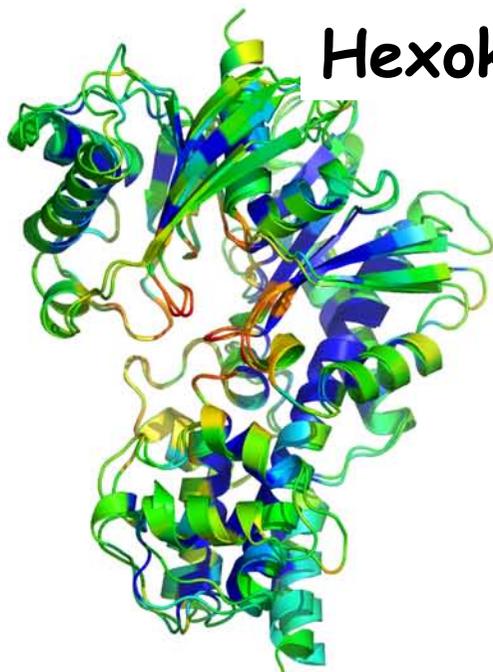
Hierarchical scoring of RBscore and related residue distance to RNA resulting energy funnel like pattern. RBscore of the residues are colored according to red to green color scheme. Minimum distances from the residues to RNA are marked. Residues further away from RNA have worse RBscore. The four residues form part of the energy funnel on the protein surface.

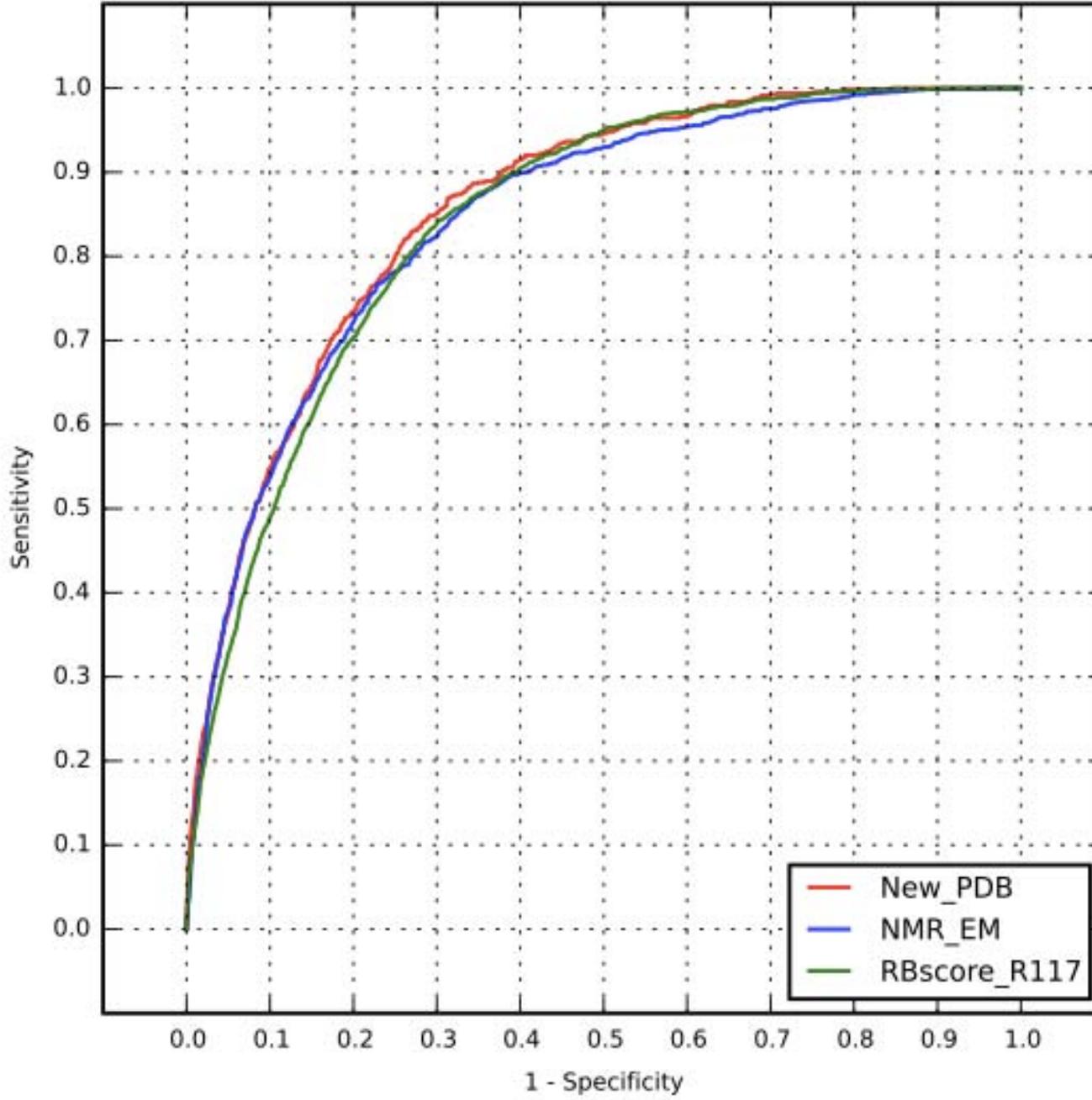
**Glycolytic  
enzymes  
As  
RNP**

## Phosphoglucose isomerase



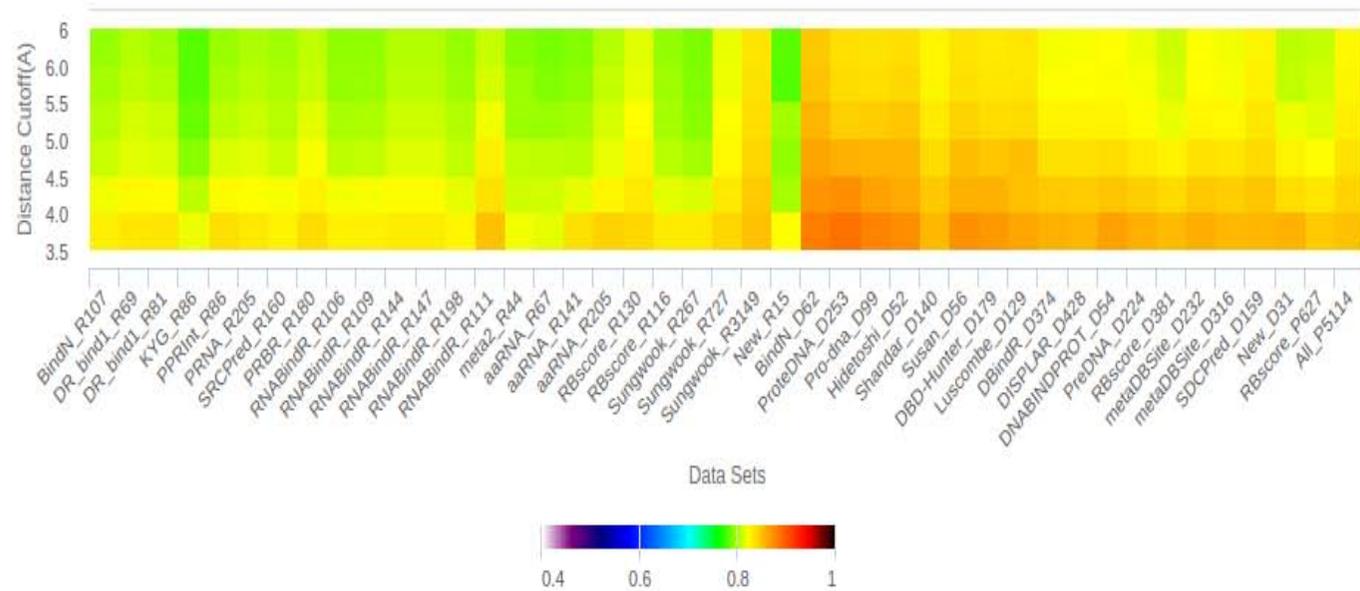
## Hexokinase



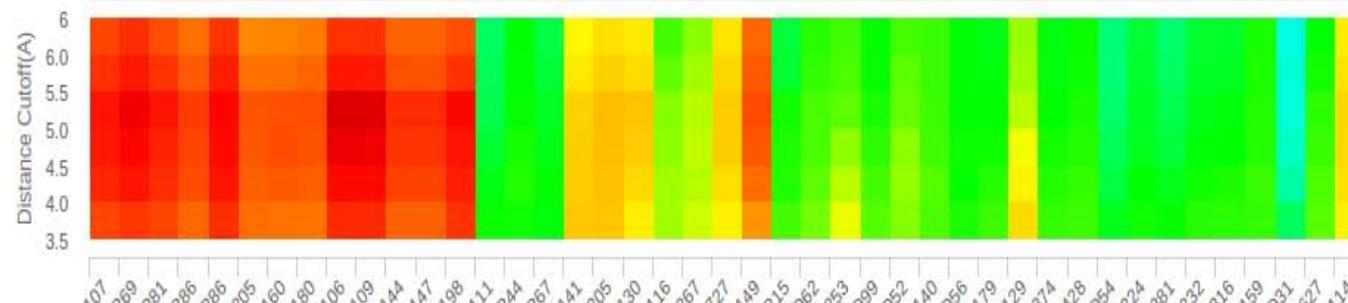


RBscore on NMR and EM structures (first model in the PDB file). Those were excluded from previous training and tests.

## aaRNA



## RNABindRPlus



<http://ahsoka.u-strasbg.fr/nbench/>



# RNA Binding score

RBscore: a web server to predict RNA binding sites on protein

## RBscore Pages

- [Introduction Page](#)
- [Predict With Structure](#)
- [Predict With Sequence](#)
- [Find a Result](#)
- [Datasets of RBscore](#)
- [Help Page](#)

## Institute

- [IBMC](#)
- [Architecture et Réactivité de L'ARN](#)
- [CNRS](#)

## Related tools

- [BindN+](#)
- [BindN](#)
- [RNABindRPlus](#)
- [RNABindR2.0](#)
- [PPRIInt](#)

## Predict

### Input protein information to prediction nucleic acid binding sites

- I have a PDB structure
- I only have a protein sequence

♣ Please input a **PDB code**:

Or **Upload a PDB file**:  No file chosen

♣ **chain**:

♣ are required as inputs

Your **Email Address**:  (optional)



[Privacy & Terms](#)

<http://ahsoka.u-strasbg.fr/rbscore/>

UPR 9002 du CNRS,  
Architecture et Réactivité de l'ARN,  
Institut de Biologie Moléculaire et Cellulaire,  
Université de Strasbourg



**Z. Miao (IBMC)**

French Government grant  
ANR-10-BINF-02-02 " BACNET "

