

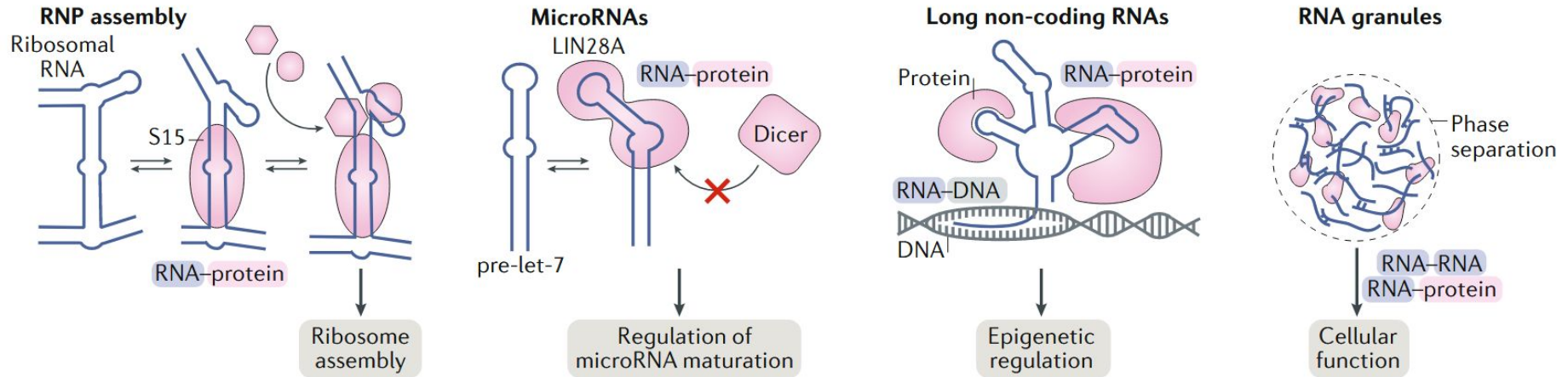
Grigory Nikolaev

ClaPNAC: a Classifier of Protein – Nucleic Acid Contacts

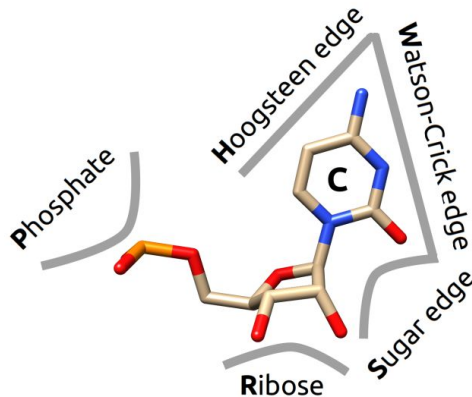
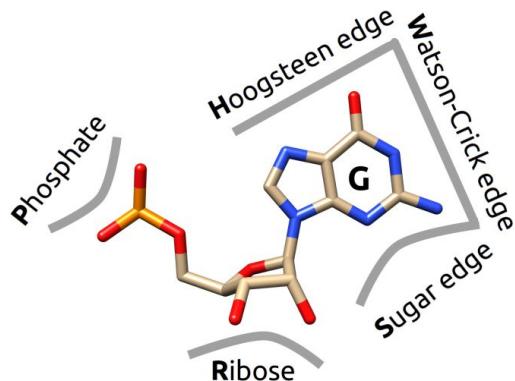
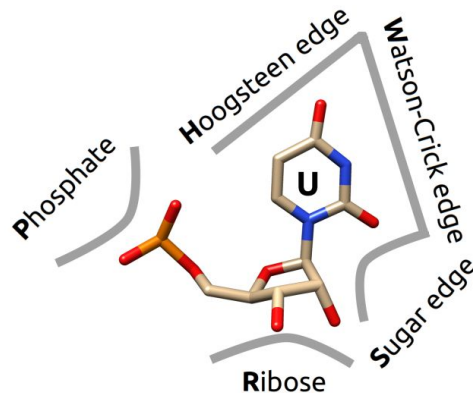
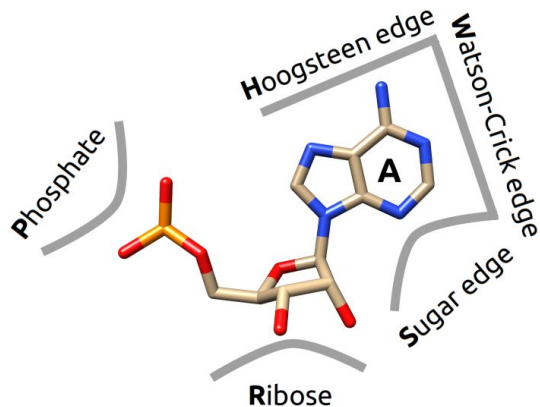
Laboratory of Bioinformatics and Protein Engineering (prof. Janusz M. Bujnicki)

Benasque 13 July 2026

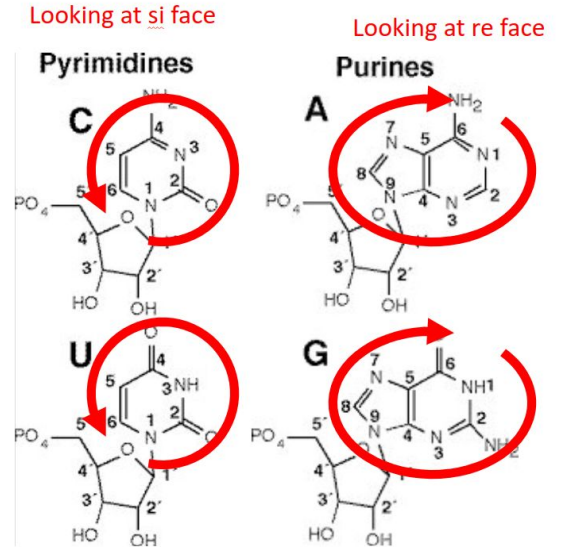
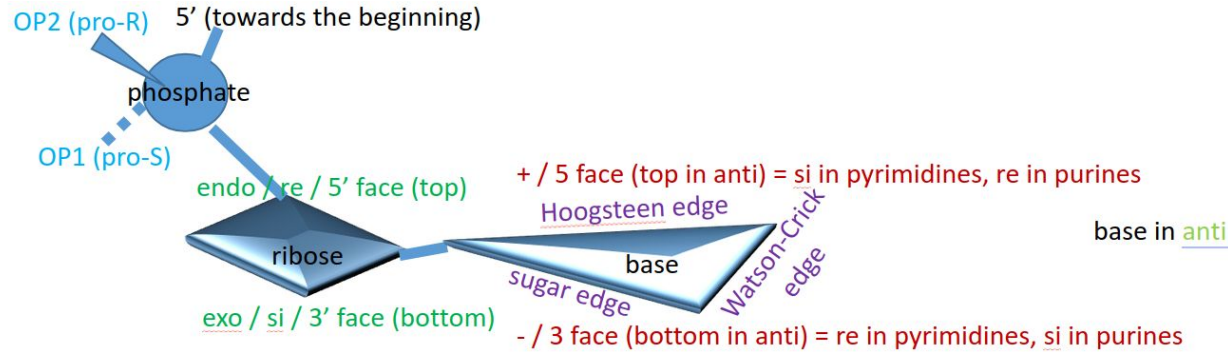
Protein-RNA interactions are essential



Interaction sites: base edges and backbone moieties



Interaction sites in ribonucleotides: base faces

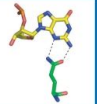
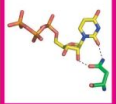


Classification of RNA-protein contacts

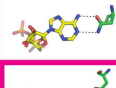
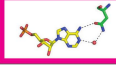
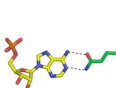
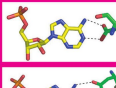
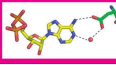
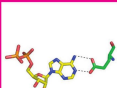
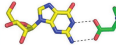
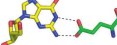
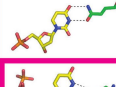
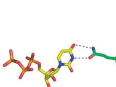
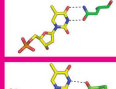
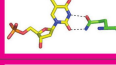
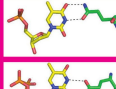
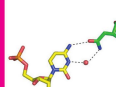
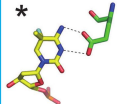
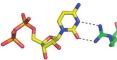
Hoogsteen edge pseudo pairs

	Asn	Gln	Asp	Glu	Arg
A					
G					
U					
T					
C					

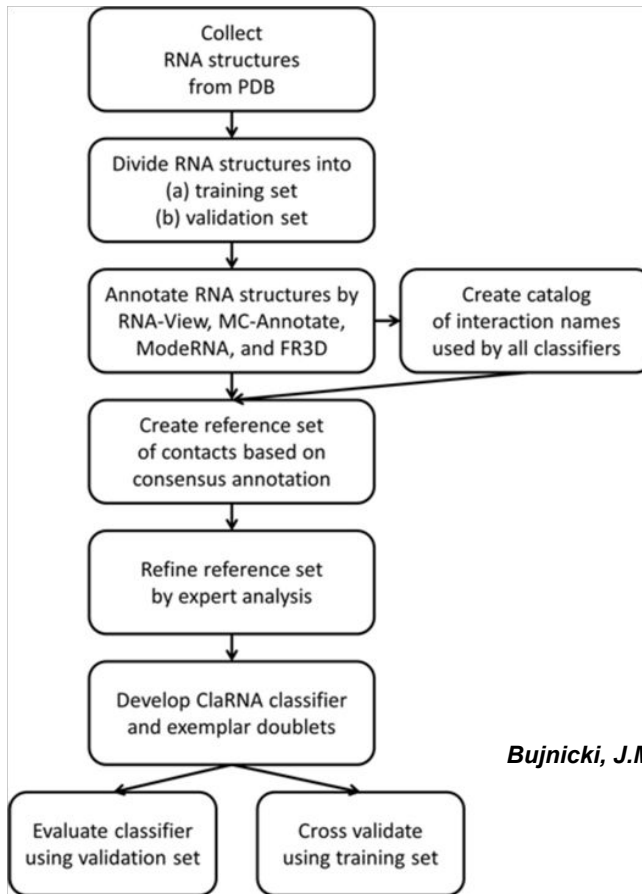
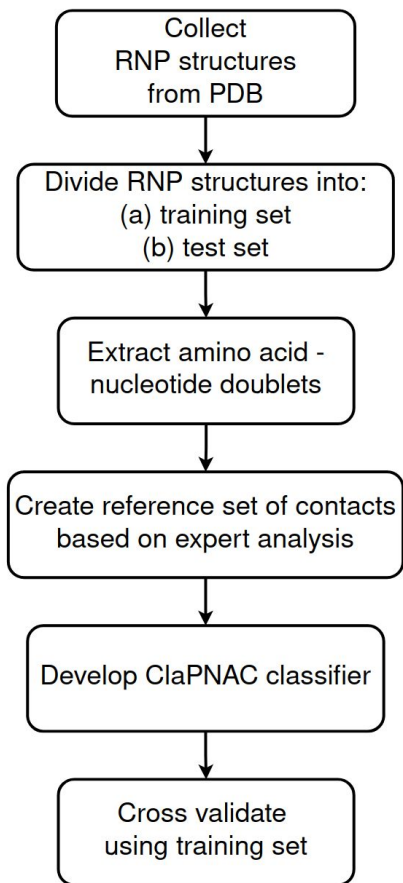
Sugar edge pseudo pairs

	Asn	Gln	Asp	Glu	Arg
A					
G					
U					
T					
C					

Watson–Crick edge pseudo pairs

	Asn	Gln	Asp	Glu	Arg
A	 		 		
G					
U	 				
T	 	 			
C					

Pipeline for ClaPNAC development (ClaRNA based approach)



Bujnicki, J.M. et.al. (2014) *Nucleic Acids Res.*, 42, e151

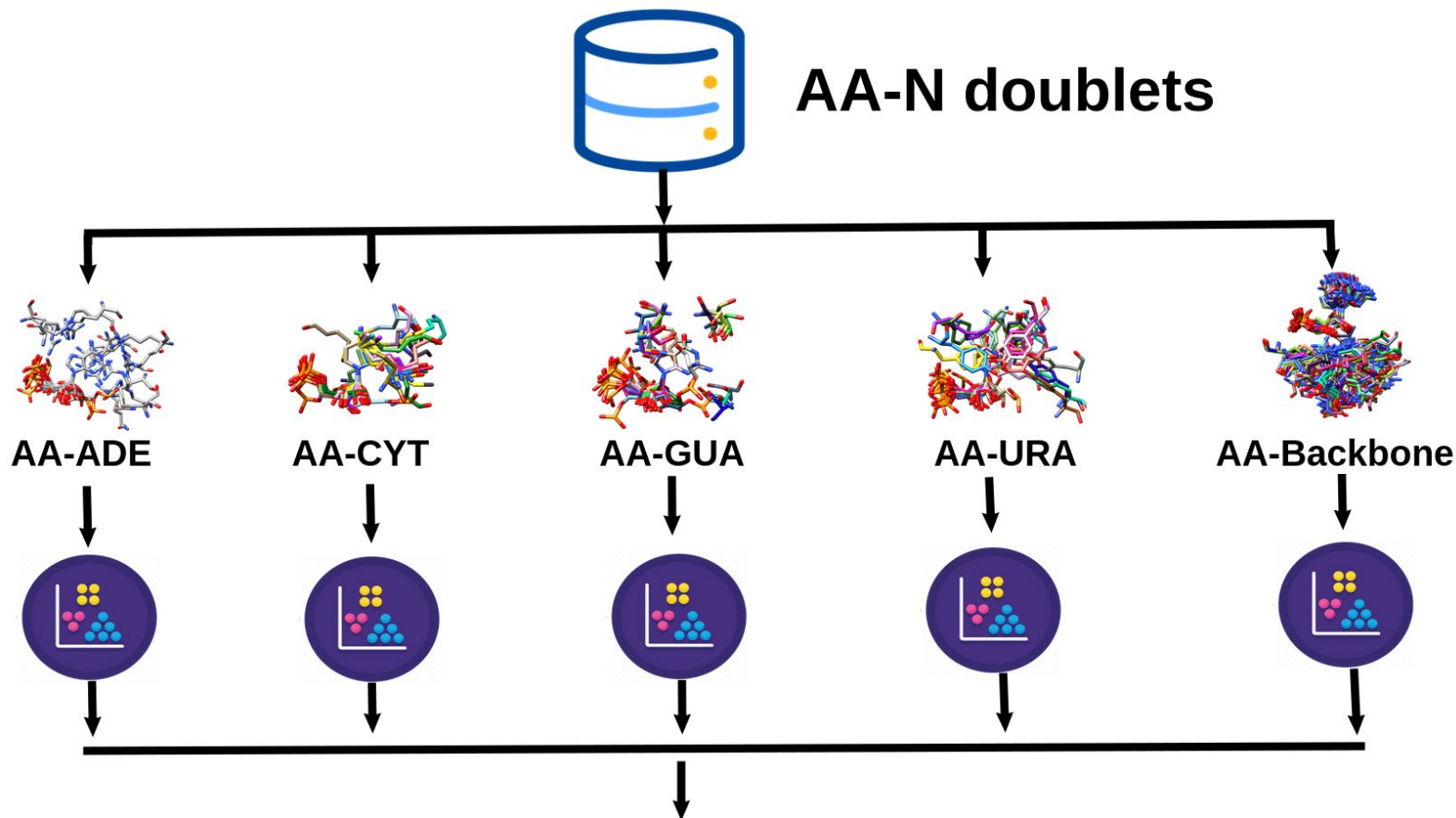
A curated dataset of biological assemblies of RNPs

	ADE	CYT	GUA	URA
ALA	769	356	768	1,012
ARG	10,203	10,518	12,135	9,325
ASN	2,179	1,574	2,069	2,733
ASP	1,353	710	2,075	1,332
CYS	86	22	65	132
GLN	1,970	1,501	2,443	2,145
GLU	1,038	834	1,808	1,177
GLY	2,065	1,558	2,371	1,695
HIS	2,026	2,110	2,406	1,753
ILE	451	386	504	508
LEU	617	523	775	768
LYS	5,306	4,780	6,637	4,463
MET	545	665	697	552
PHE	1,176	676	1,005	740
PRO	803	614	707	890
SER	3,896	2,365	3,238	2,020
THR	6,066	1,552	1,968	1,571
TRP	738	249	626	556
TYR	1,839	1,694	1,260	2,602
VAL	1,022	324	693	630
			TOTAL	158,013

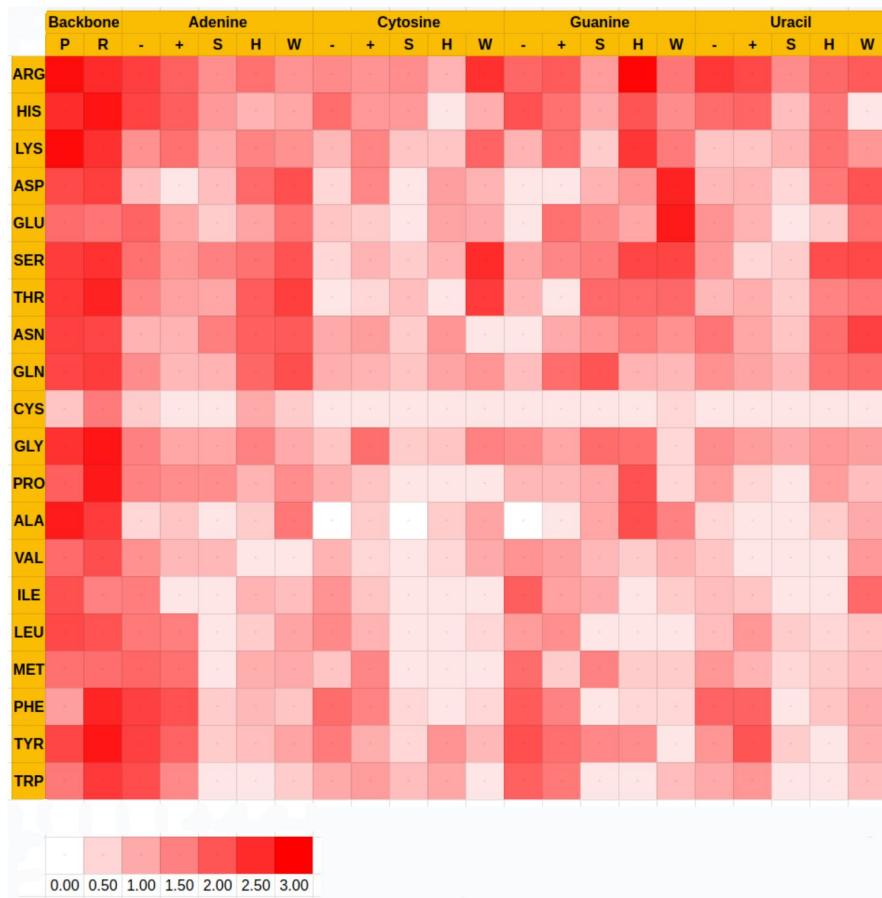
Number of nucleotide(N) - amino acid (AA) pairs:
- distance between 2 atoms of AA and
one atom of N < 3.5 Å

2,164 RNP complexes from the dataset →
extracted 158,013 doublets divided
into 20 groups by AA

ClaPNAC training dataset

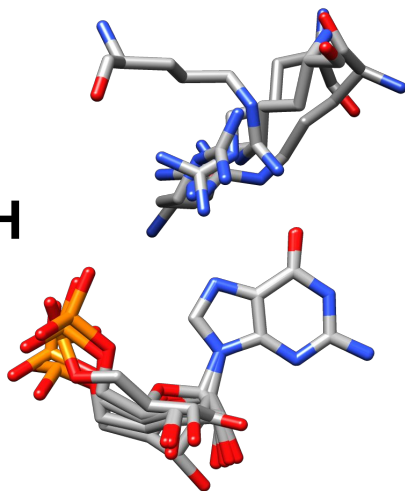


ClaPNAC training dataset

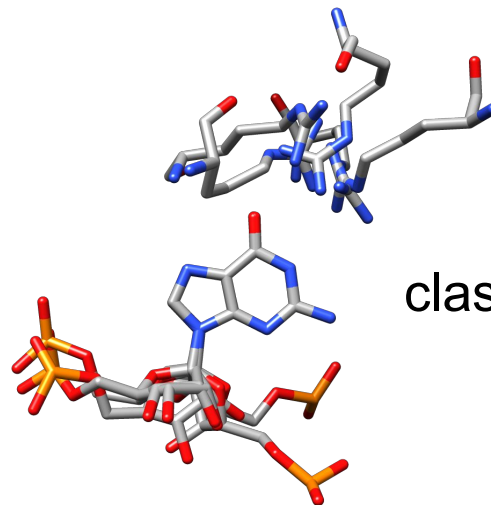


Classes for sidechain ARG-GUA base interactions

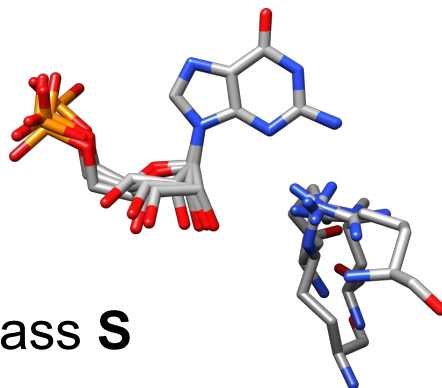
class H



class W

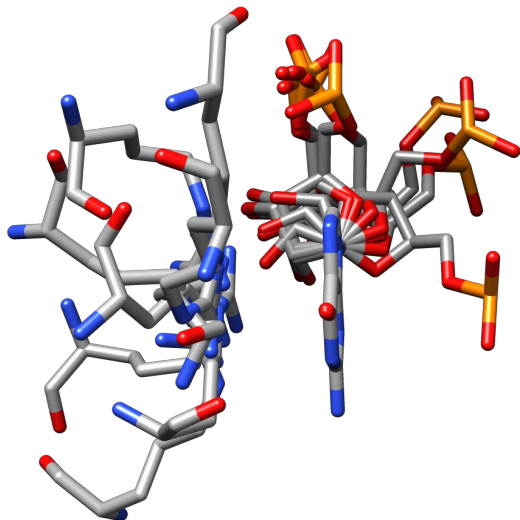


class S

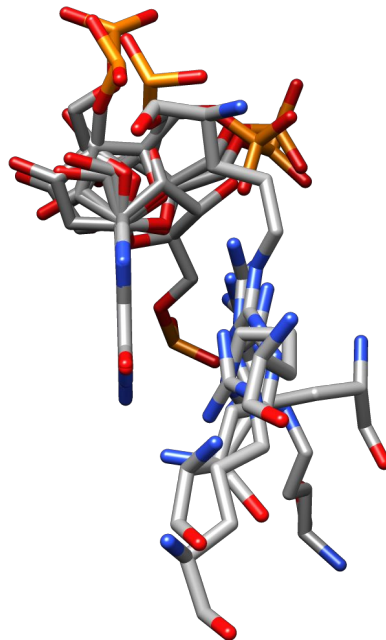


Classes for ARG-GUA interactions with – and + base faces

class -

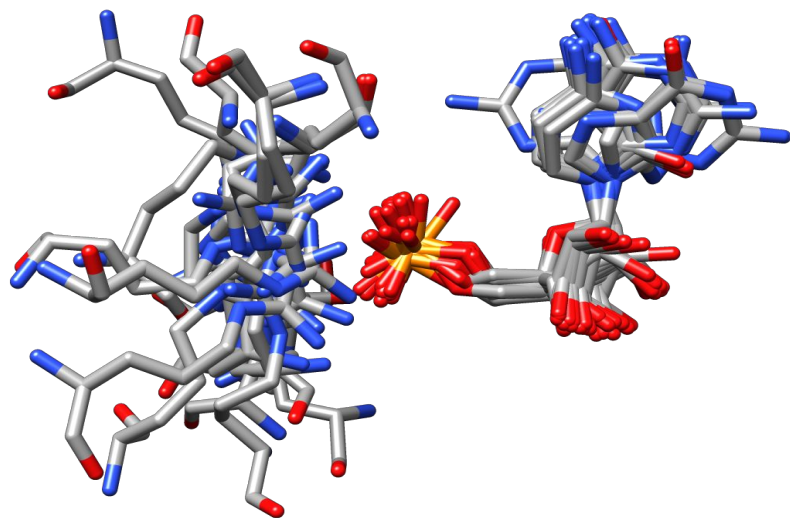


class +

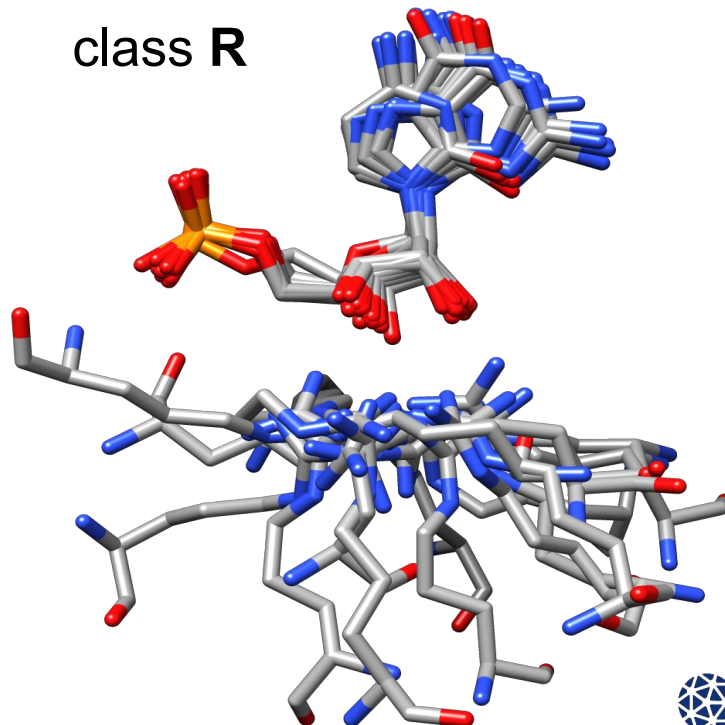


Classes for sidechain ARG-backbone interactions

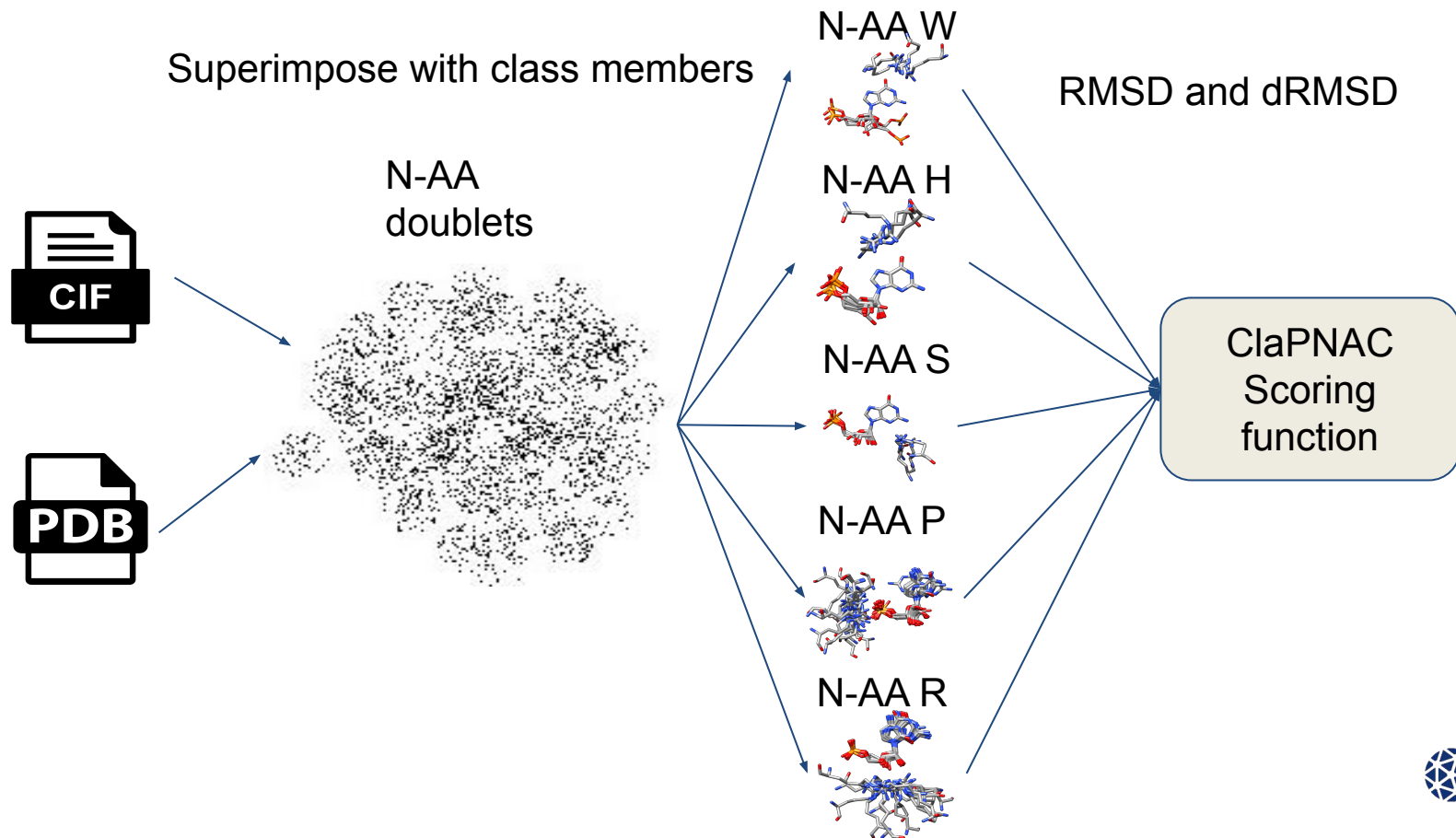
class P



class R



N-AA classification workflow



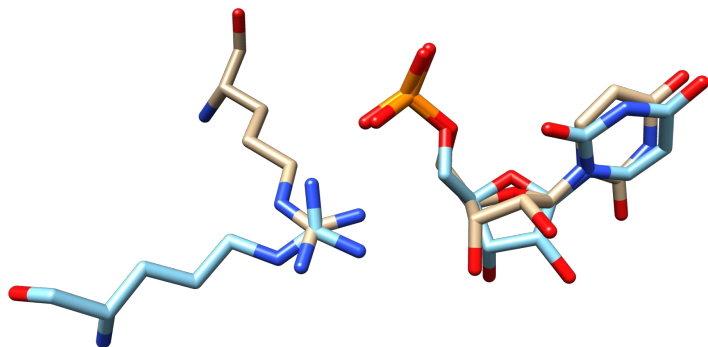
Distance calculation

Root Mean Square Deviation (**RMSD**)

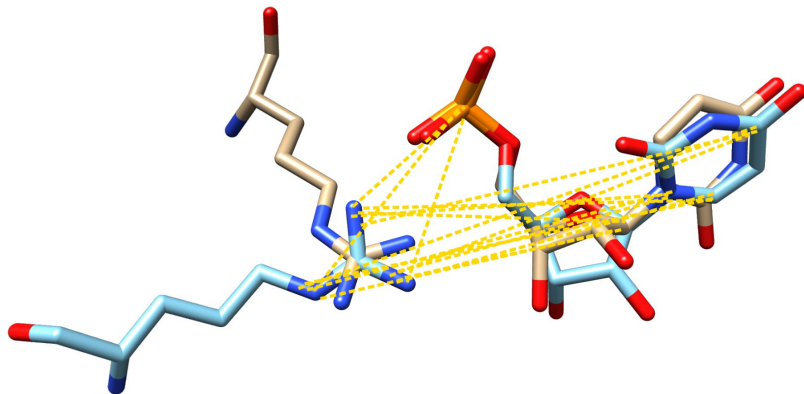
dissimilarity between two groups of atomic coordinates

distance-based RMSD (**dRMSD**)

dissimilarity between intramolecular atomic distance matrices

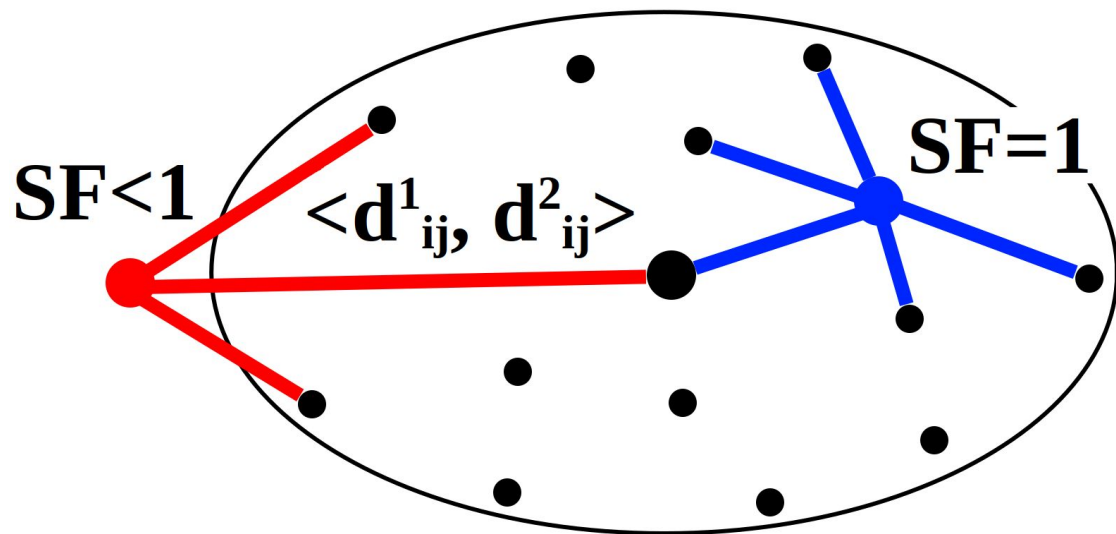


RMSD



dRMSD

Scoring function

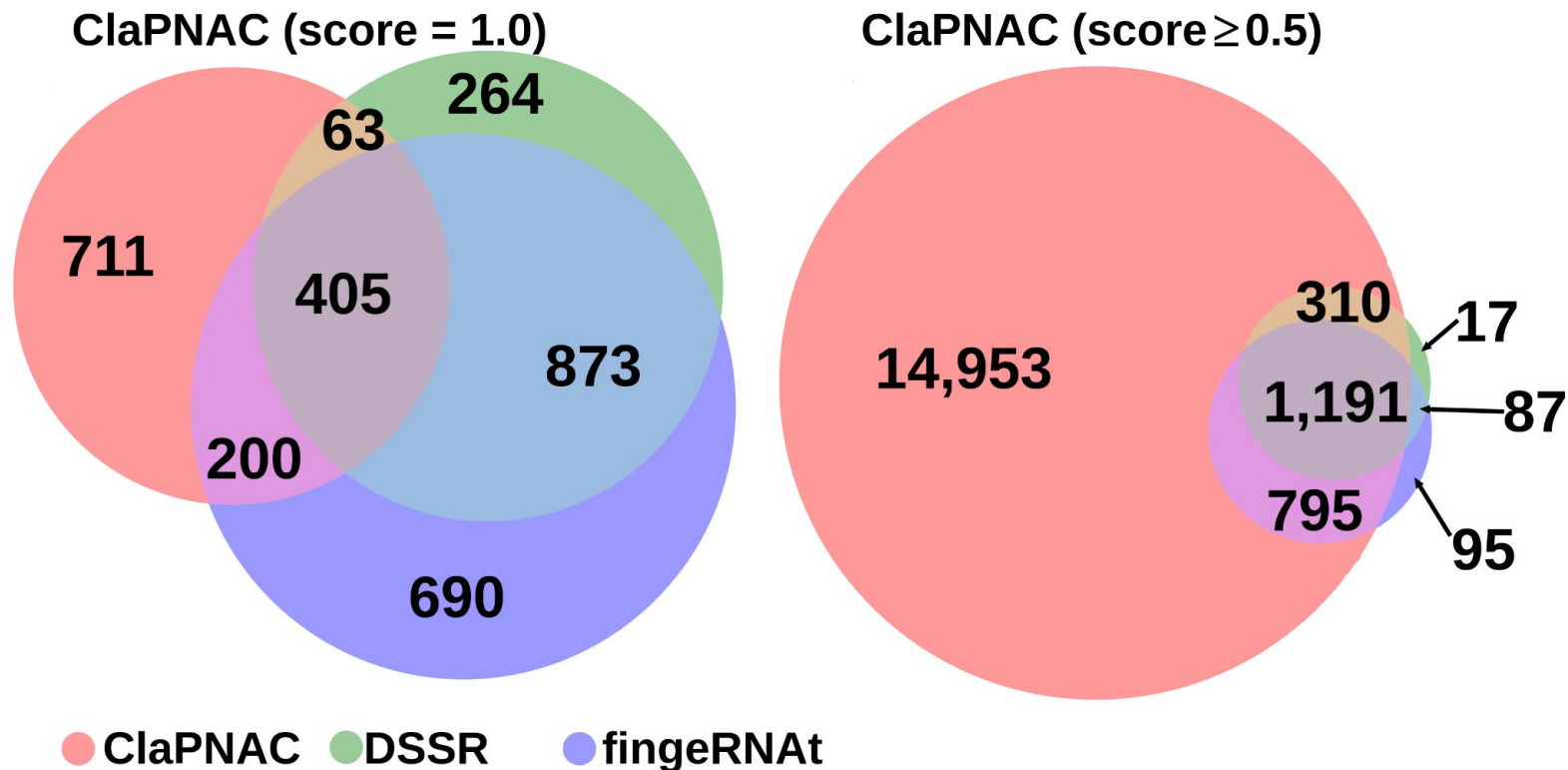


$$SF = \frac{SF^1 + SF^2}{2}$$

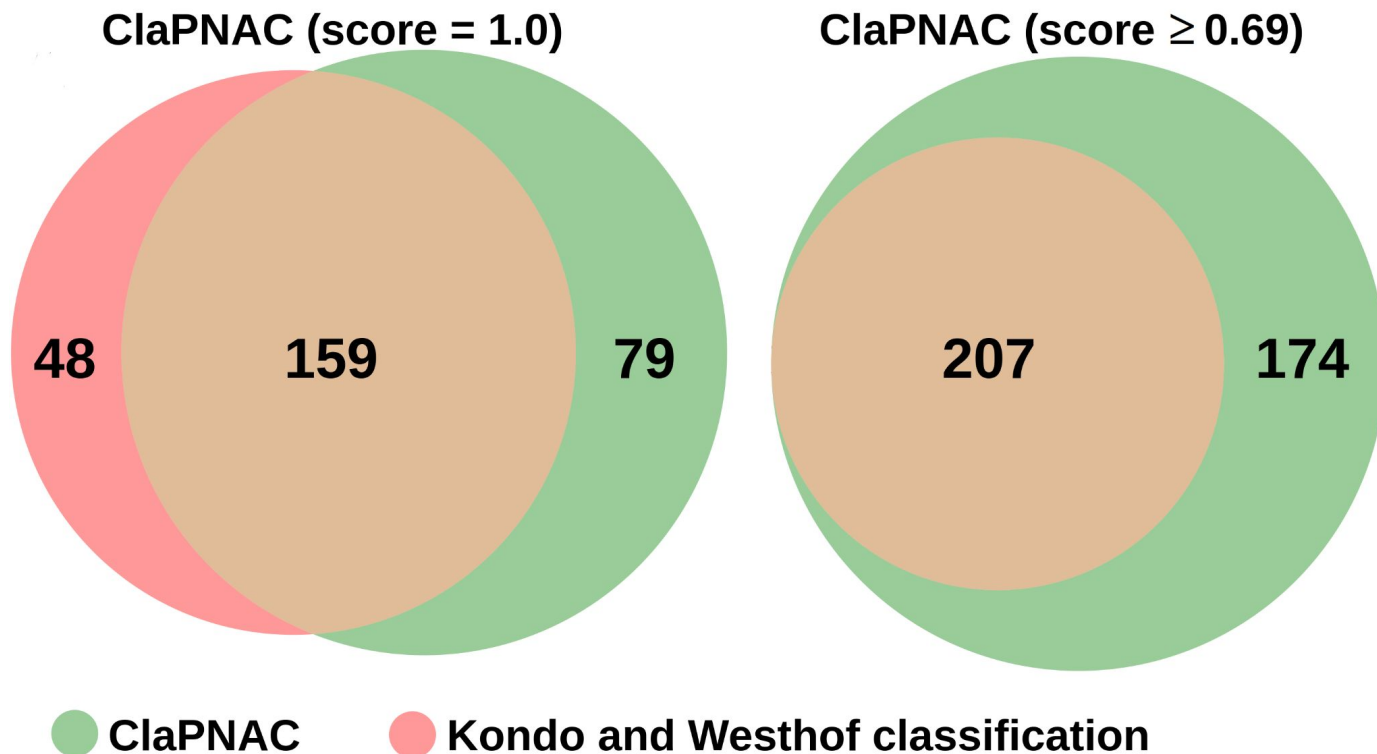
$$SF^y_{y \in 1,2} = \frac{\min_{i,j \in 1..n} (d^y_{i,j})}{\min_{i \in 1..n} (d^y_{i,n+1}) + \min_{i,j \in 1..n+1} (d^y_{i,j})}$$

ClaPNAC vs fingerRNAat and DSSR

Test dataset:
100 randomly selected complexes
from 2,164 RNPs



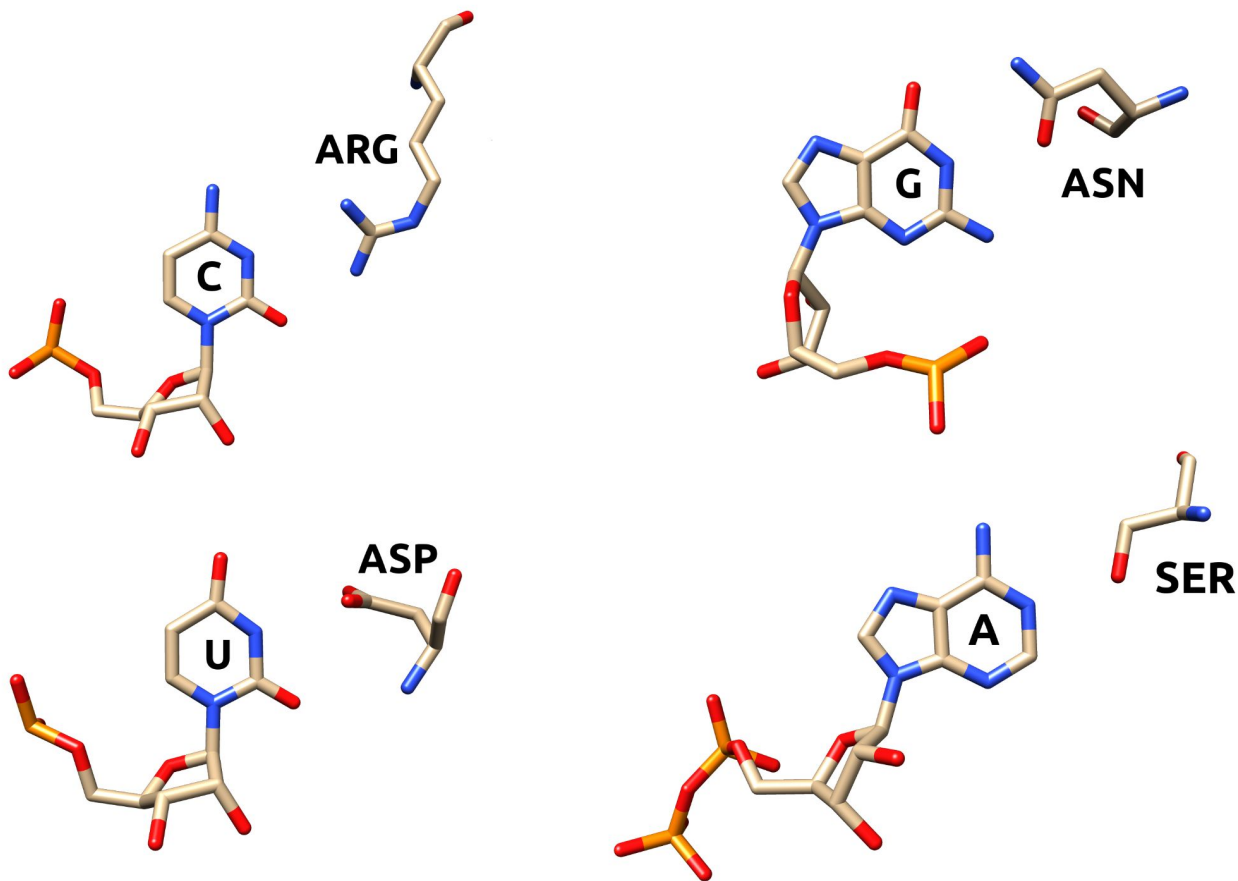
ClaPNAC vs Kondo & Westhof dataset



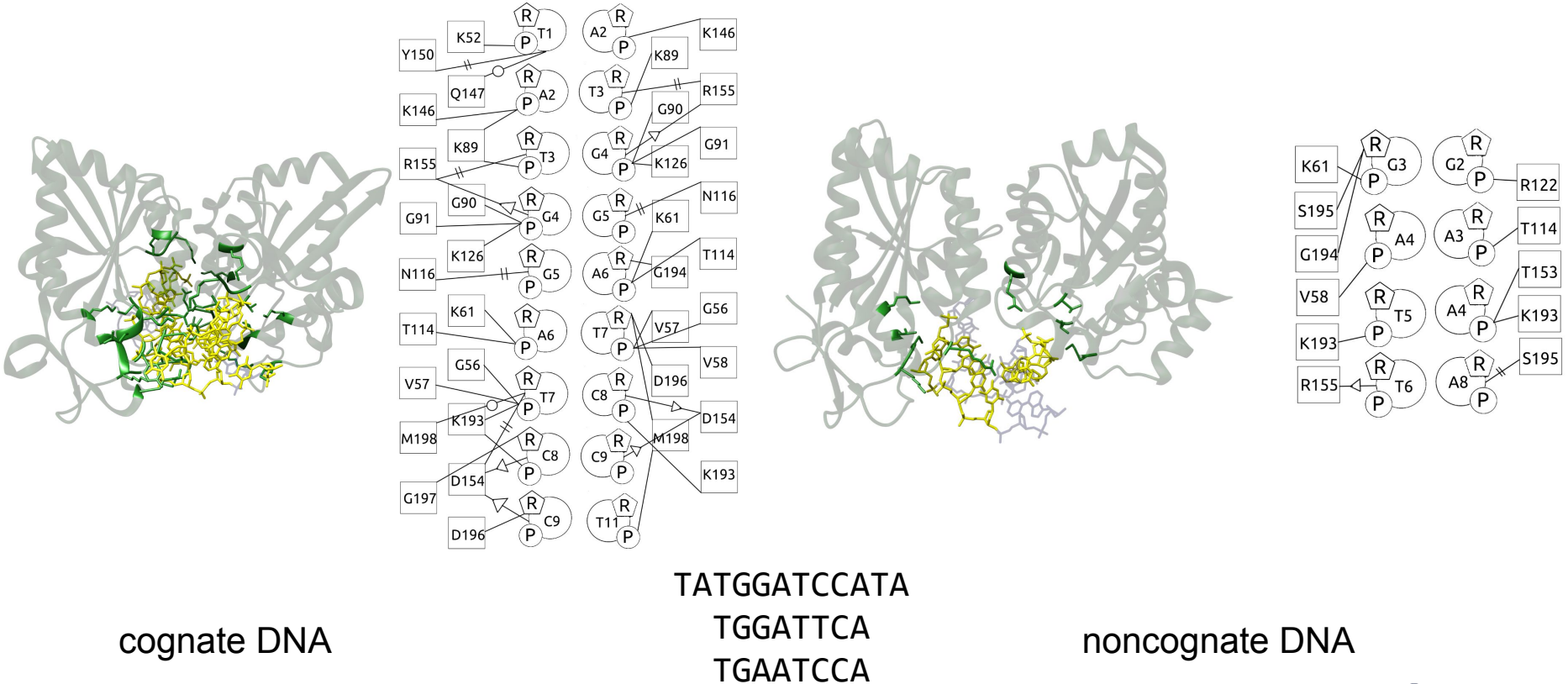
Kondo & Westhof dataset

428 RNP complexes not used for ClaPNAC training

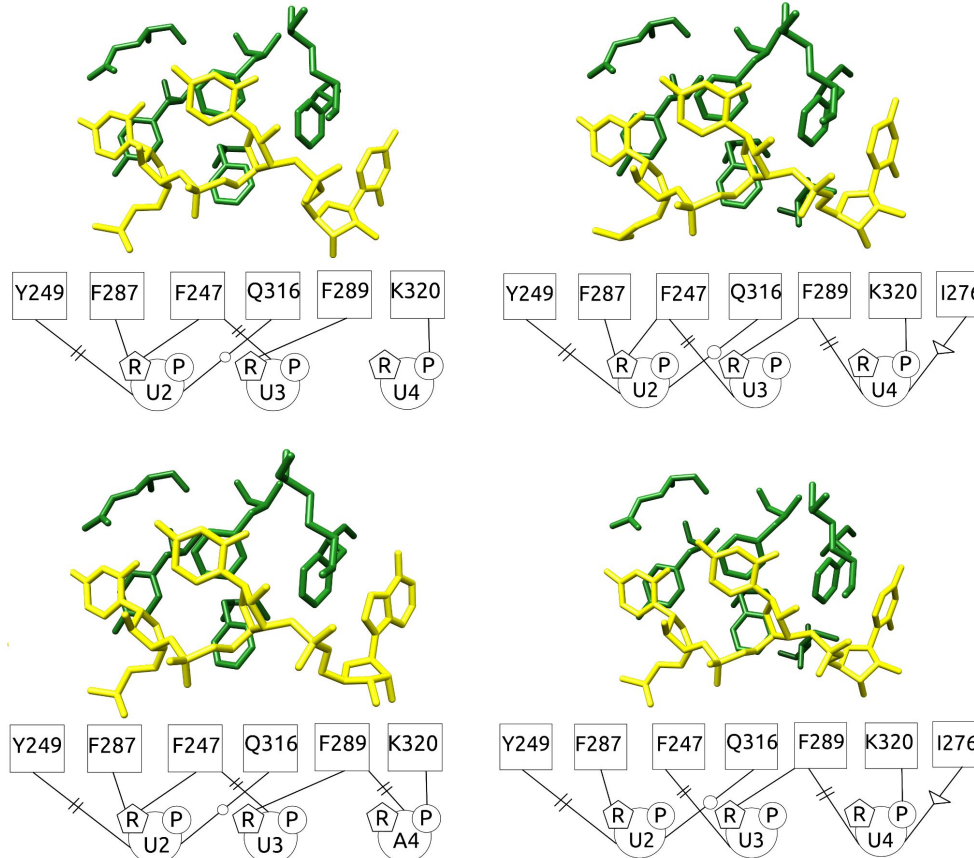
Cases not identified by Kondo & Westhof



Analysis of restriction endonuclease BamHI complex with DNA

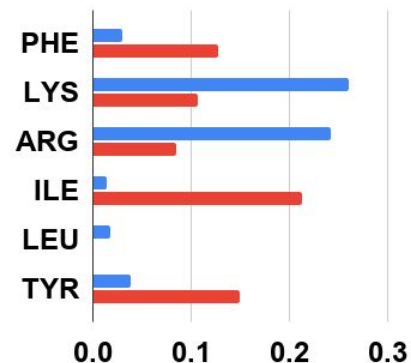
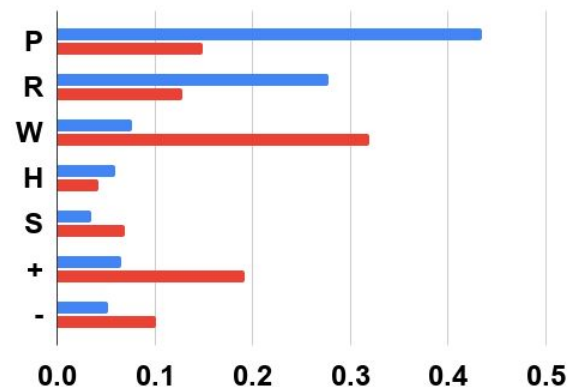
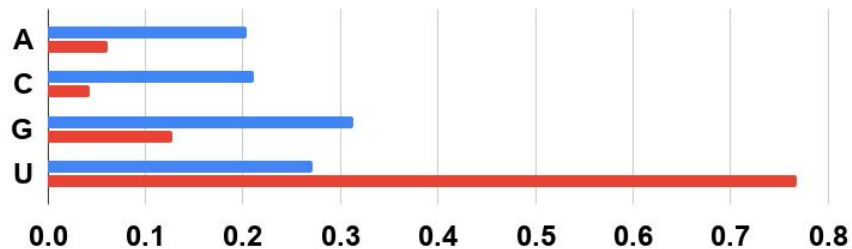
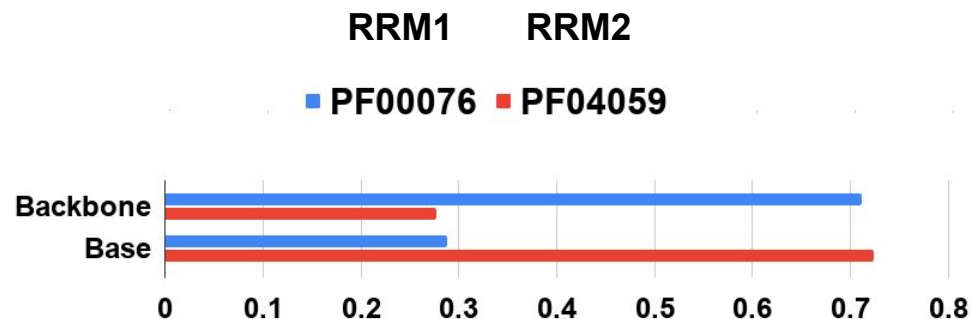


Analysis of the AU-rich element recognition and dimerization by the HuR C-terminal RRM domain



ClaPNAC is able to identify (A/U)UU(A/U) recognition motifs for RRM3

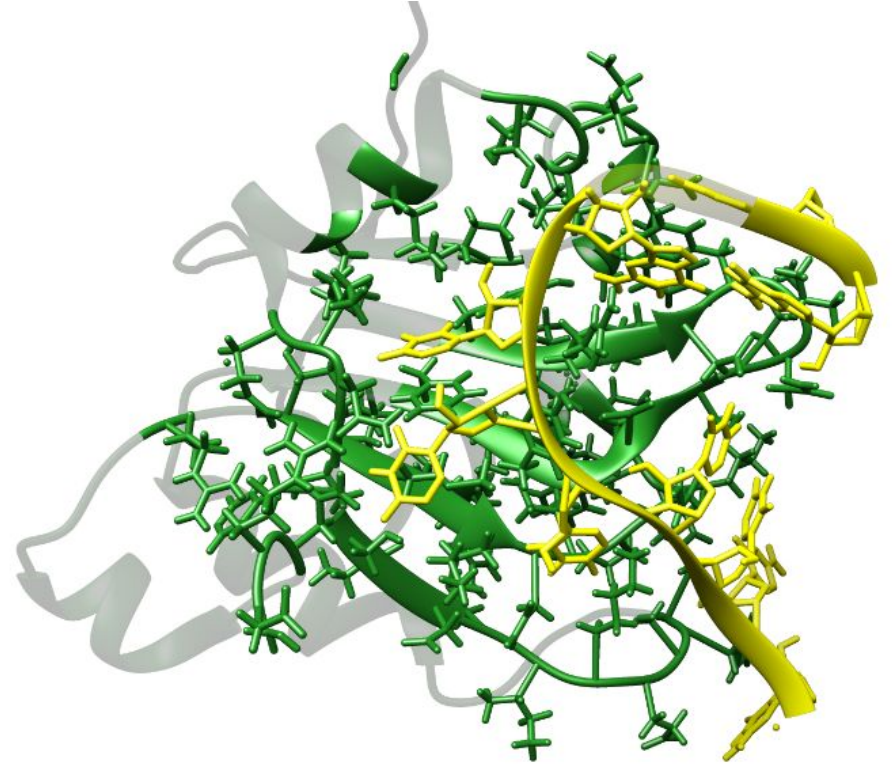
Analysis of Pfam families



Analysis of Pfam families

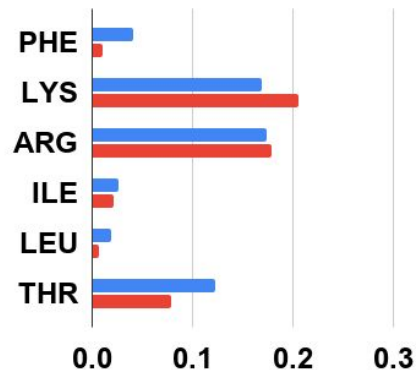
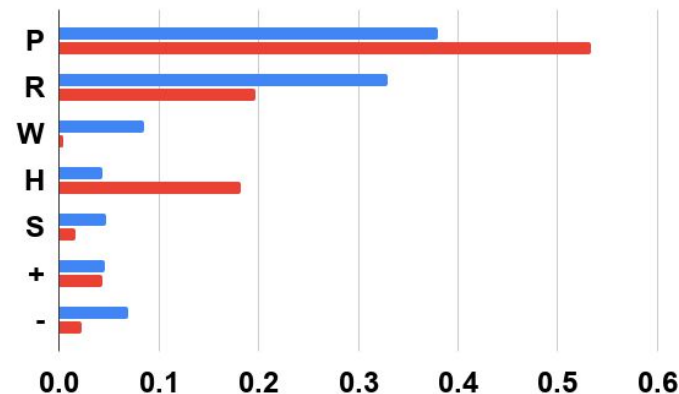
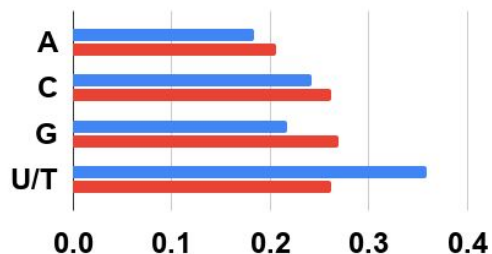
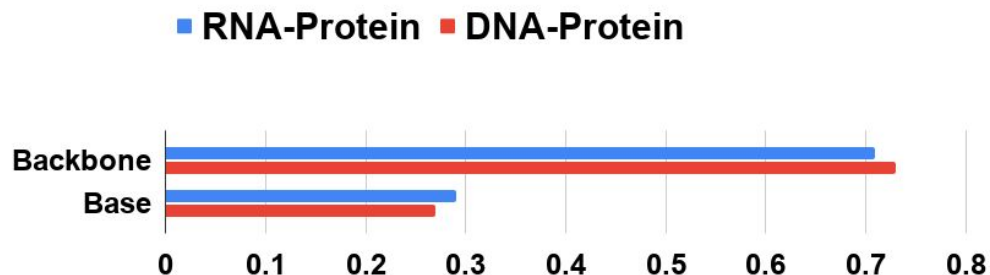


RRM1



RRM2

Comparative analysis of RNA-protein and DNA-protein complexes



Summary

- 1) ClaPNAC: novel method for contact classification in NA-protein complexes
- 2) Uses geometry (like ClaRNA) instead of hydrogen bonding etc.
- 3) Distinguishes nucleotide edges, ribose, and phosphate otide side
- 4) Distinguishes backbone and sidechain on the amino acid side
- 5) Can detect previously unclassified aa-nt pairs



Acknowledgements



Janusz M Bujnicki

Eugene Baulin

Satyabrata Maiti

Filip Stefaniak

and all the members of the
Bujnicki lab at IIMCB in
Warsaw