

# Design of oligonucleotides as RNA-based drugs

**prerequisites and opportunities for *in silico* methods within  
the framework of fragment-based approaches**

# Outline

- **Introduction**
  - RNA-based drugs
  - Fragment-based approaches
  - Proof of concept for RNA molecules: pros & cons
- **Preliminary results and lessons learned**
  - 3 examples of RNA-binding proteins (RBPs)
- **Benchmarking for single nucleotide fragments**
  - 120 nucleotide-protein complexes
- **Conclusions**

# RNA or nucleotide based drugs

## Modes of action

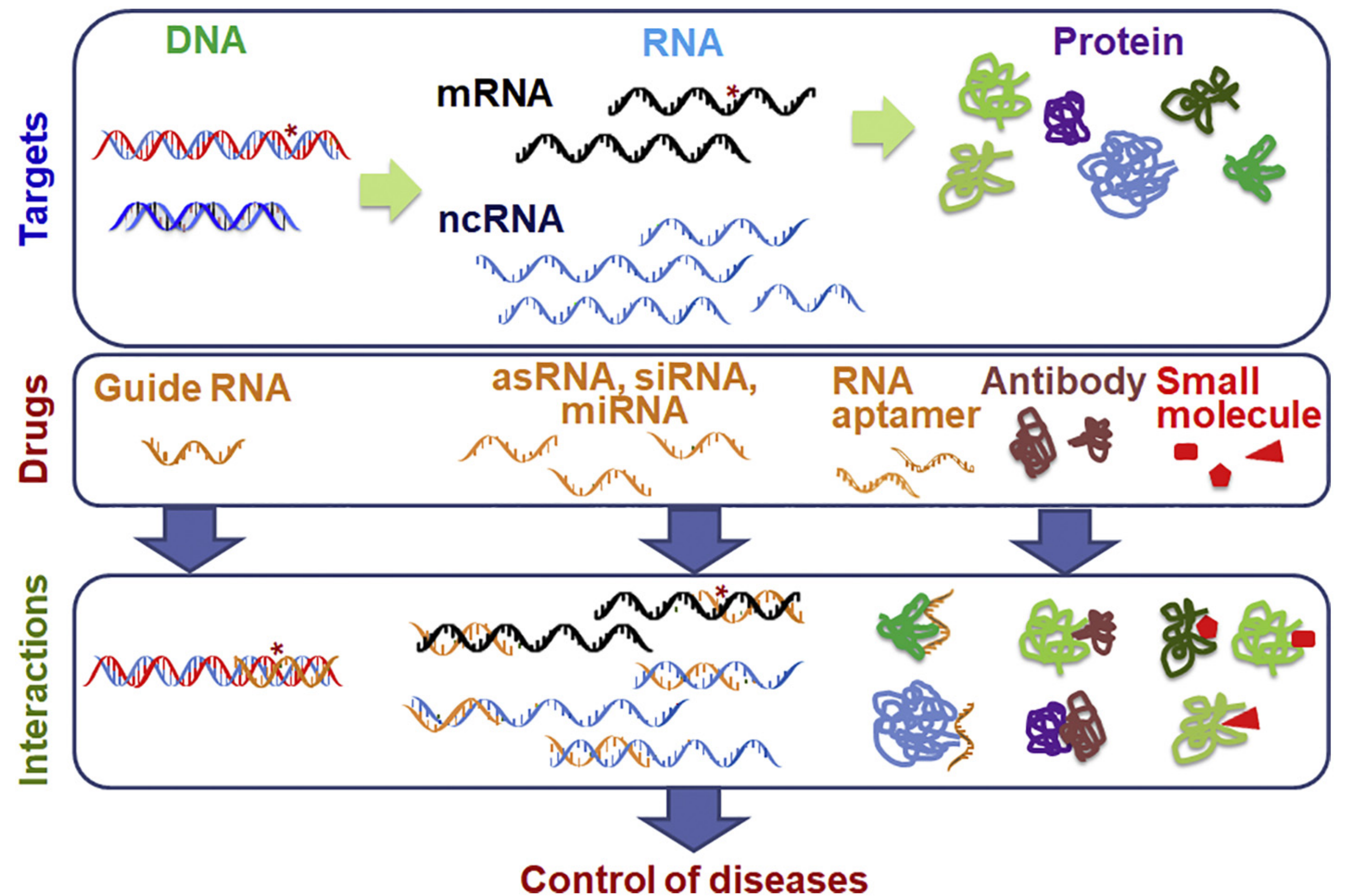
- **hybridisation-dependent (RNA target)**

- siRNA (small interfering RNA)
- miR (microRNA)
- ASO (Antisense Oligonucleotides)

- ...

- **hybridisation-independent (targets)**

- aptamers (protein targets, ...)
- immunostimulatory (CpG) oligos
- mRNA
- ...



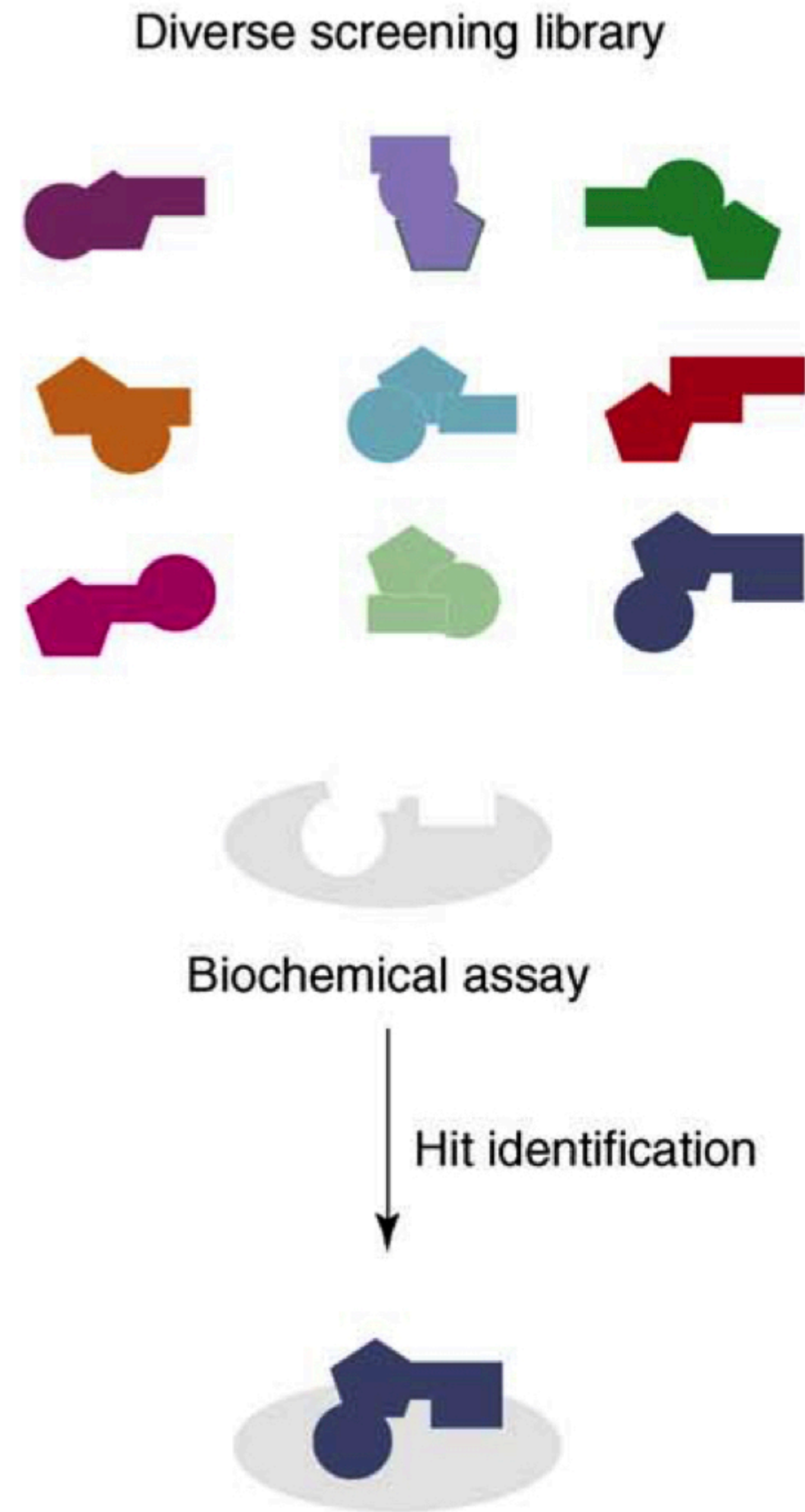






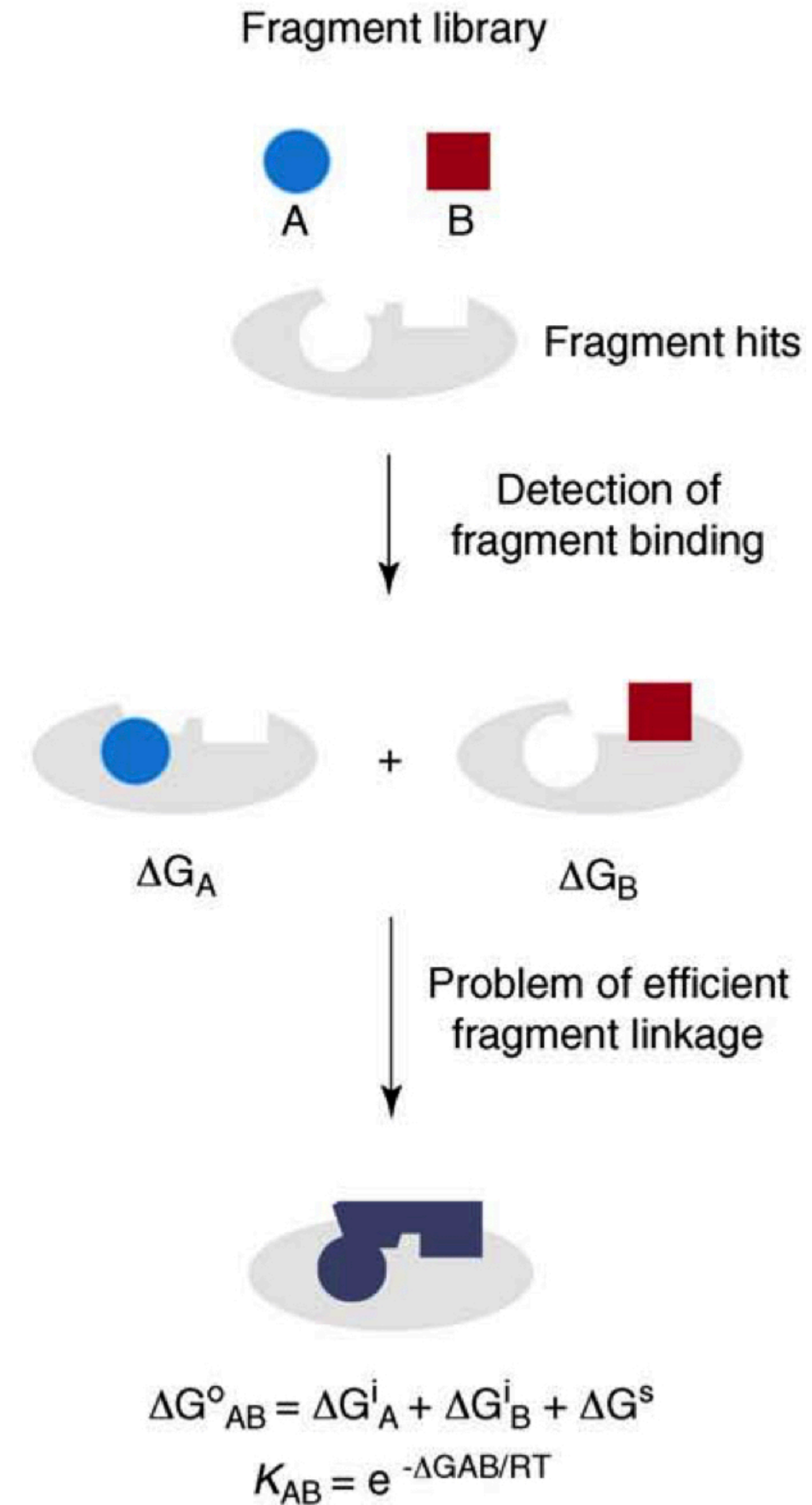
# Fragment-based design/discovery (FBD) approaches

(a) High-throughput screening



## Pros & Cons

(b) Fragment-based drug discovery



*in silico*

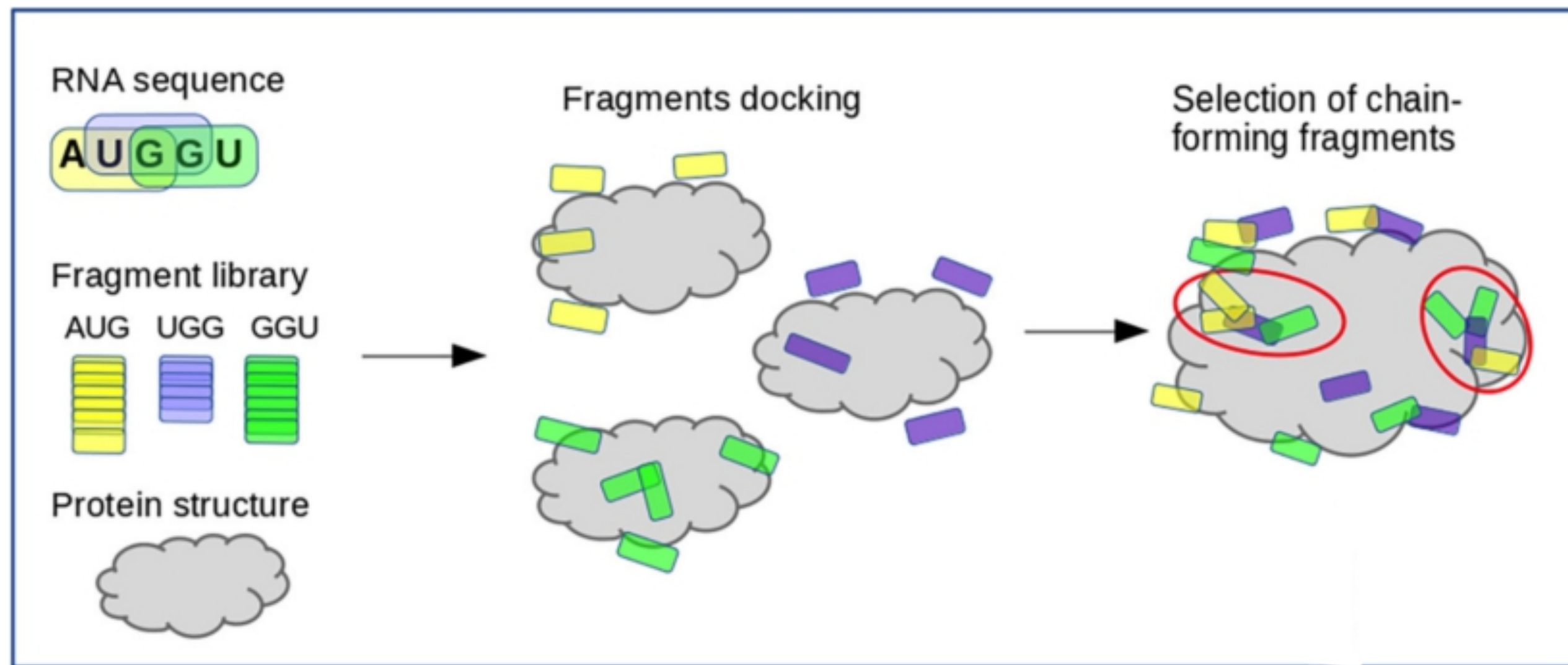
high-resolution structure

more fitted fragments

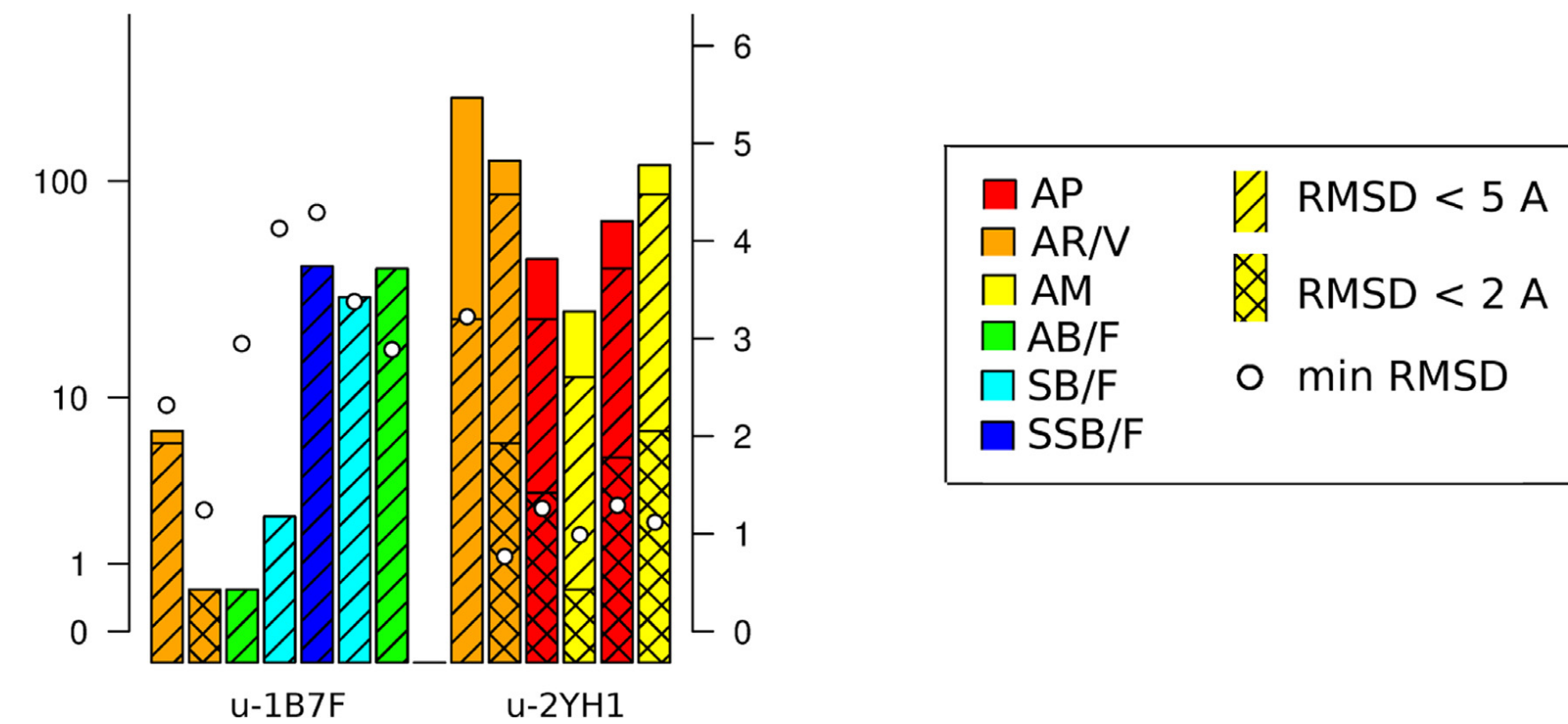
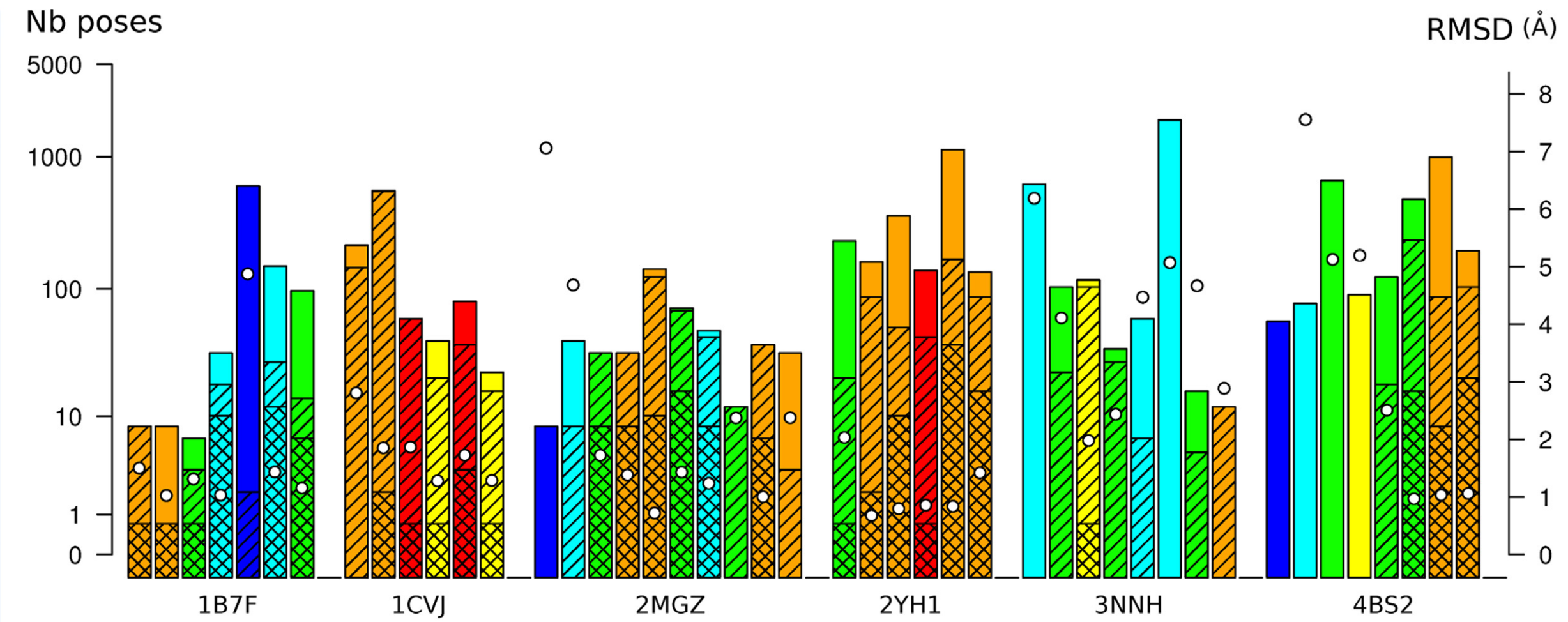
docking power (native)  
 screening power (true binders)

# FBD applied to RNA molecules

## Proof of concept



**discrete representation  
of 3-mers  
(fragment=trinucleotide)**



**Top 1000 poses**

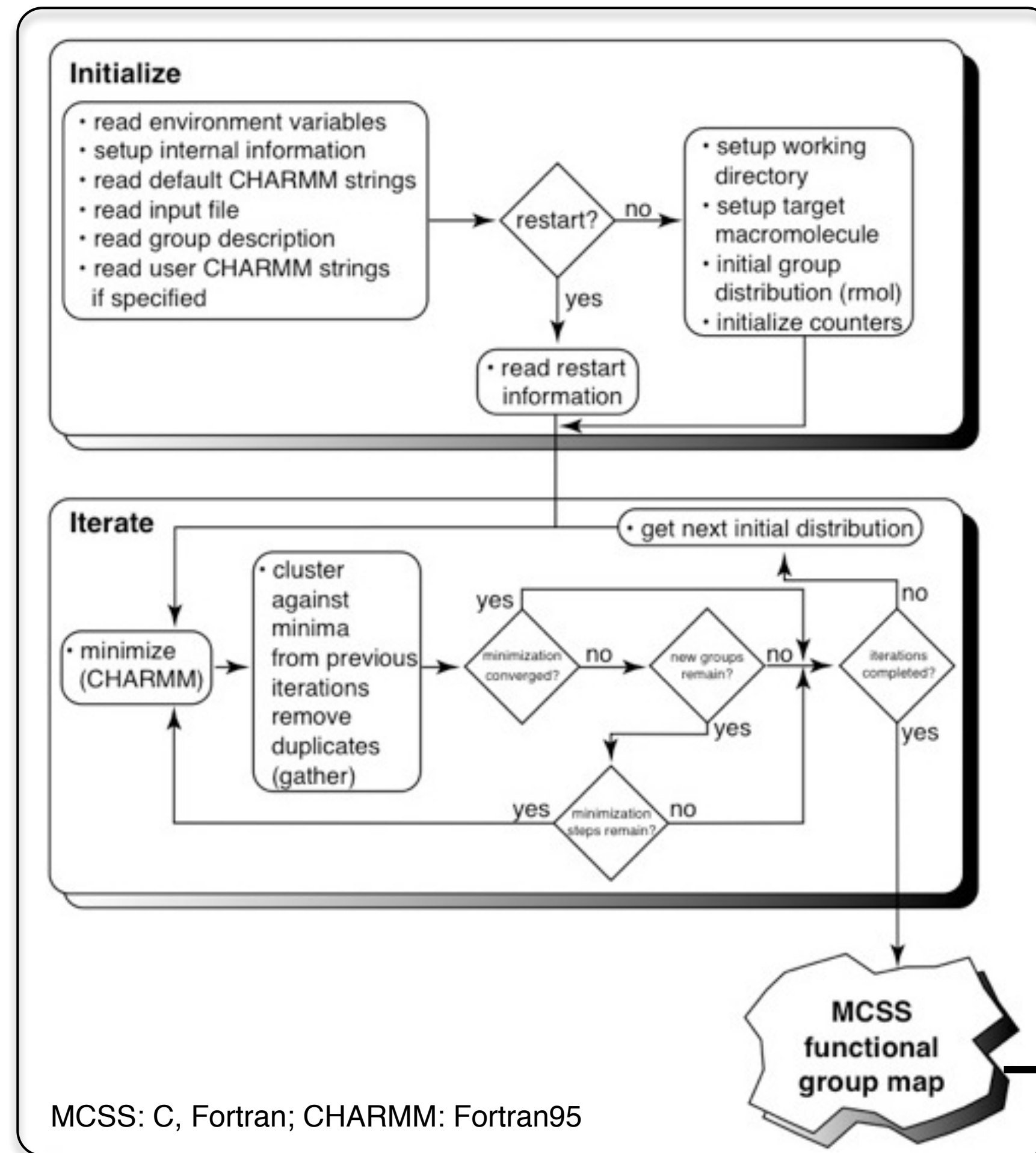
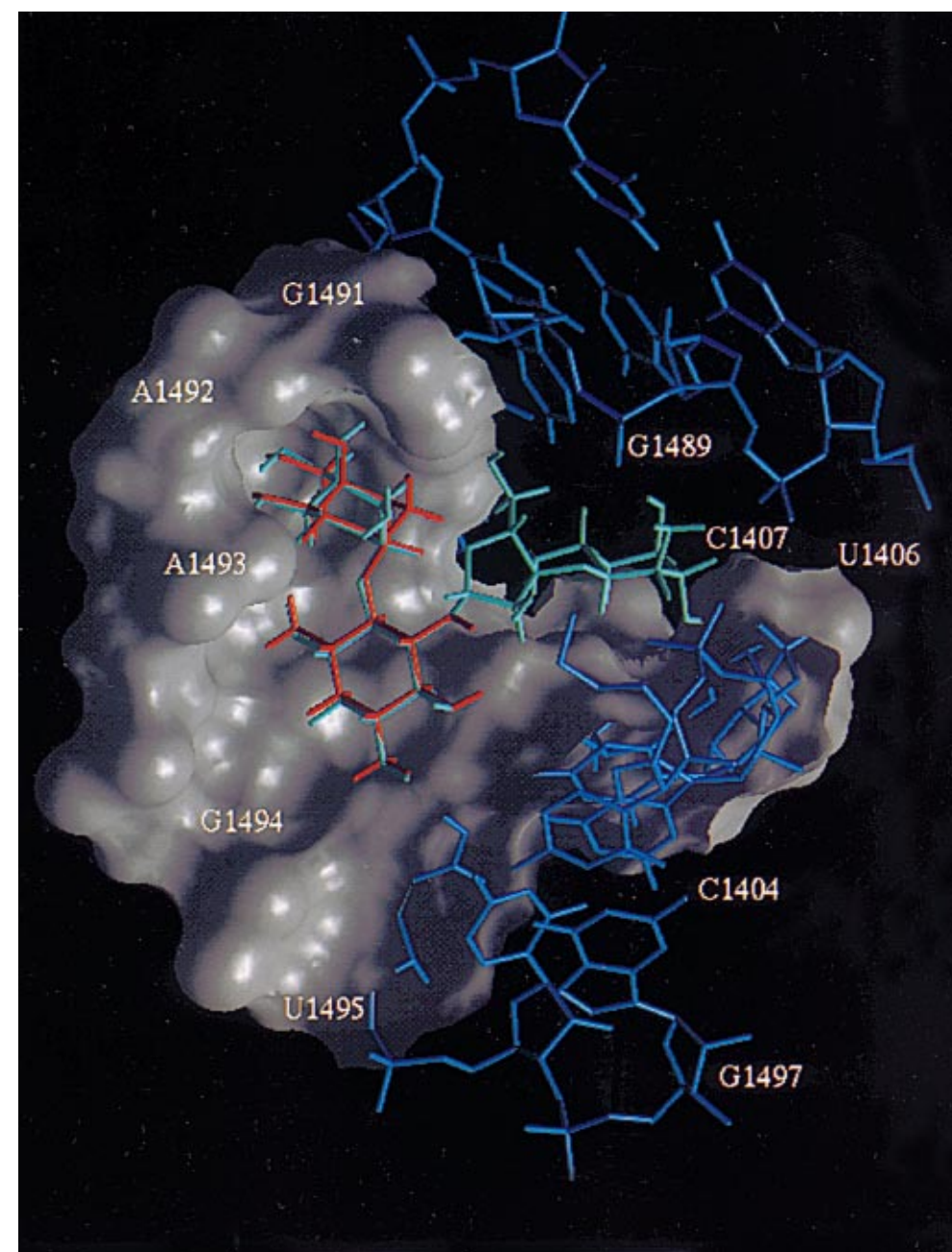


# MCSS-based Modeling of RNA Ligands

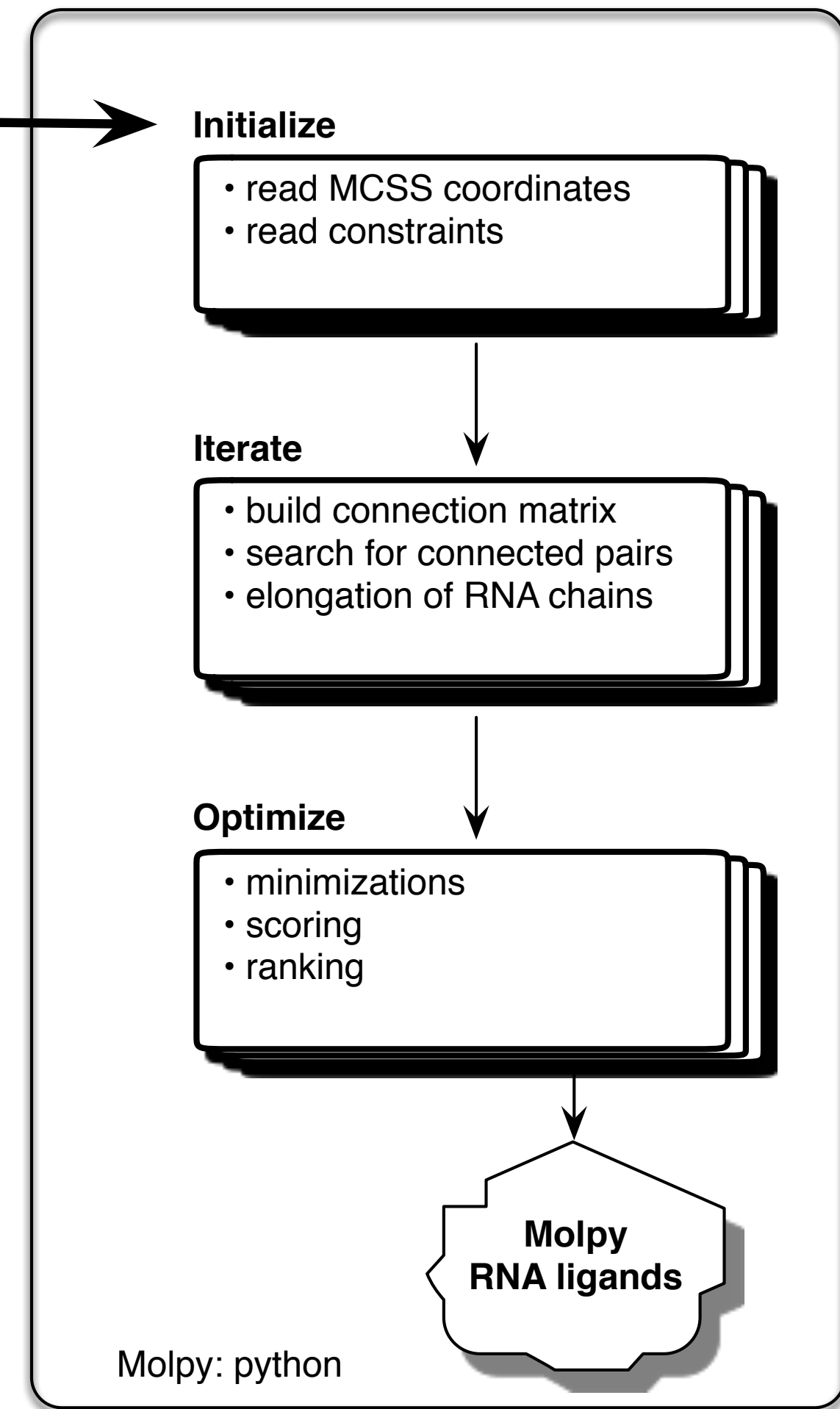
single-nucleotide fragment & forcefield based strategy

**MCSS  
(CHARMM)**

Multiple Copy Simultaneous Search



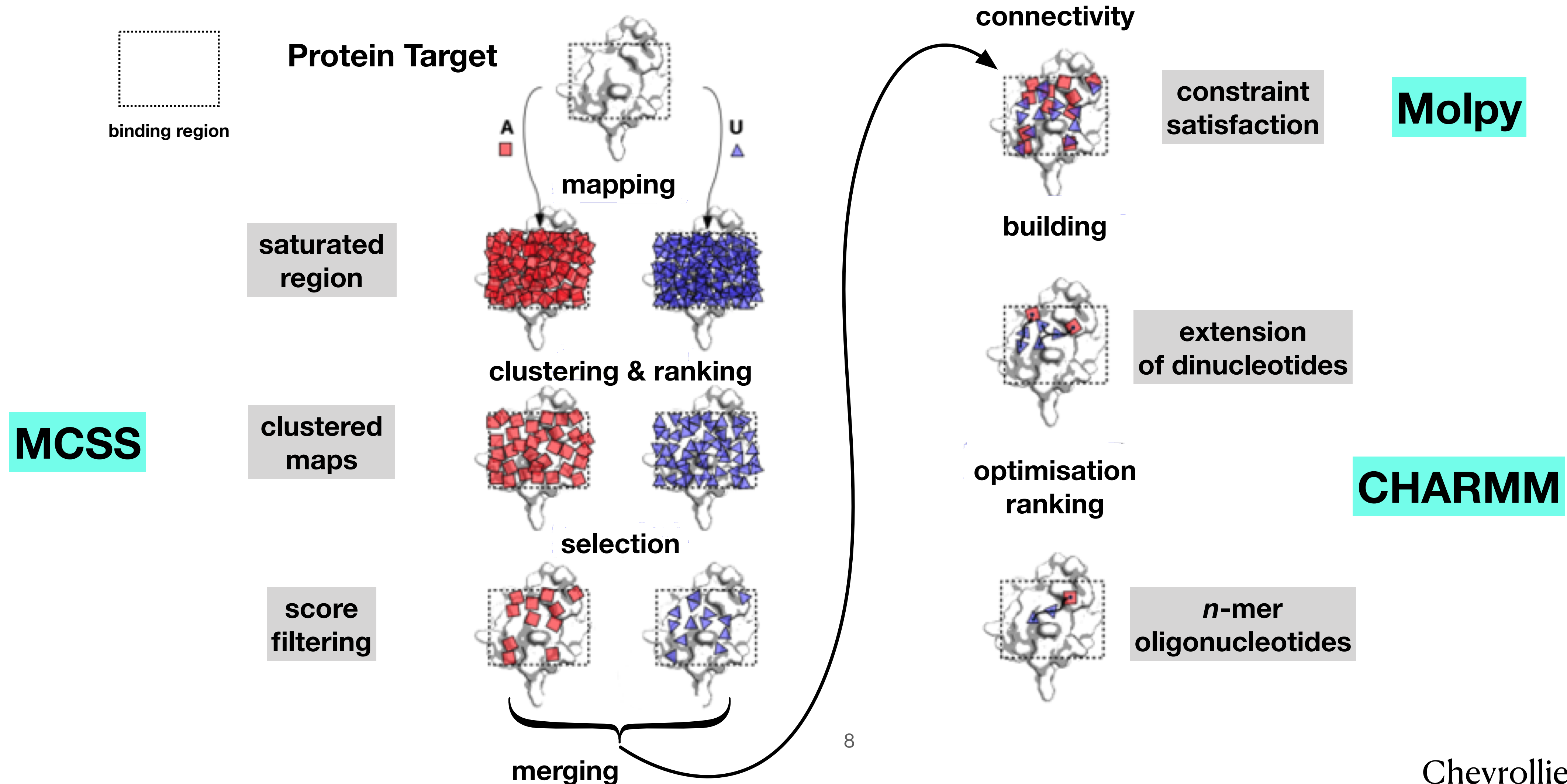
**Molpy**





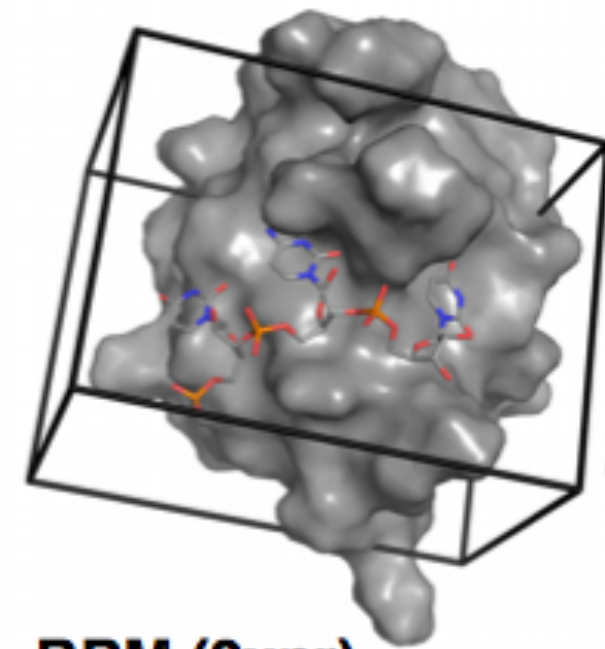
# MCSS-based workflow for FBD of oligos

Reproduce oligonucleotides binding on RBPs

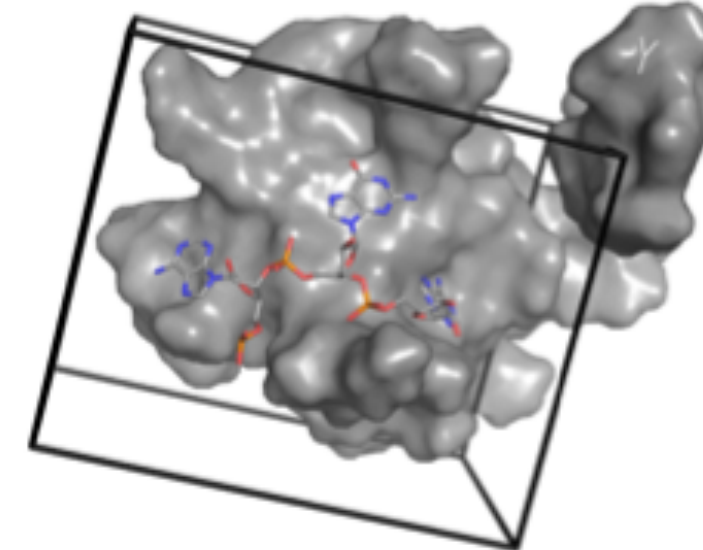


# Preliminary analysis and MCSS results on 3 RBPs

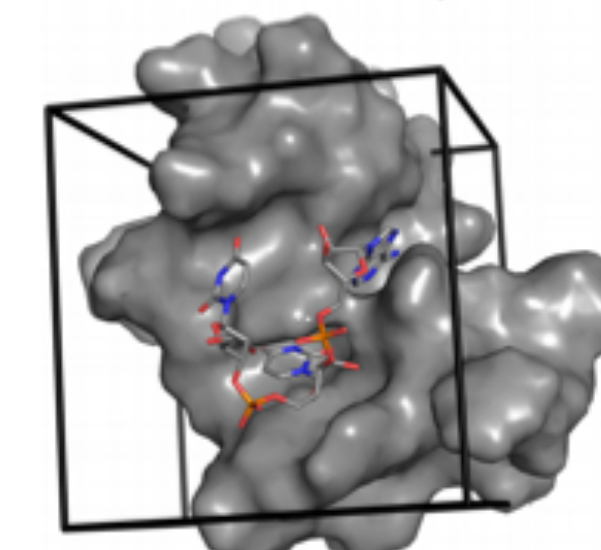
optimal/sub-optimal binding in RRM, KH and CCCH families



RRM (2xnr)  
5'-UCU-3'

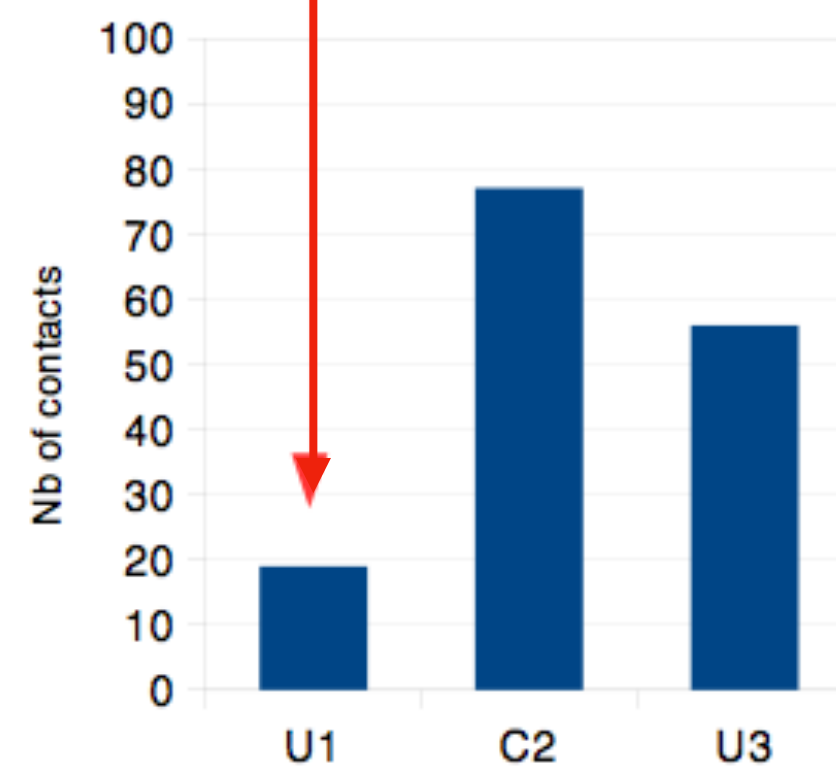


KH (5wwx)  
5'-AGA-3'

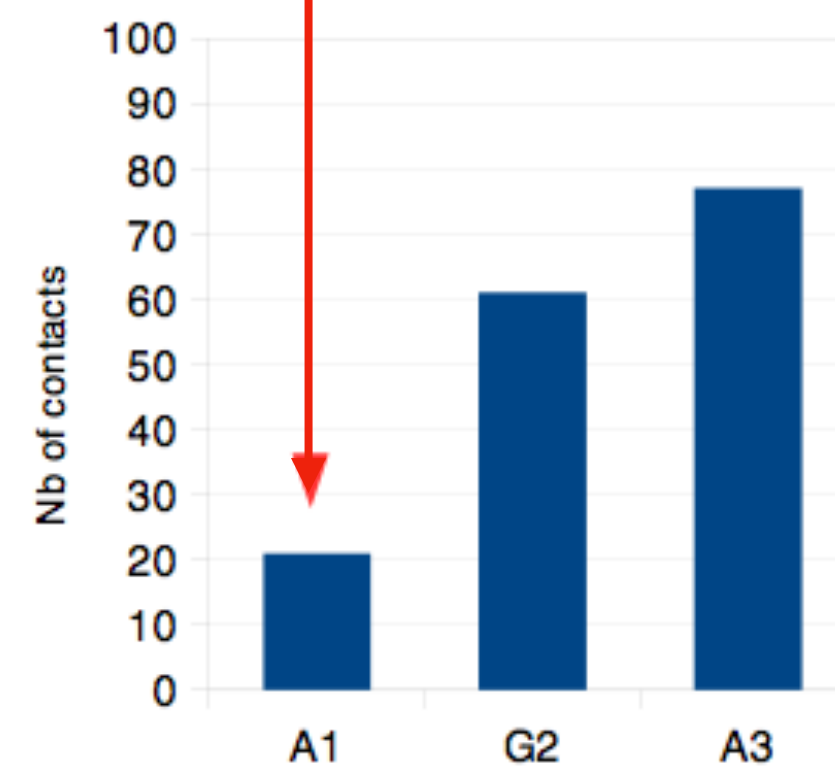


Zn-CCCH (5elh)  
5'-UUA-3'

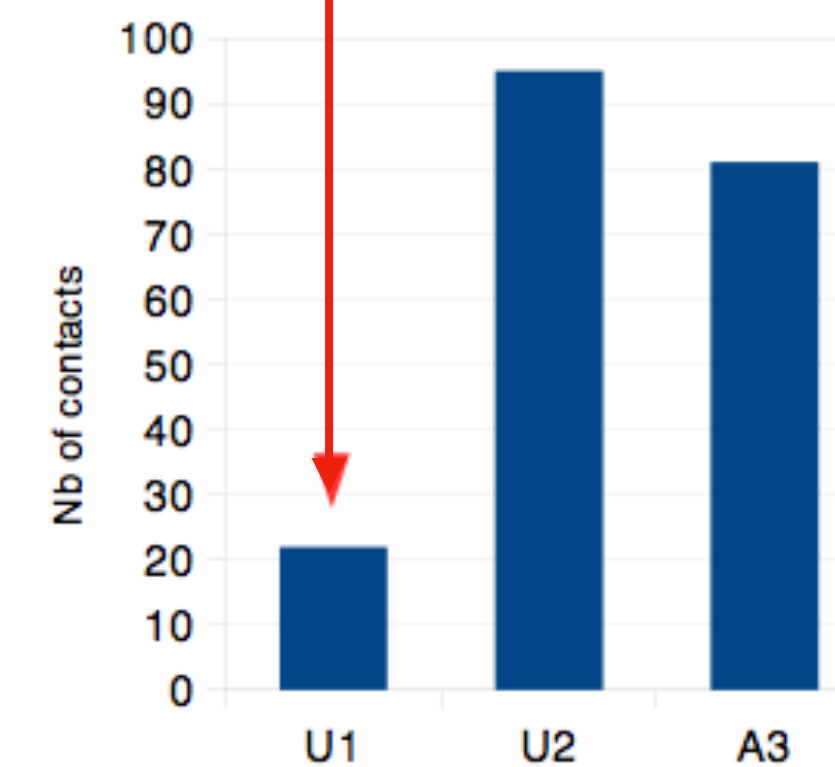
U1	C2	U3
155	1	9



A1	G2	A3
691	7	58



U1	U2	A3
490	3	1



**BINANA**



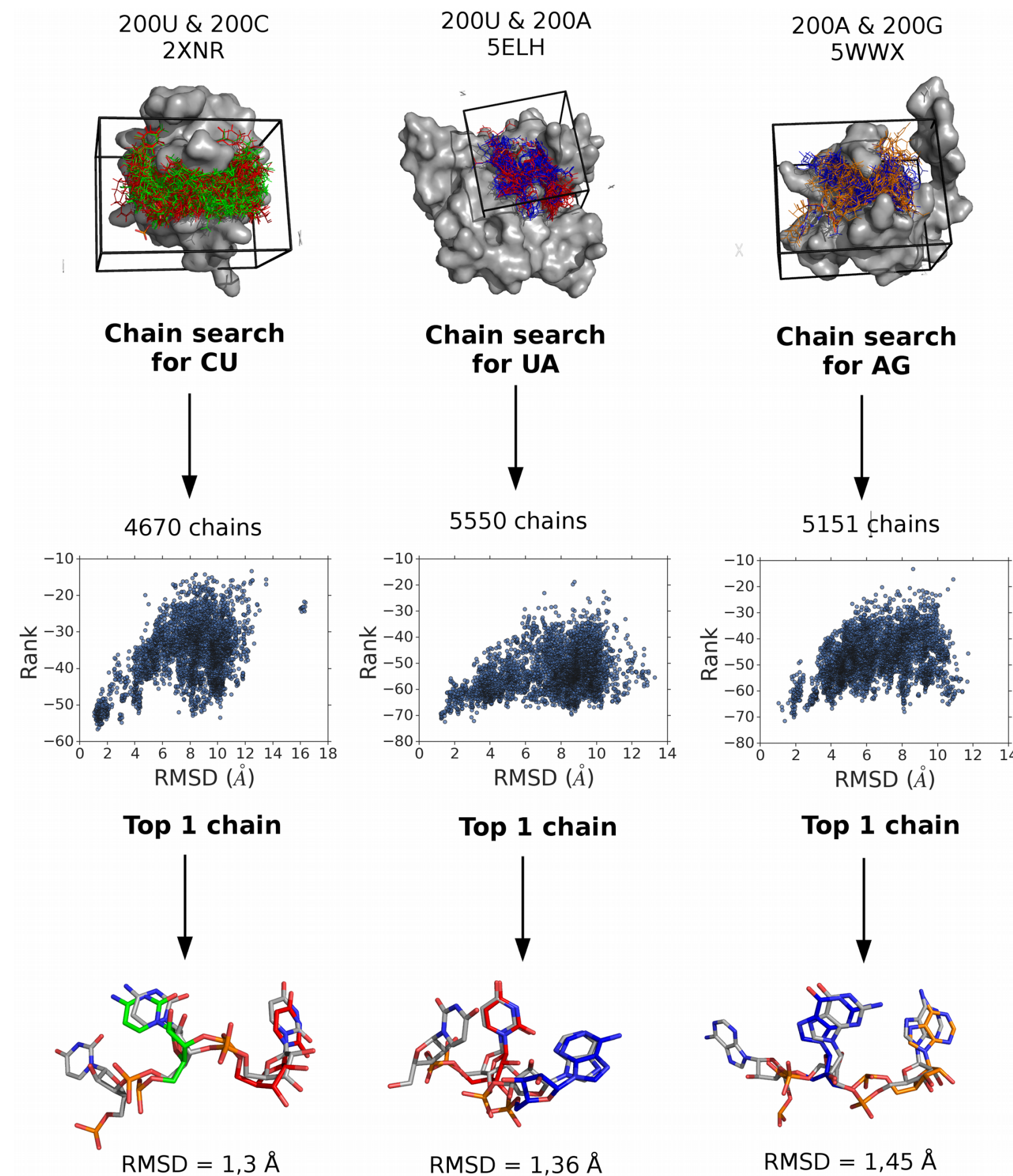
# Preliminary MCSS+Molpy results on 3 RBPs

## 2-mers/3-mers oligonucleotides

200 initial poses

~5000 dinucleotides

Top 1 2-mers



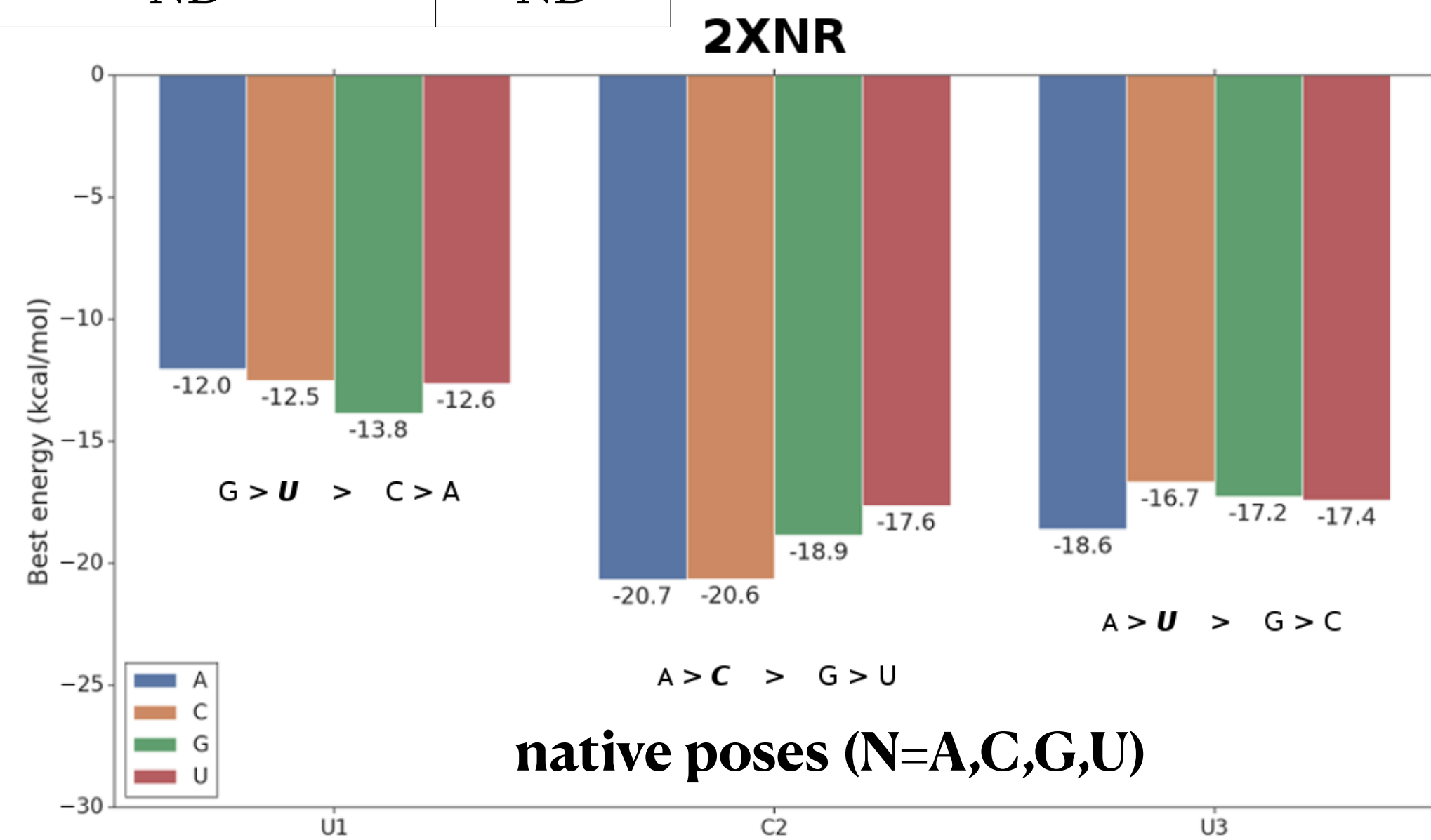
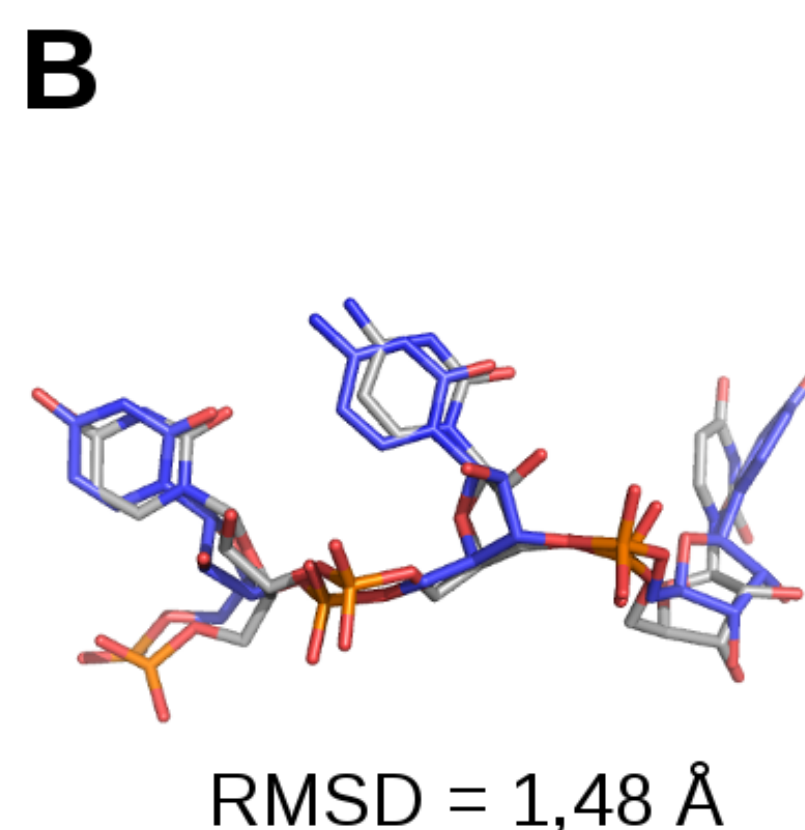
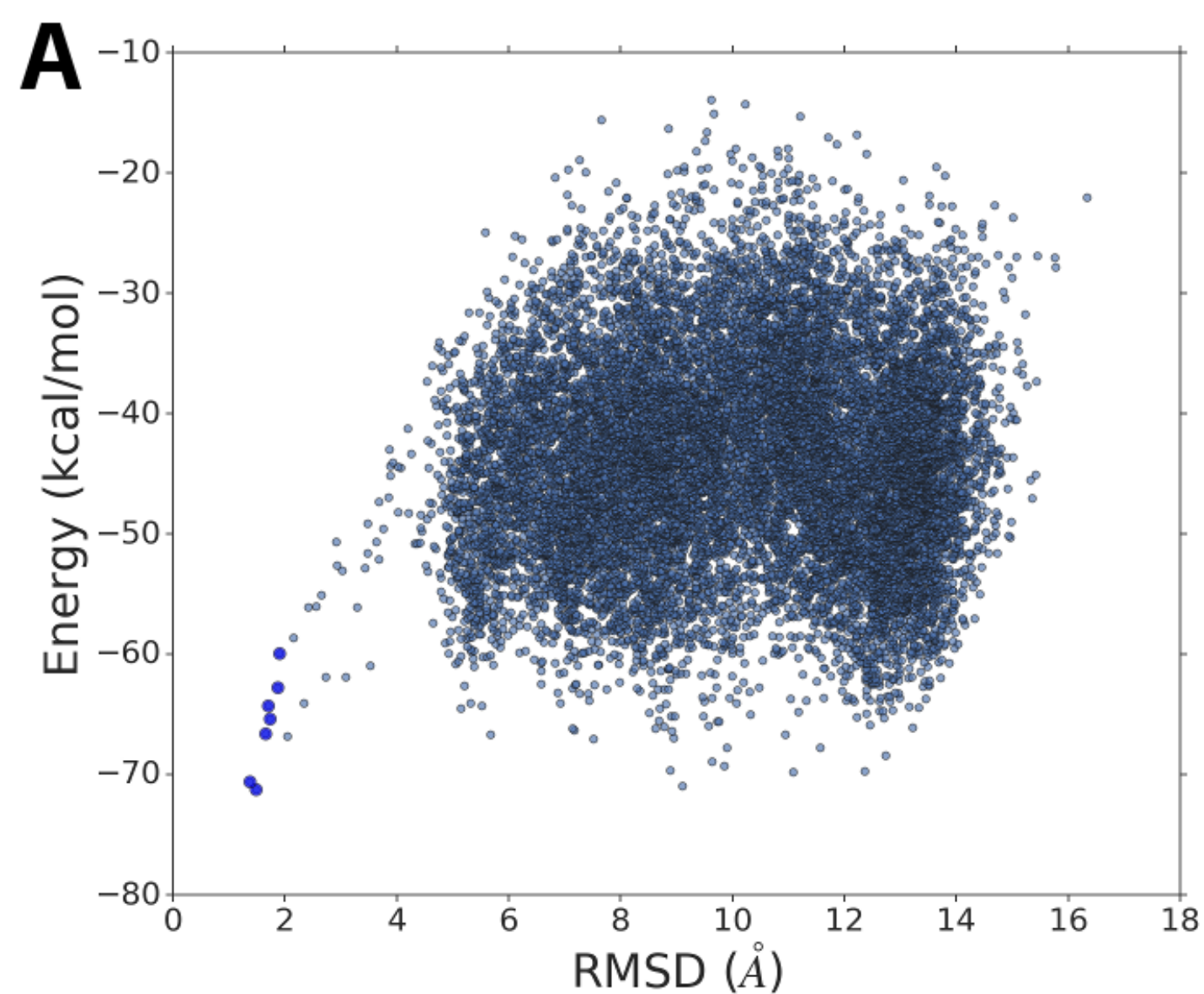
MCSS  
+  
Molpy  
+  
CHARMM



# Preliminary MCSS+Molpy results on 3 RBPs

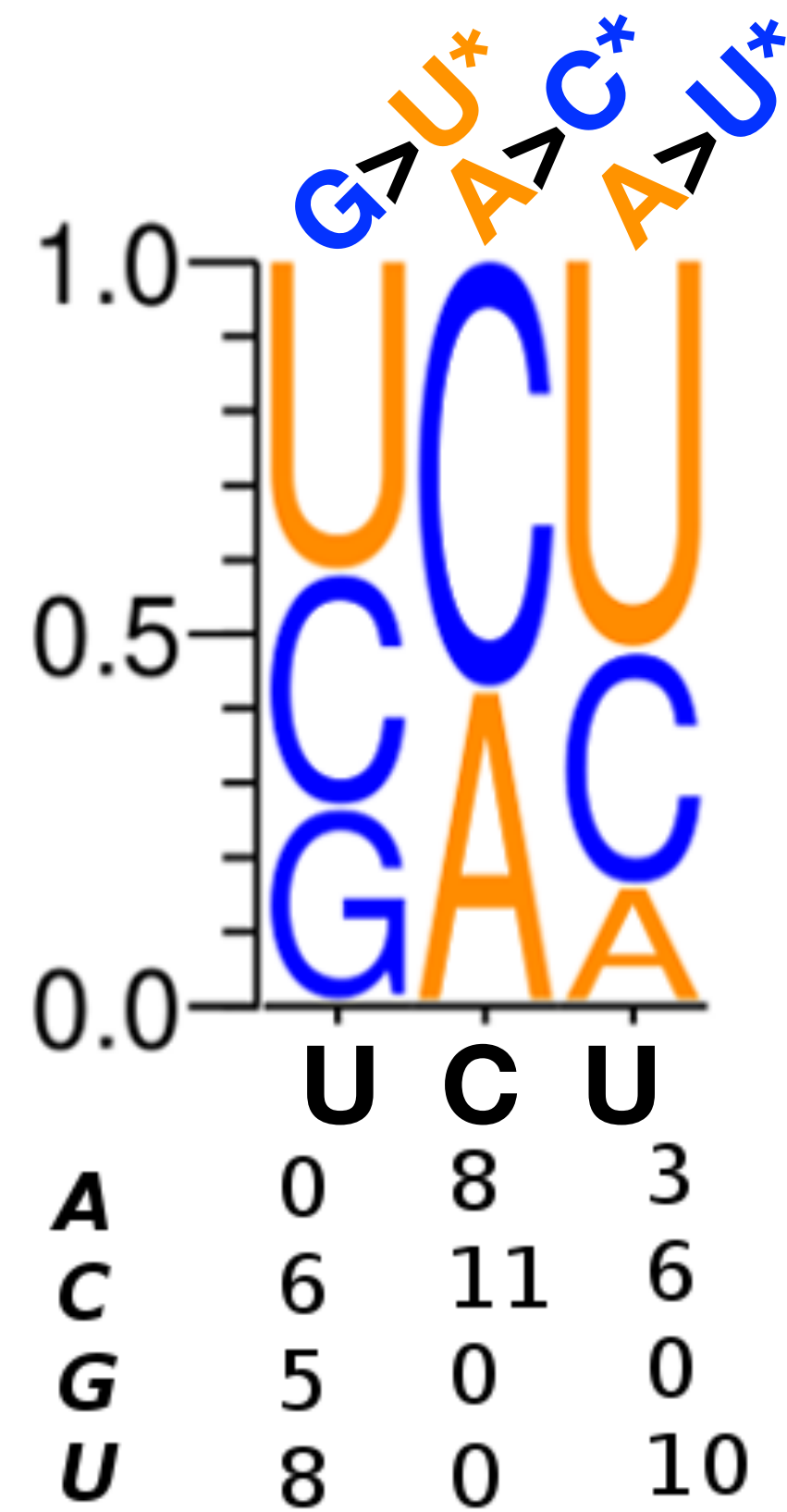
## 3-mers oligonucleotides & nucleotide-type screening

PDB ID	selected poses				sequence constraint	generated chains	native chains ( $\leq 2$ Å)	
	A	C	G	U			rank	RMSD (Å)
2XNR	/	155	/	155	UCU	13 782	1	1,48
5ELH	490	/	/	490	UUA	1 584 619	ND	ND
5WWX	691	/	691	/	AGA	2 433 270	ND	ND

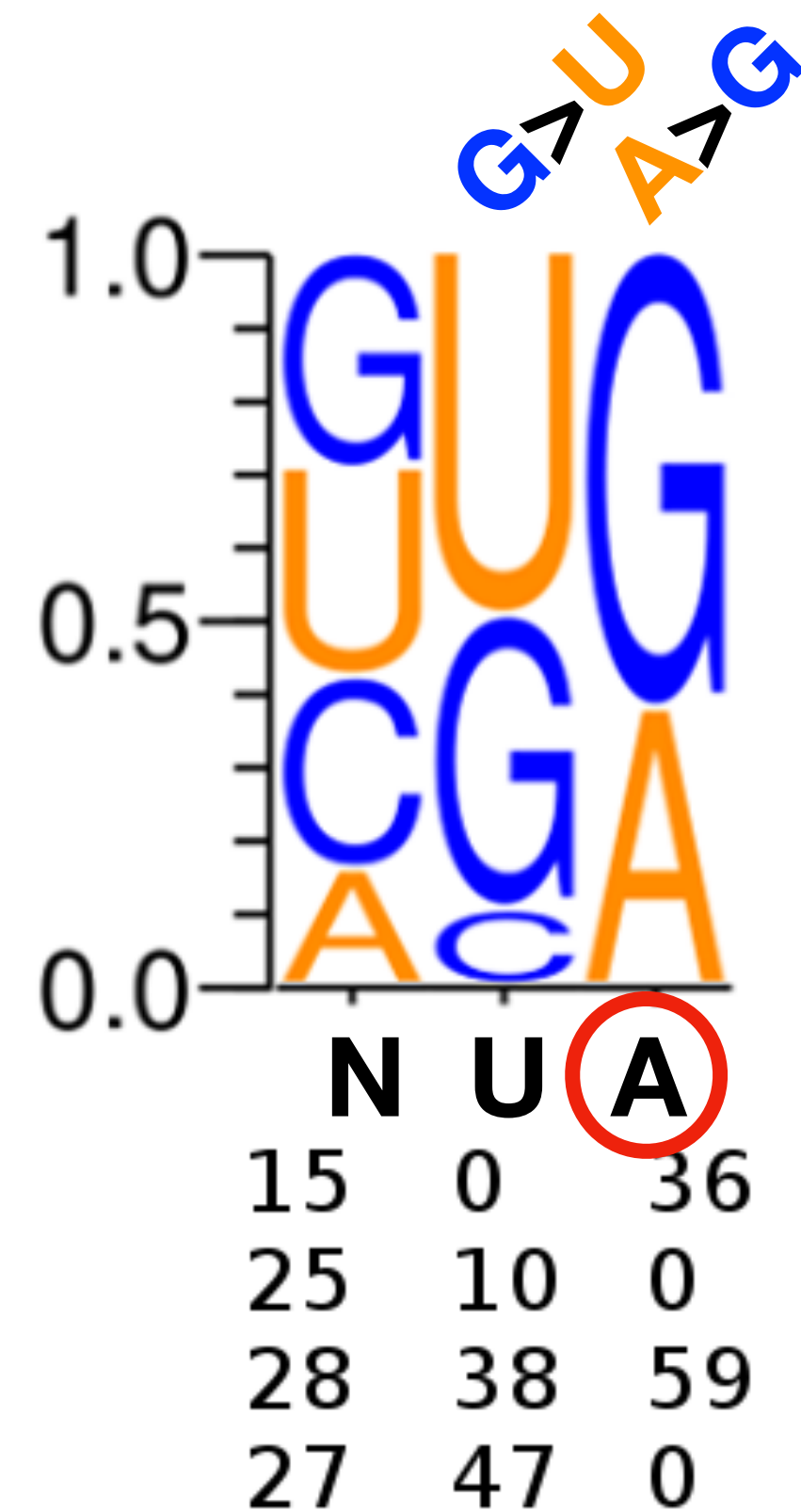


# Preliminary MCSS+Molpy results on 3 RBPs

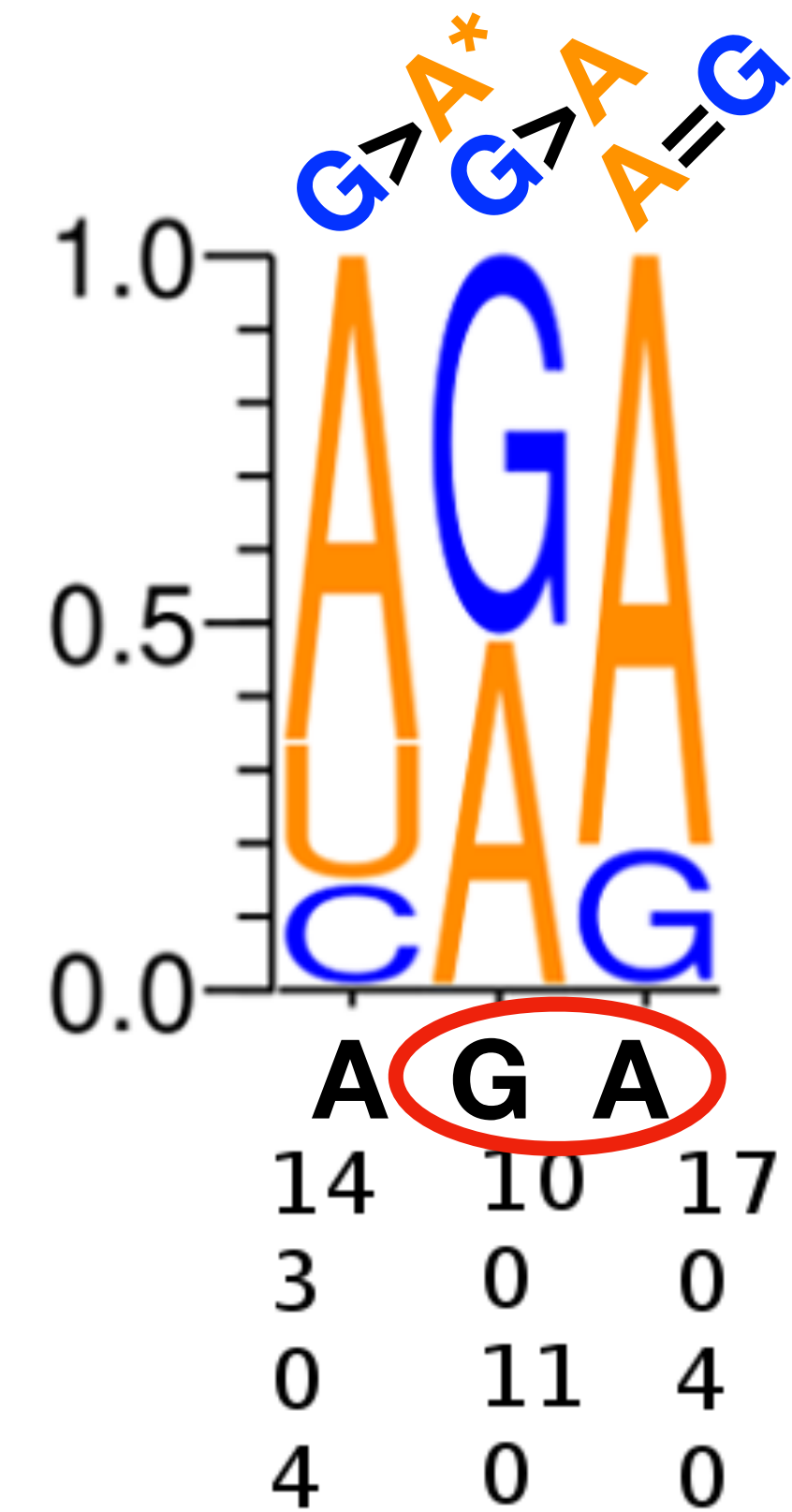
sequence-free predictions



2XNR



5ELH



5WWX

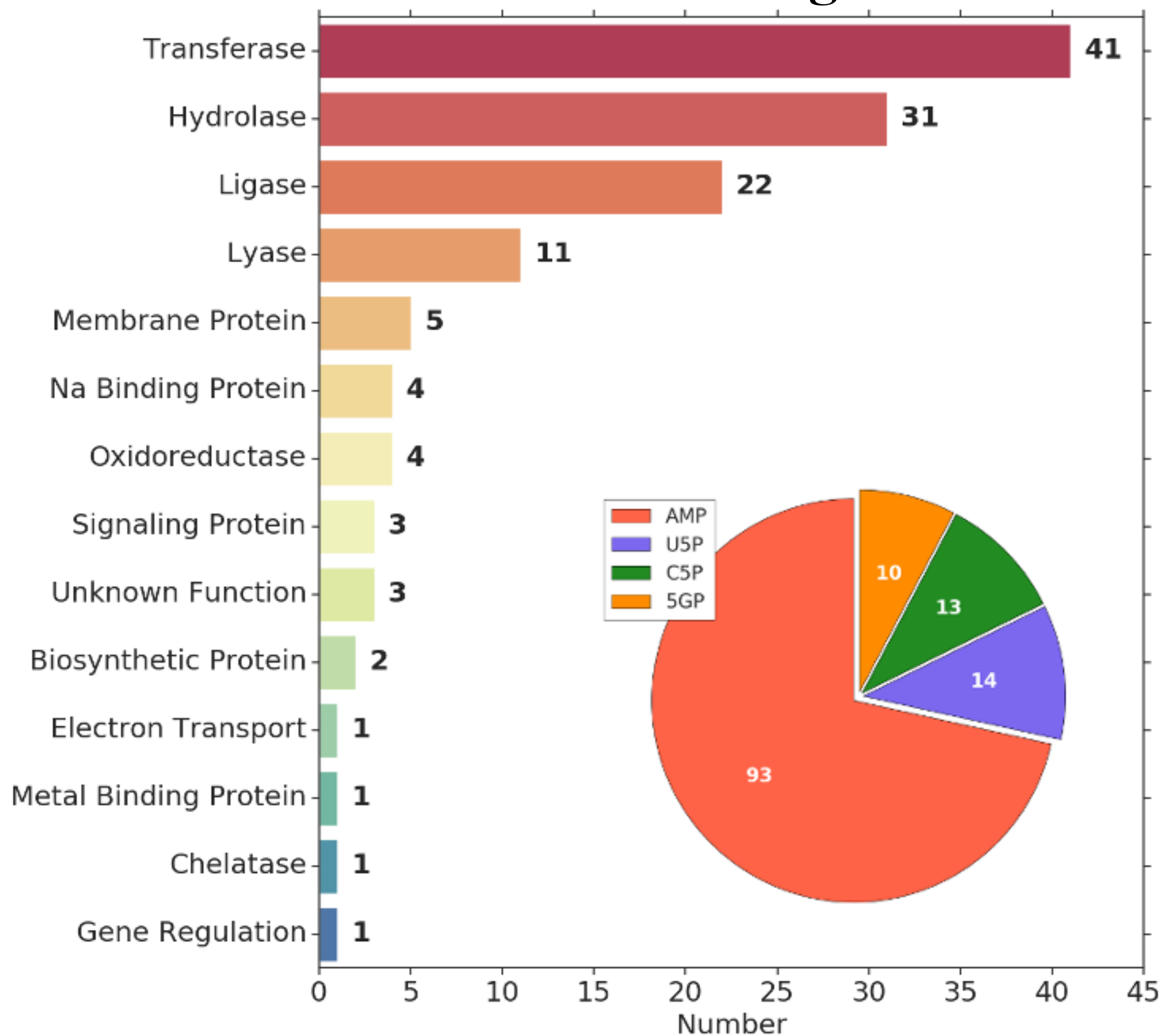
score-based nucleotide preferences



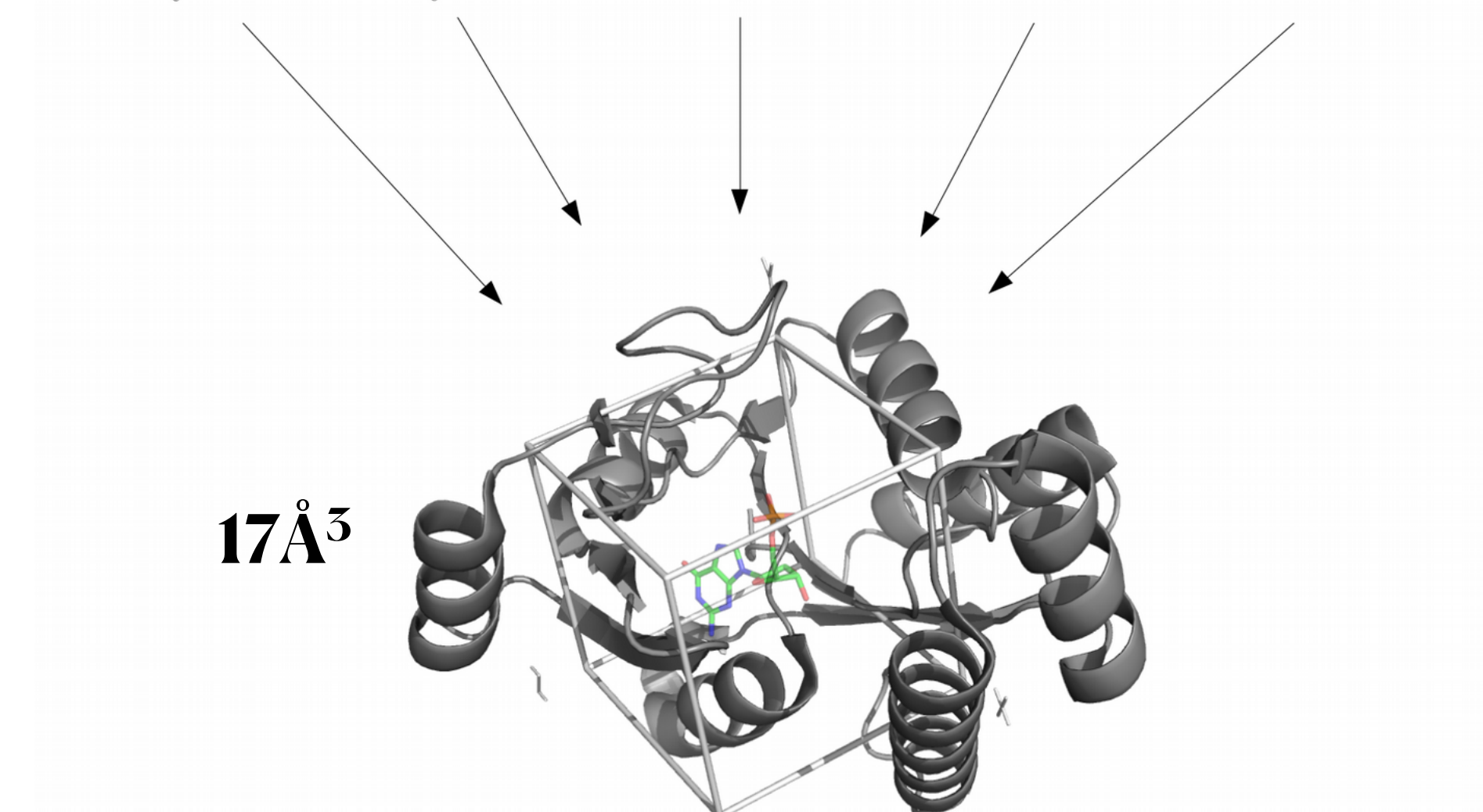
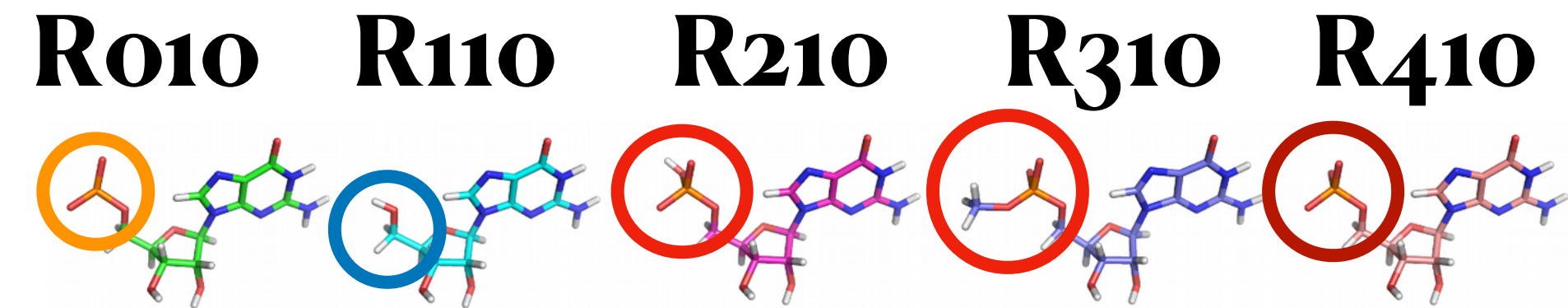
# Nucleotide-Protein Complexes

## Nucleotide Binding Sites

120 benchmark



5' patches  
R=A,C,G,U



**Minimized protein  
without ligand**

**native pose:  $\text{RMSD}_{(\text{mcSS}/\text{exp})} \leq 2.0\text{\AA}$**

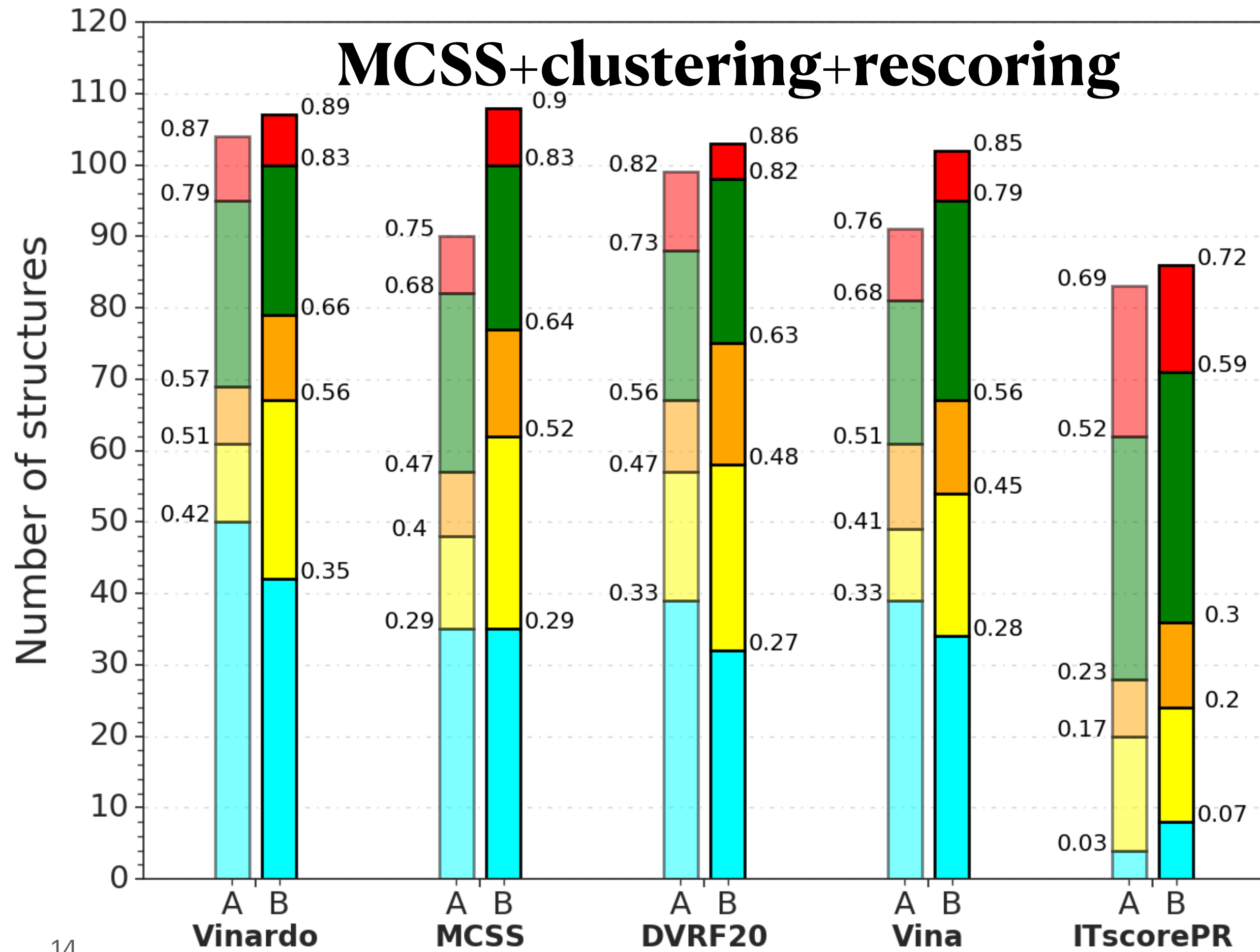
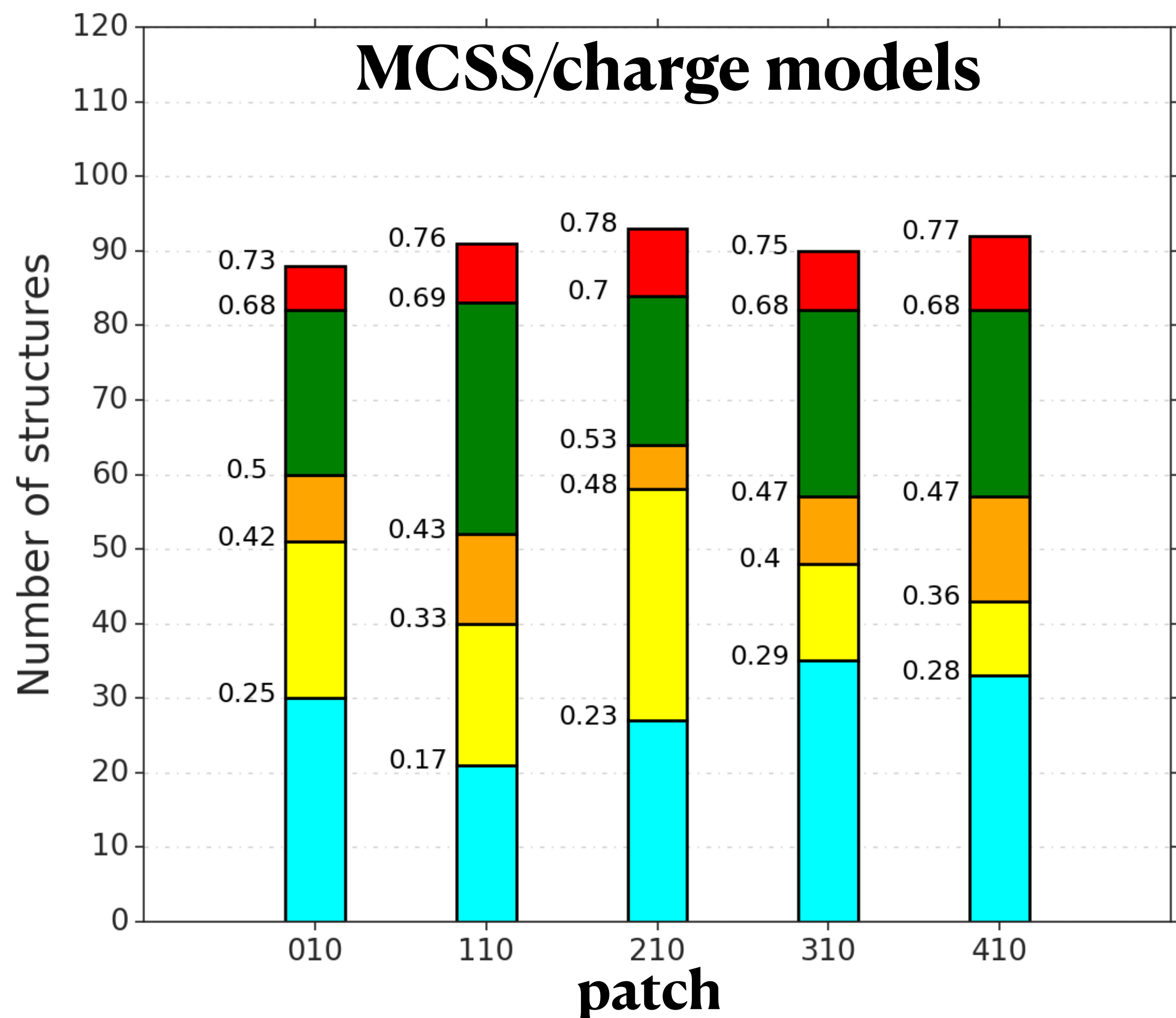
high resolution, non-redundant, etc



# MCSS Charges & Implicit Model / Scoring Functions

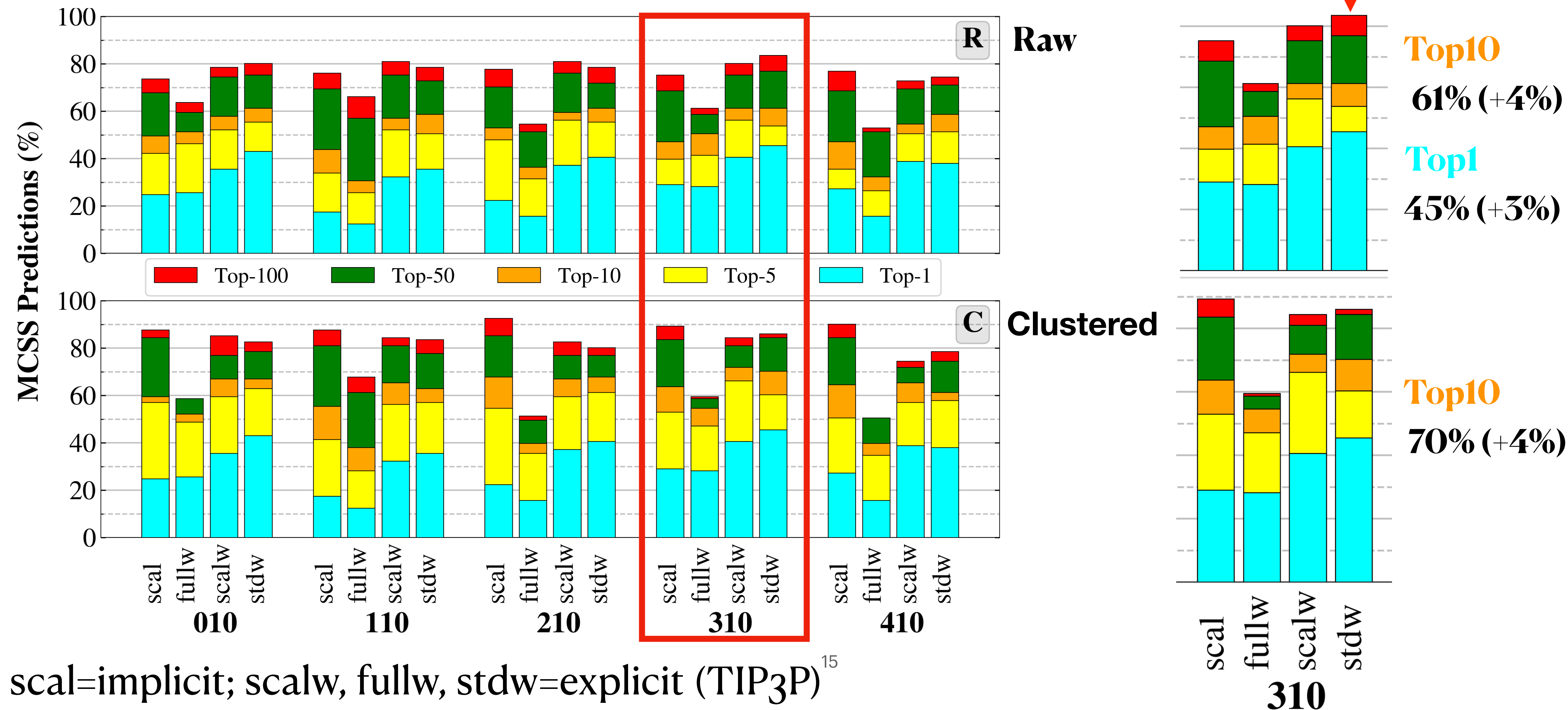
## General Performance on 120 benchmark

■ Top-100   
 ■ Top-50   
 ■ Top-10   
 ■ Top-5   
 ■ Top-1



# MCSS Charges & Solvent Models: Implicit/Explicit

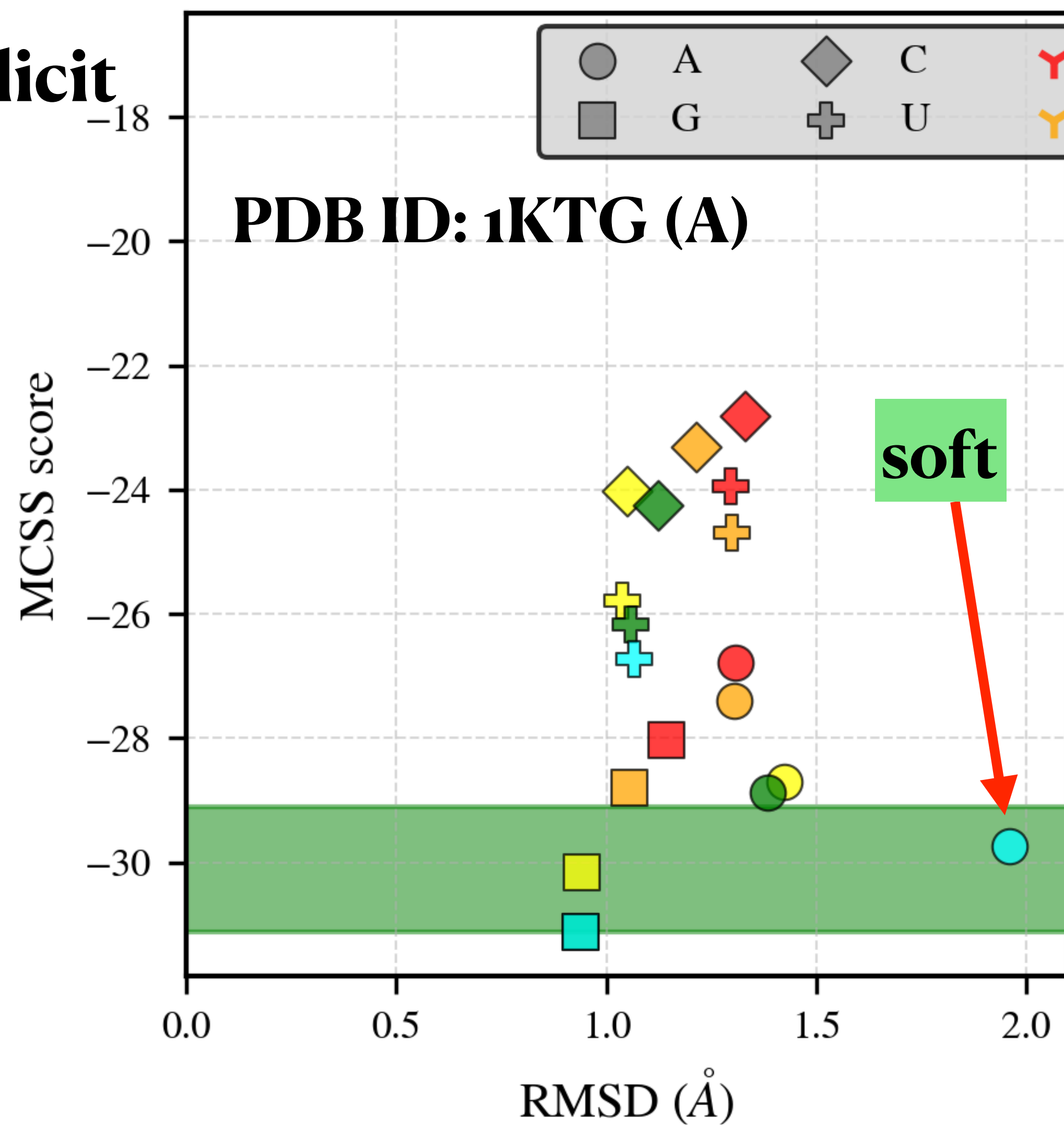
## General Performance on 120 benchmark



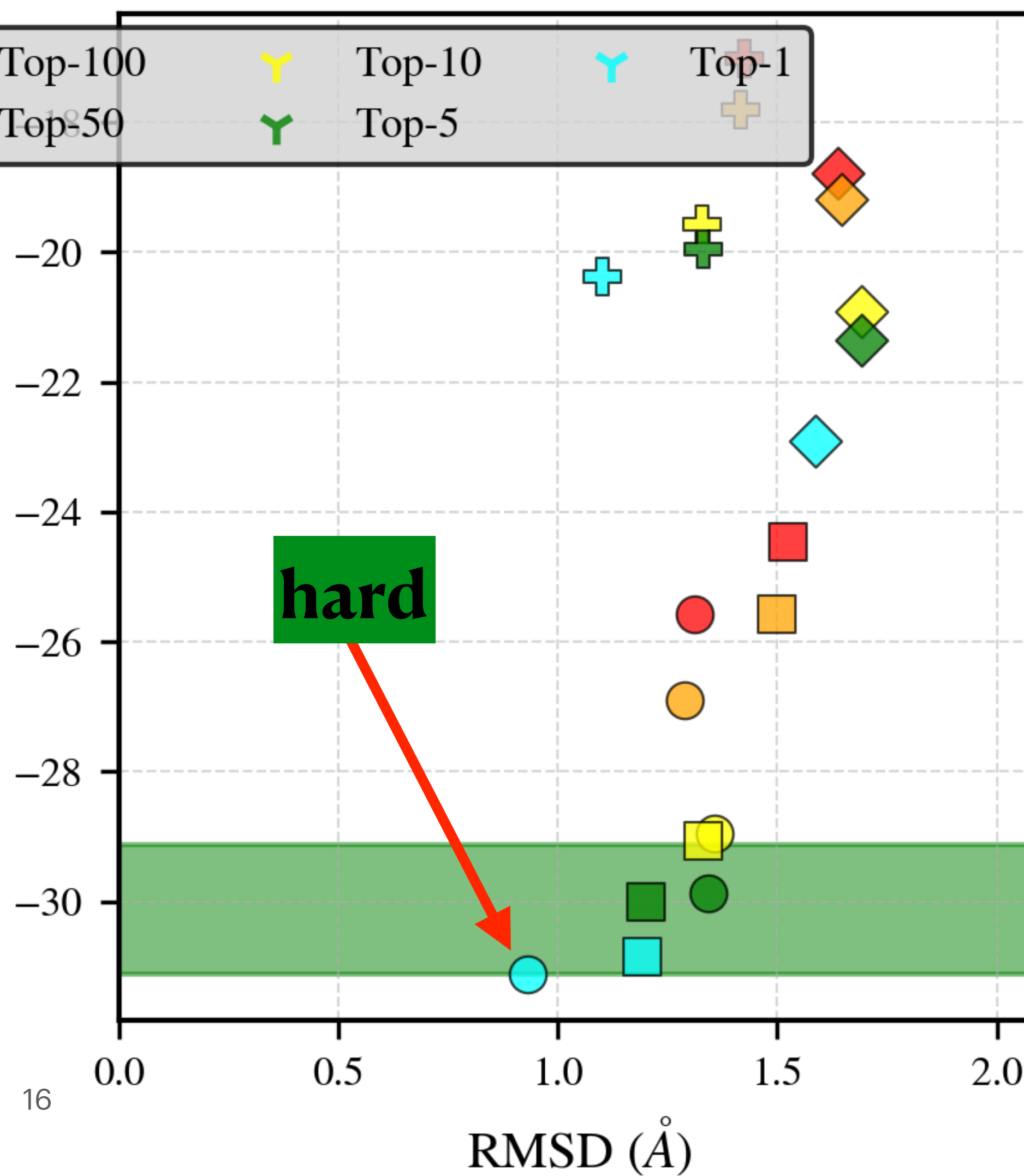
# MCSS Charges & Solvent Models: Implicit/Explicit

## Hard/Soft Predictions

implicit



explicit



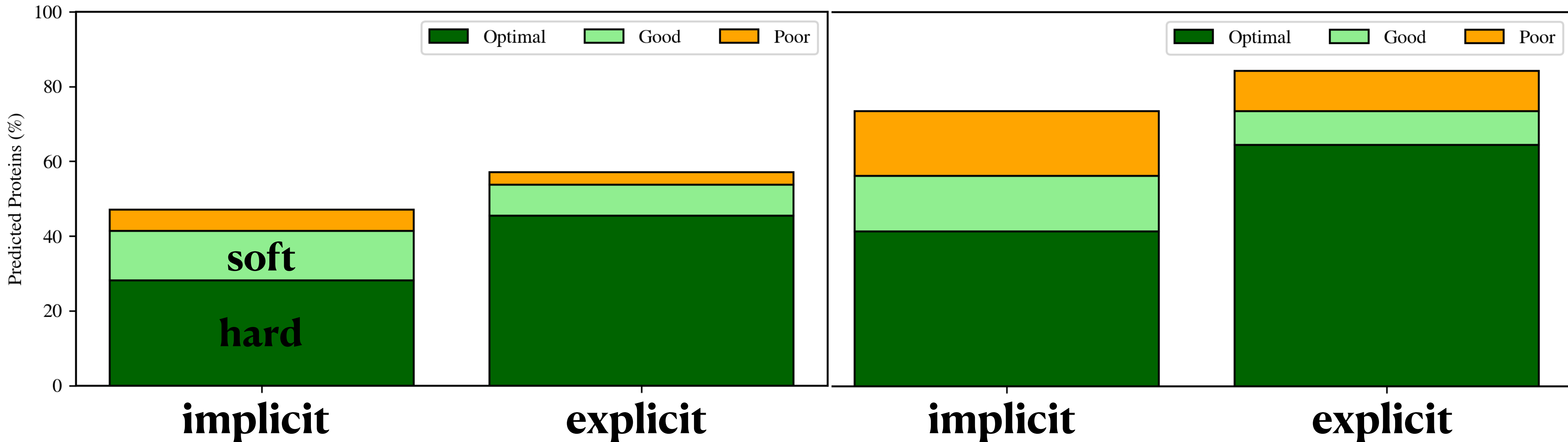


# Implicit/Explicit Models & Hard/Soft Predictions

Optimal (hard) / Good (soft) / Poor / No prediction

Top 10

Top 100



# No (soft, poor) Predictions: Structural Features

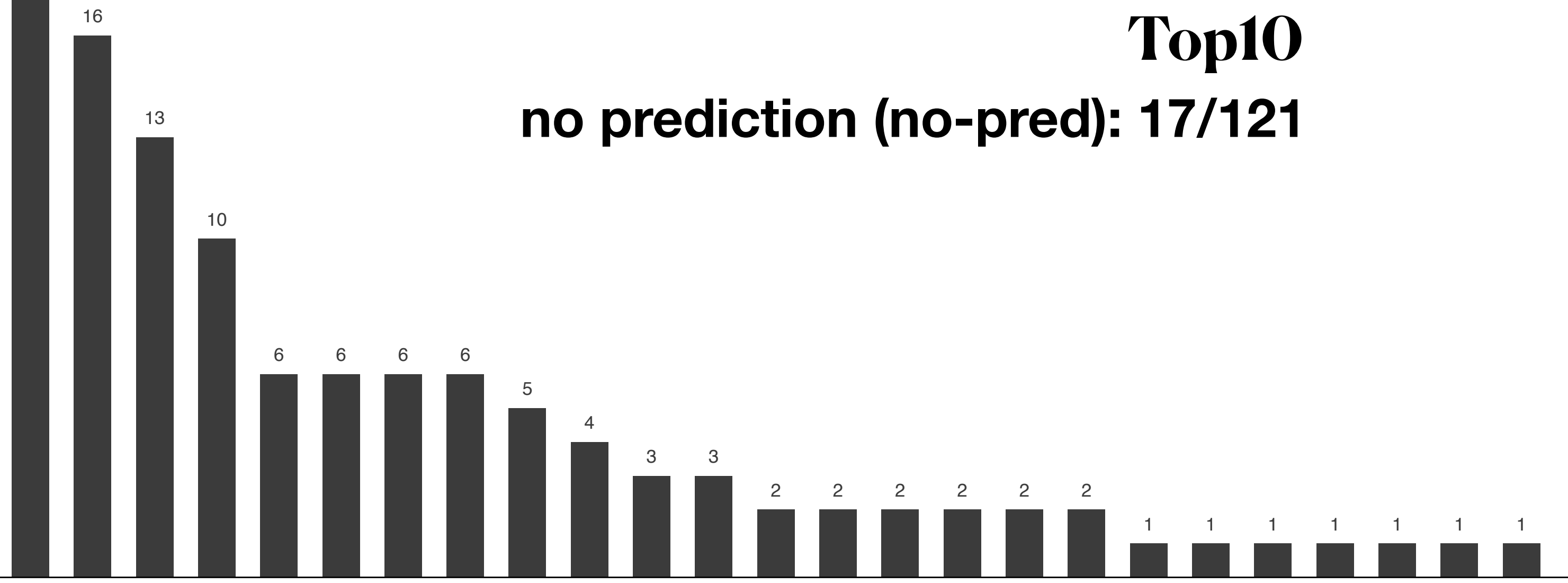
A qualitative analysis

Top10

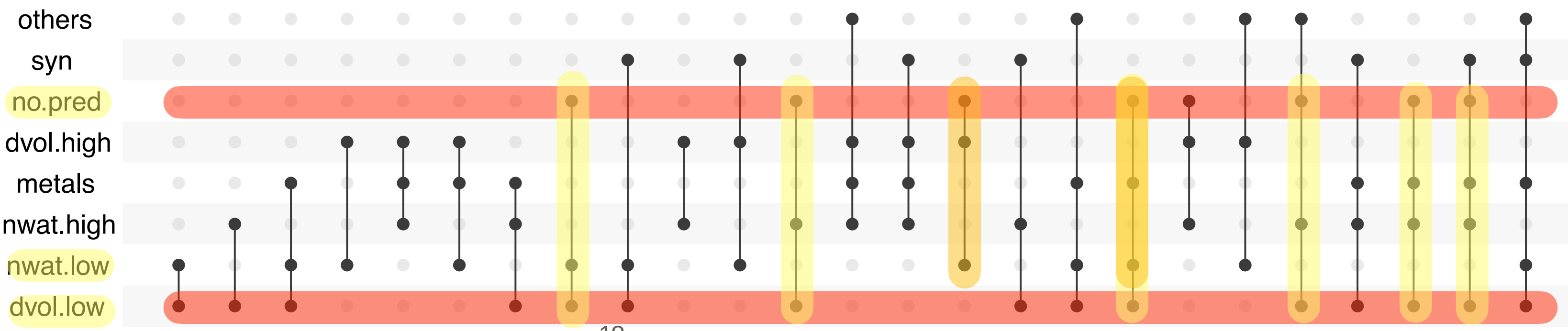
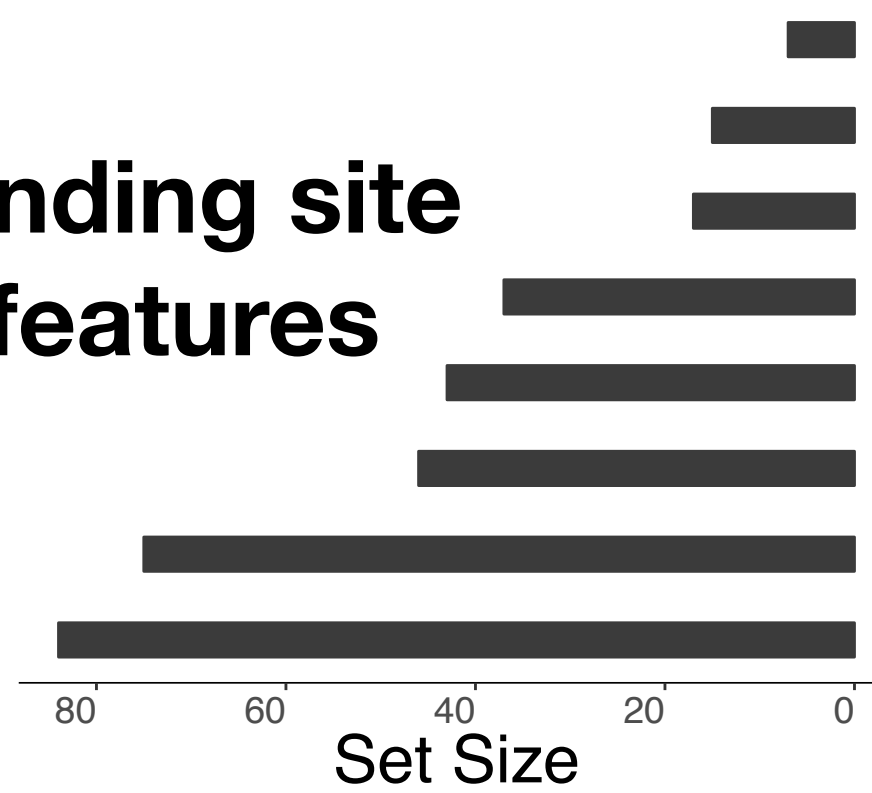
no prediction (no-pred): 17/121

Intersection Size

20  
10  
0



binding site features



18



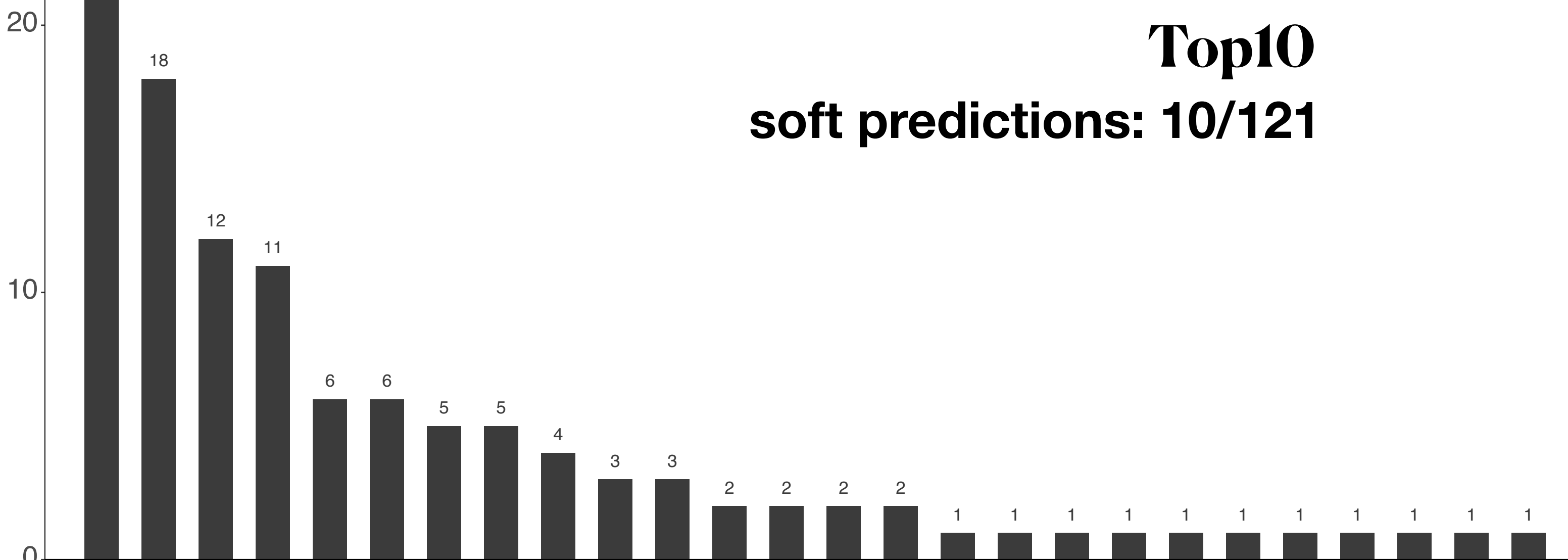
# Soft Predictions: Structural Features

A qualitative analysis

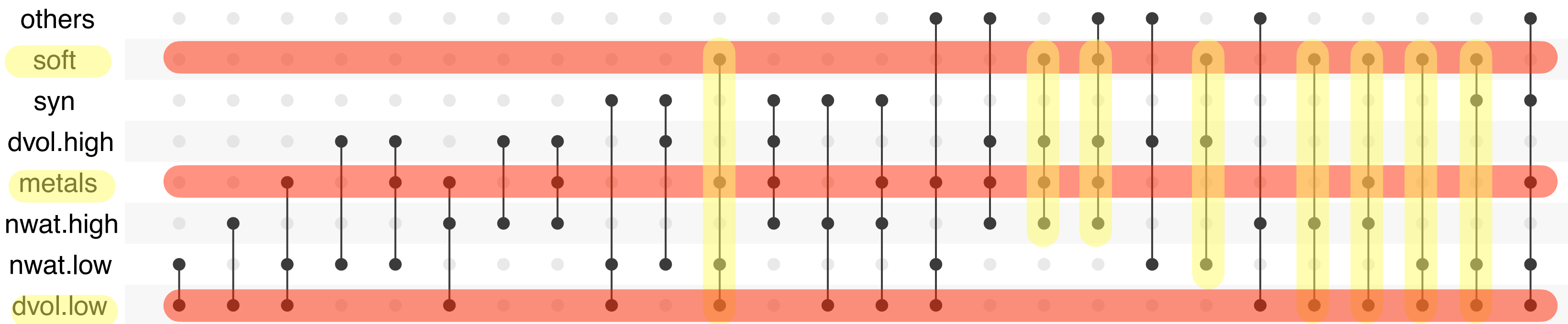
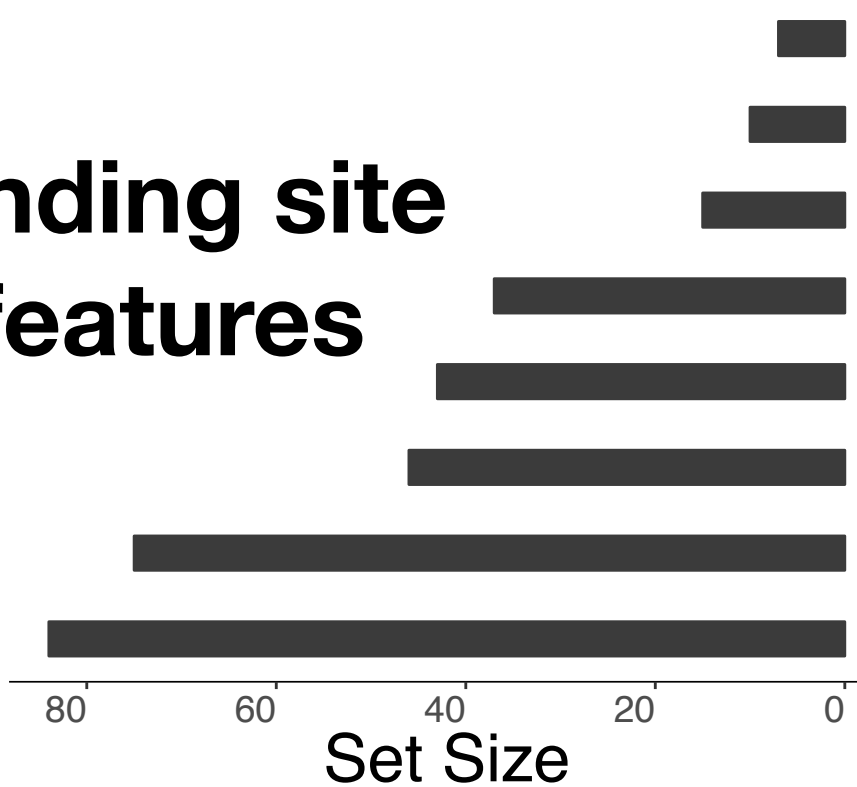
Top10

soft predictions: 10/121

Intersection Size



binding site features



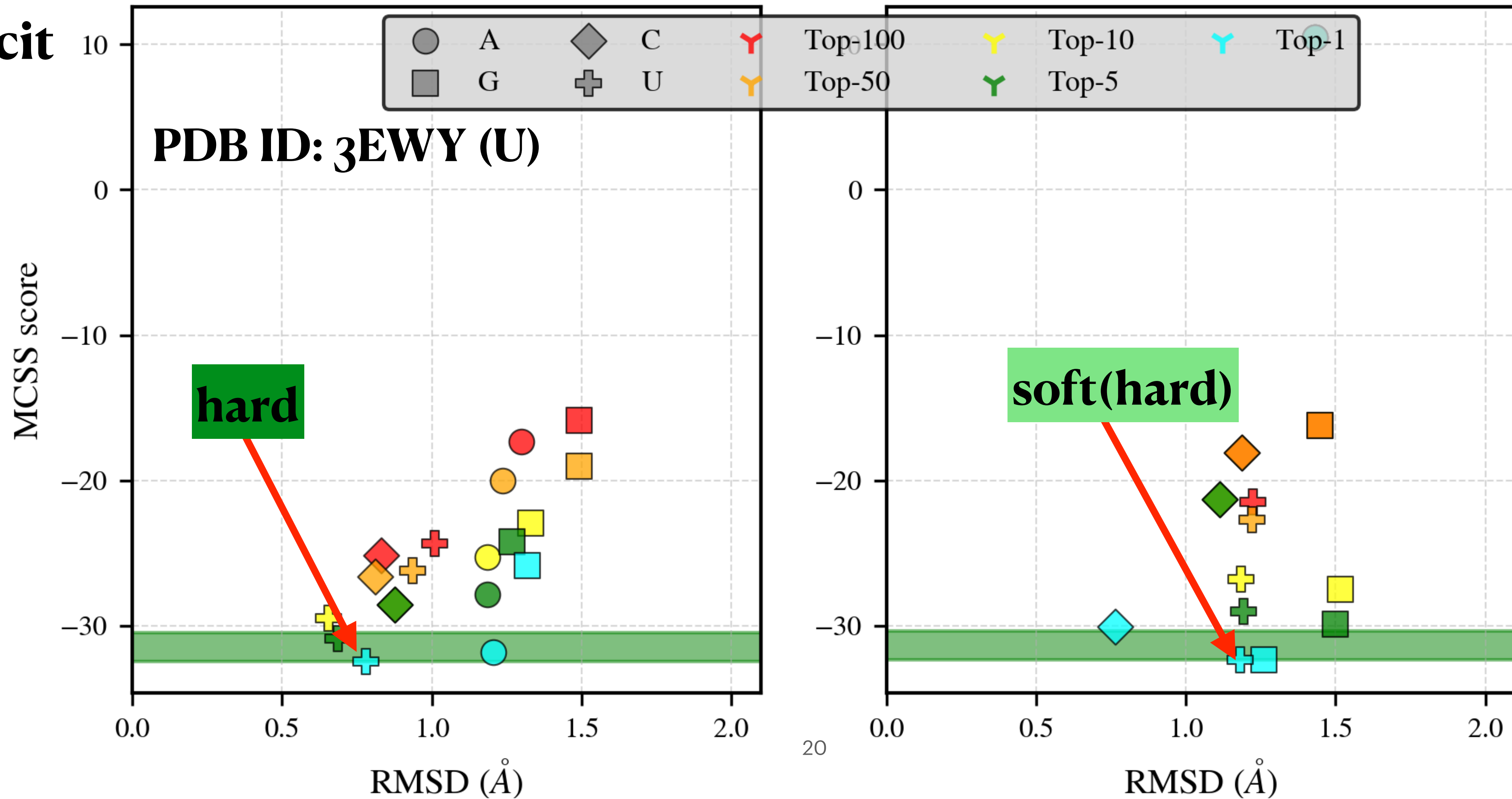
19

# Illustration of soft (poor) predictions

Why soft (poor, no) ?

implicit

explicit





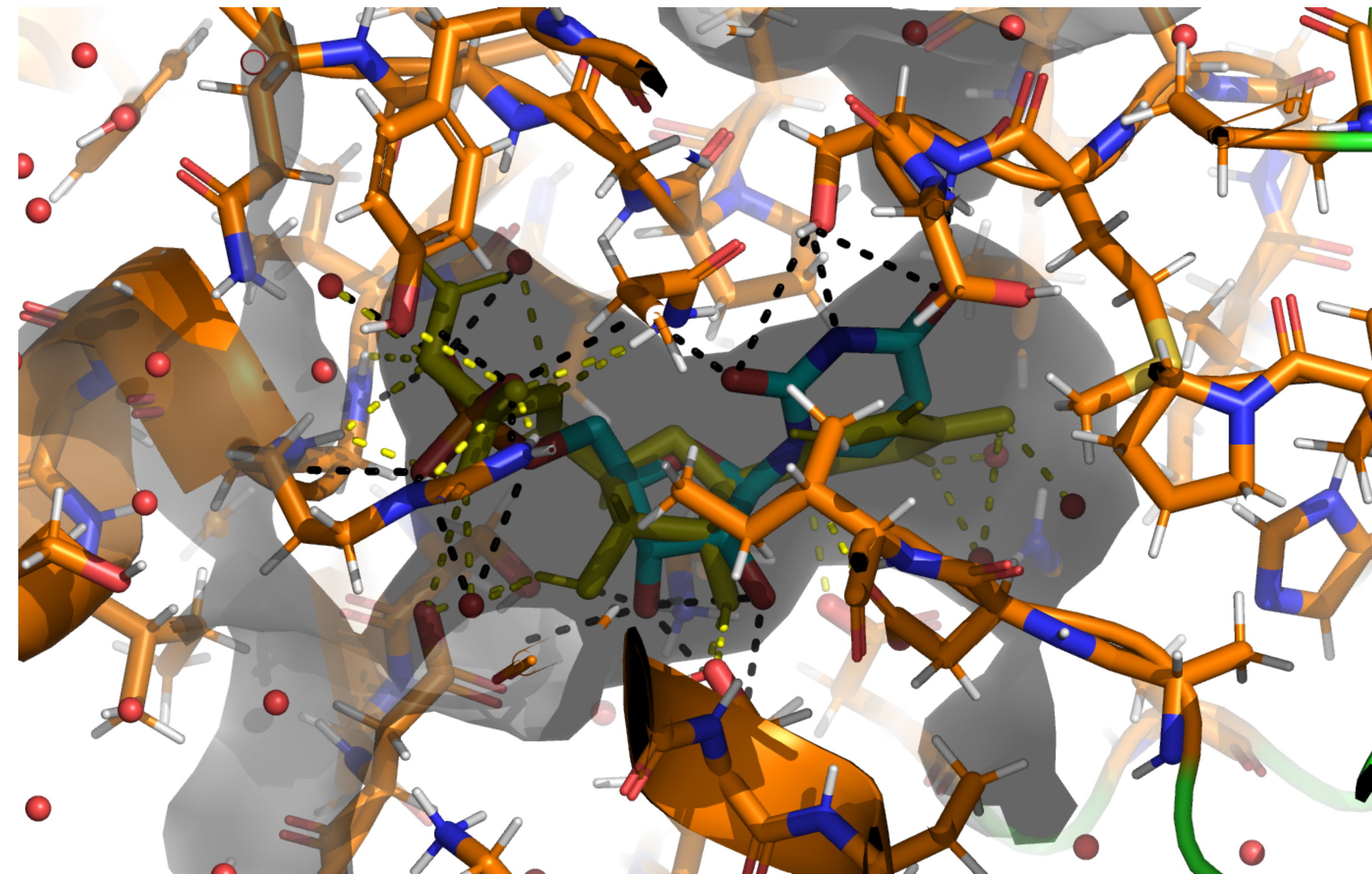
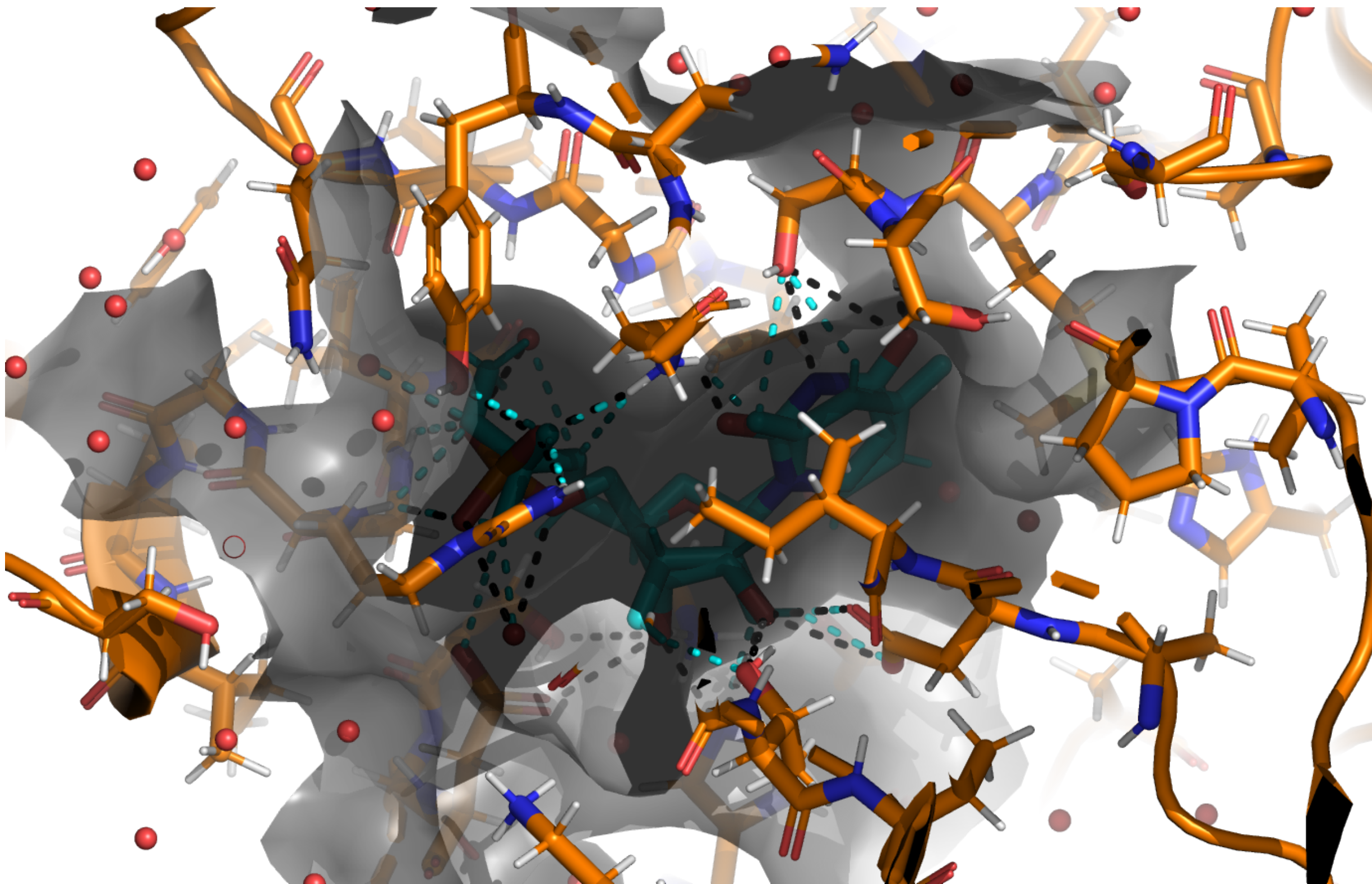
# Illustration of soft (poor) predictions

Binding modes

PDB ID: 3EWY (U)

implicit

explicit

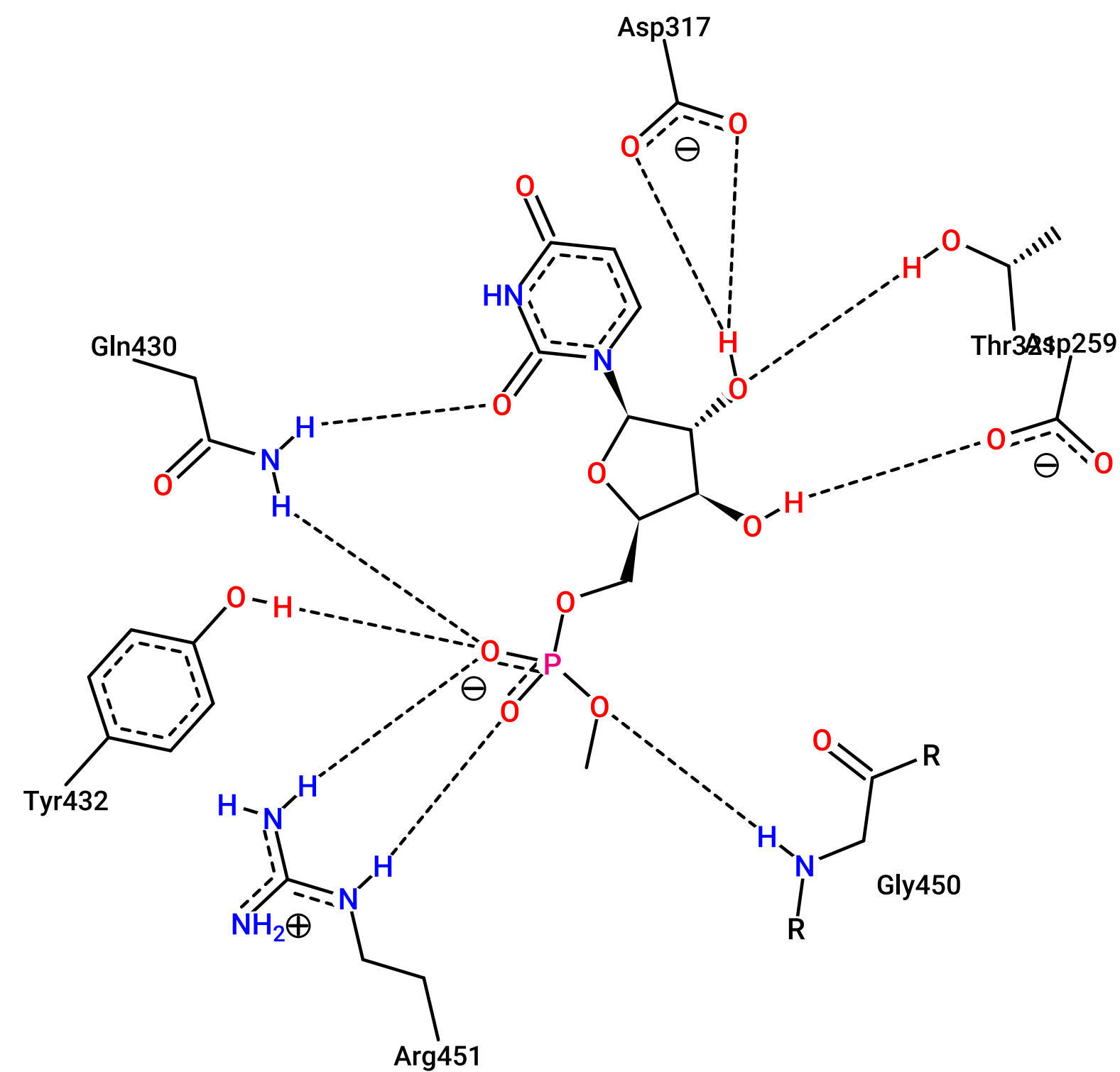




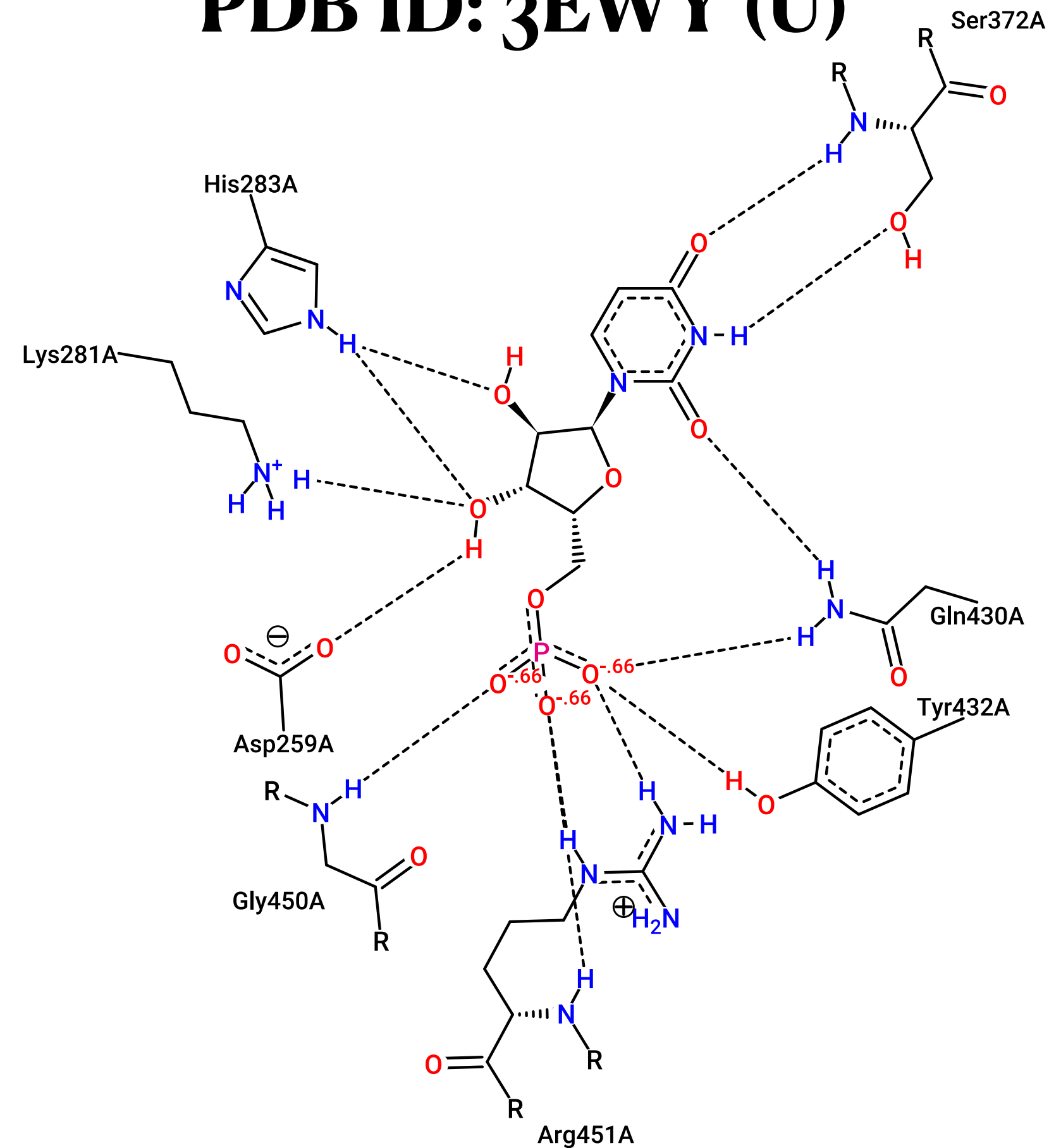
# Illustration of soft (poor) predictions

## Loss of native contacts

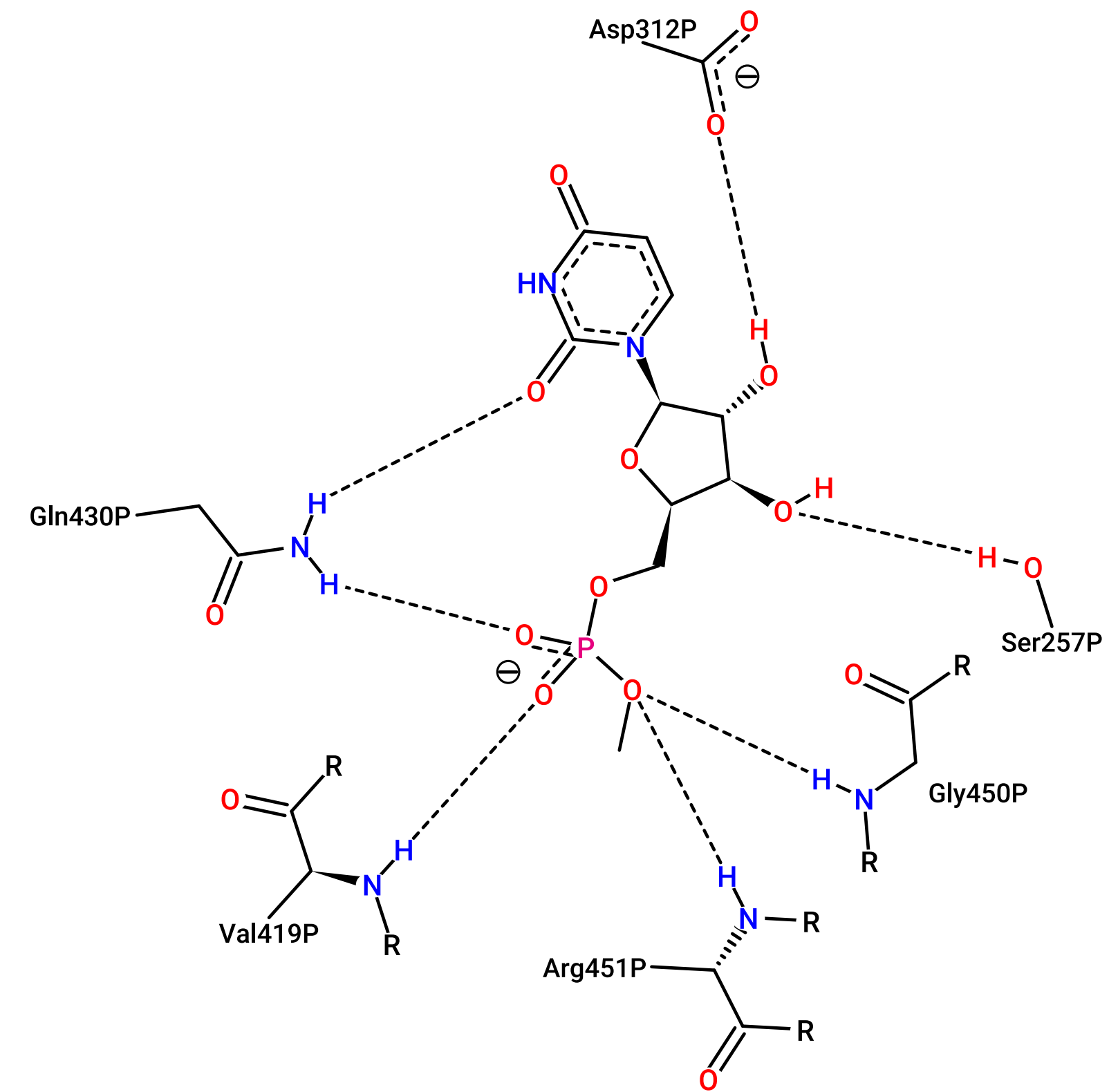
implicit



PDB ID: 3EWY (U)



explicit



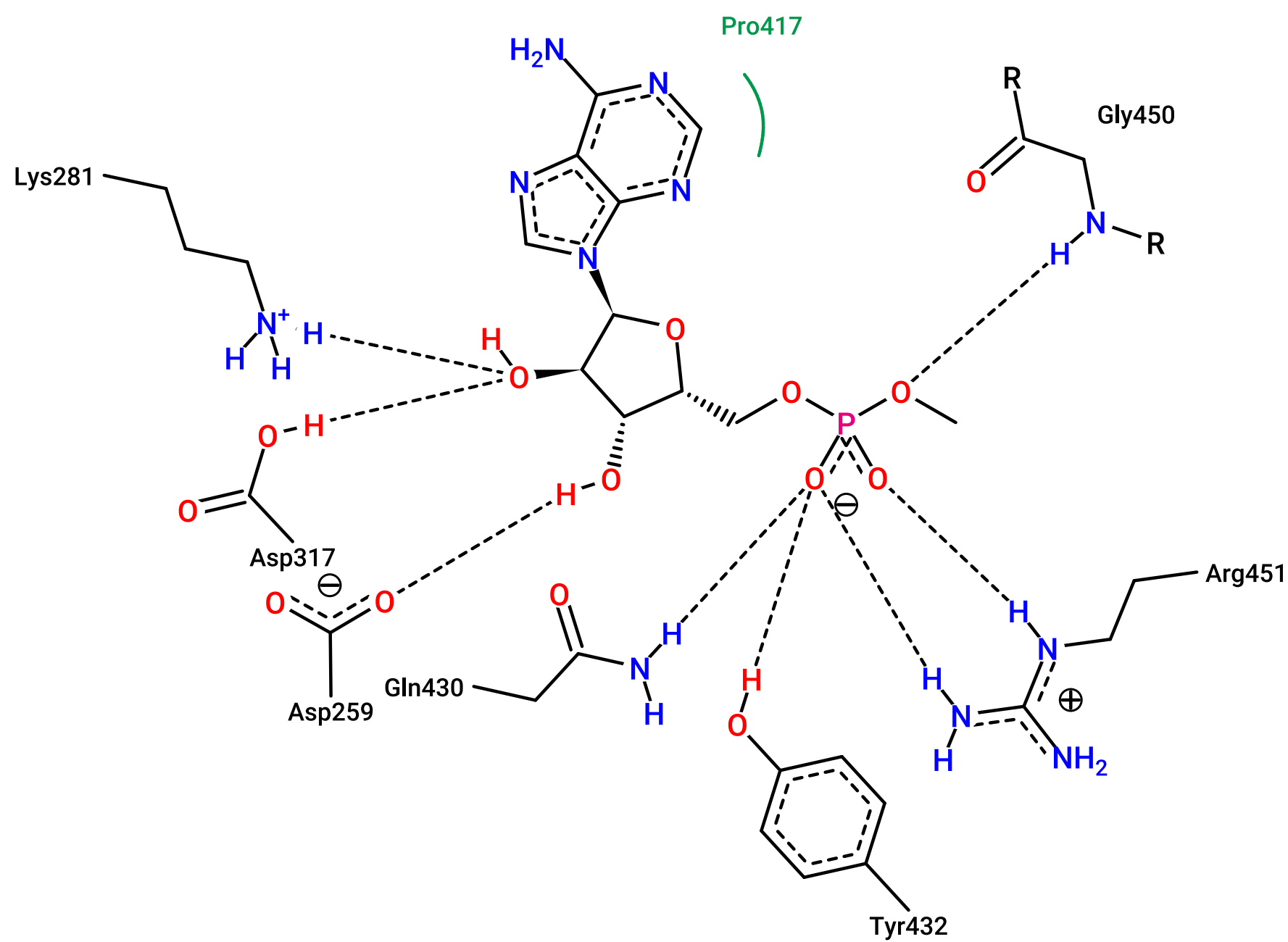


# Illustration of soft (poor) predictions

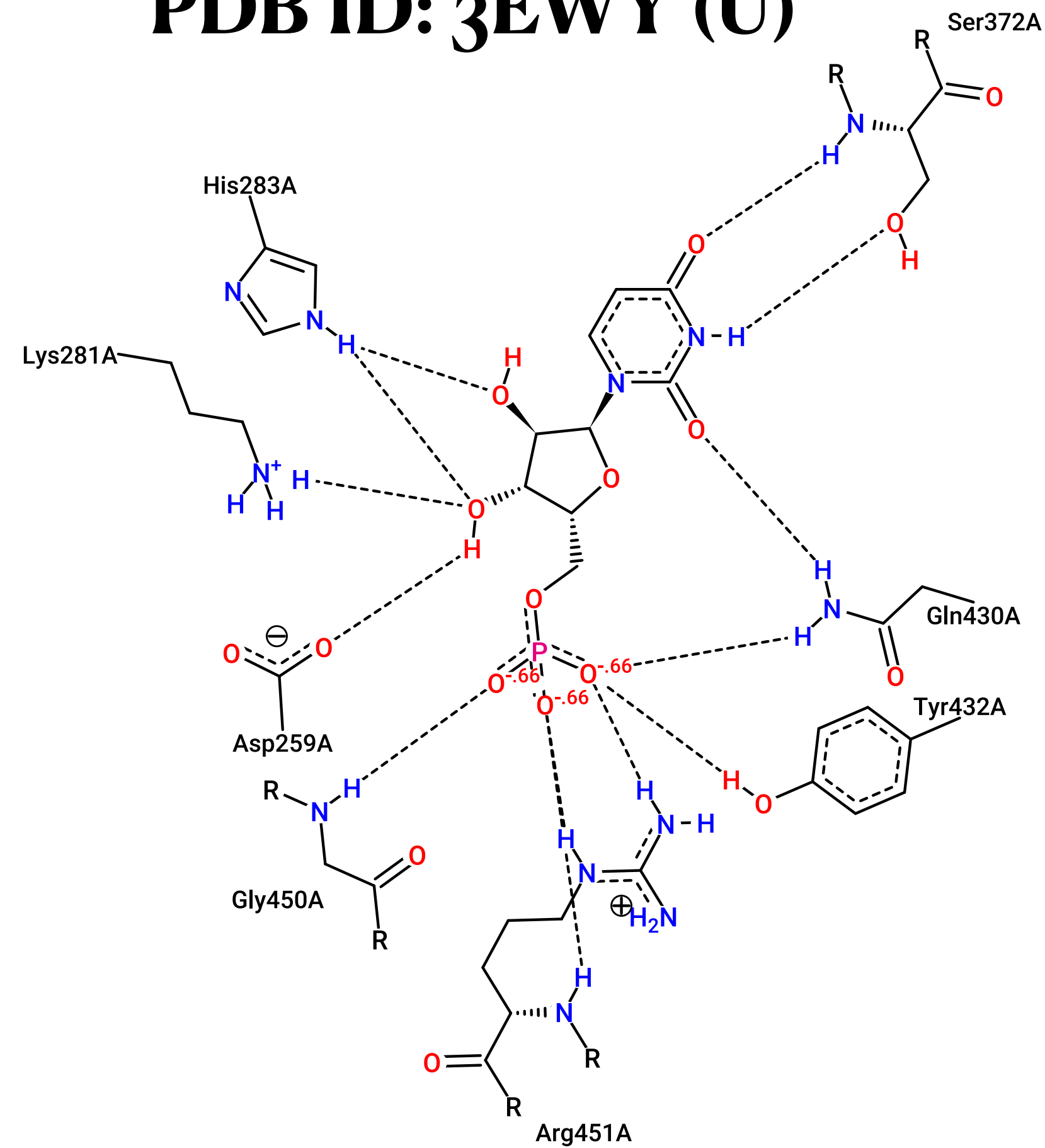
## Non-native contacts

implicit

non-native A

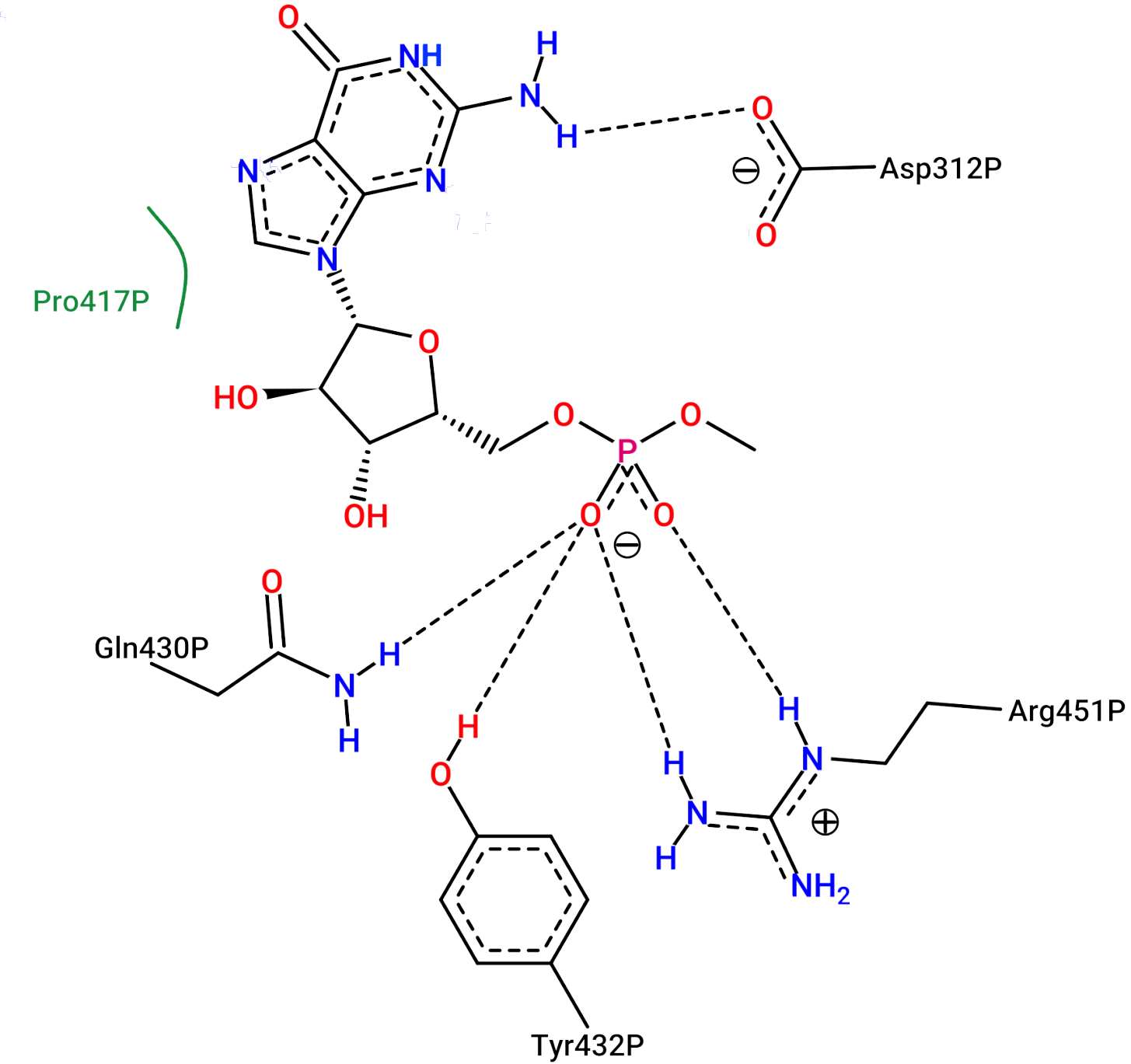


PDB ID: 3EWY (U)



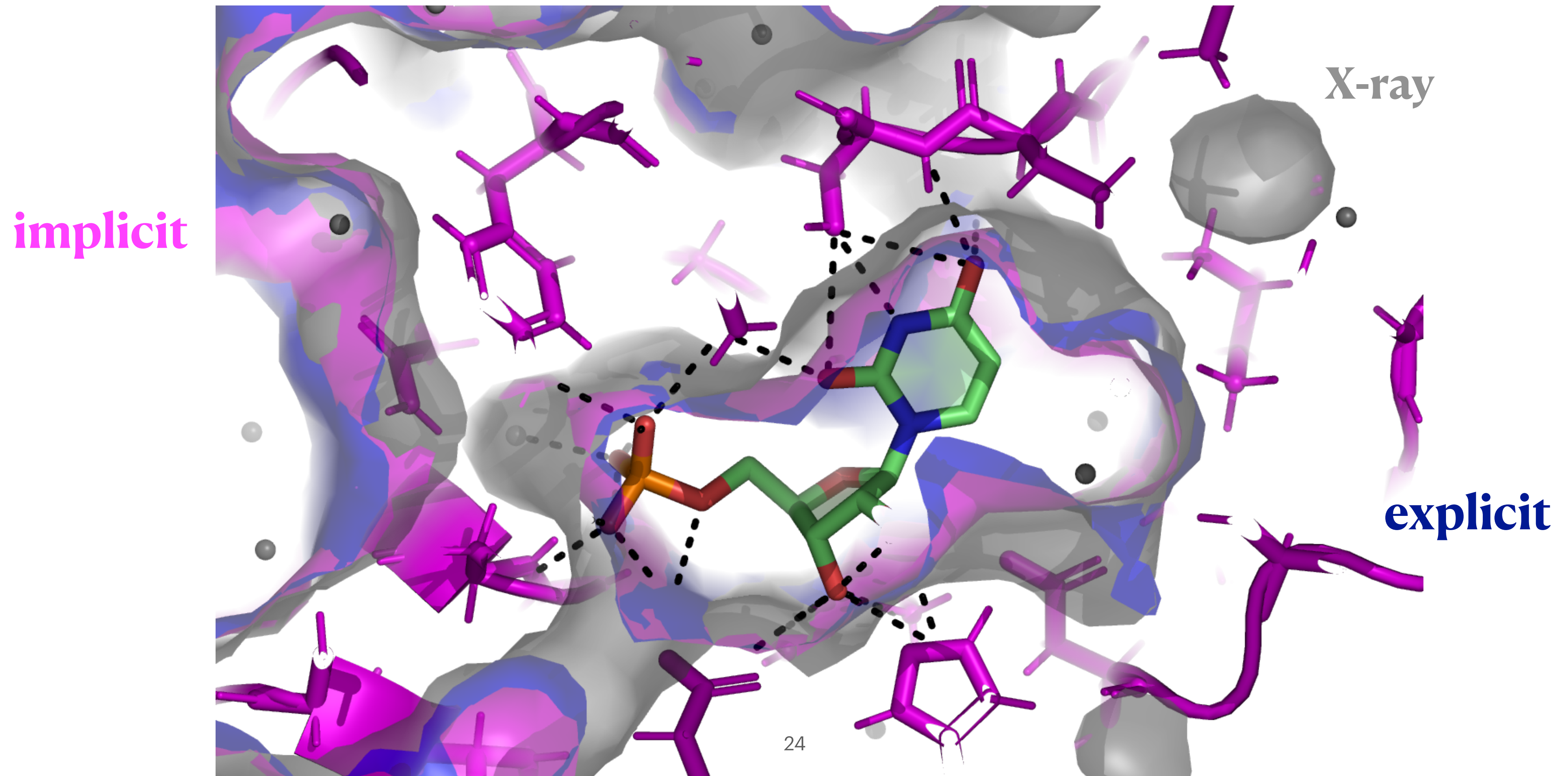
explicit

non-native G



# Illustration of soft (poor) predictions

“shrinking” of the binding site



# Conclusions

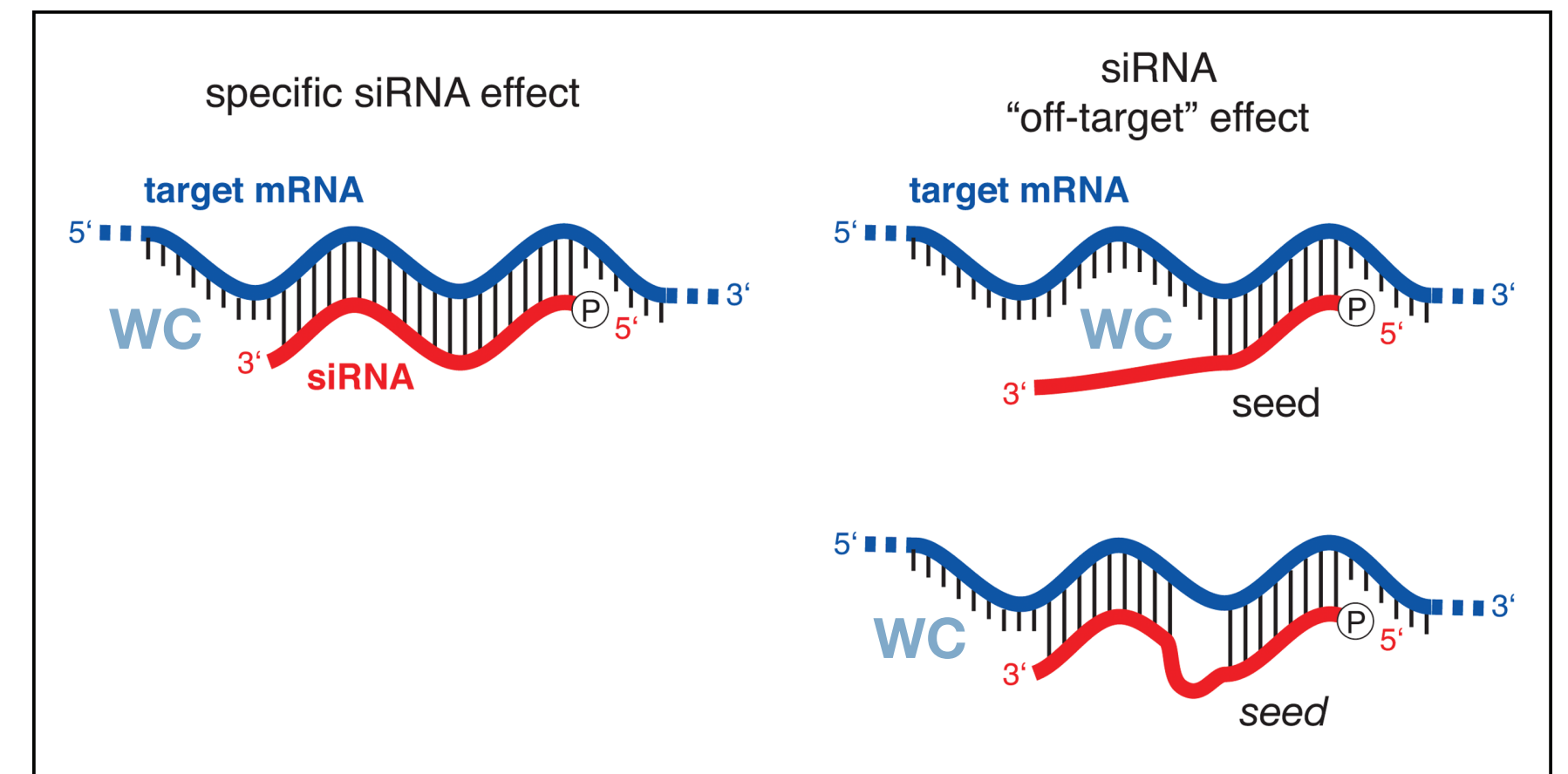
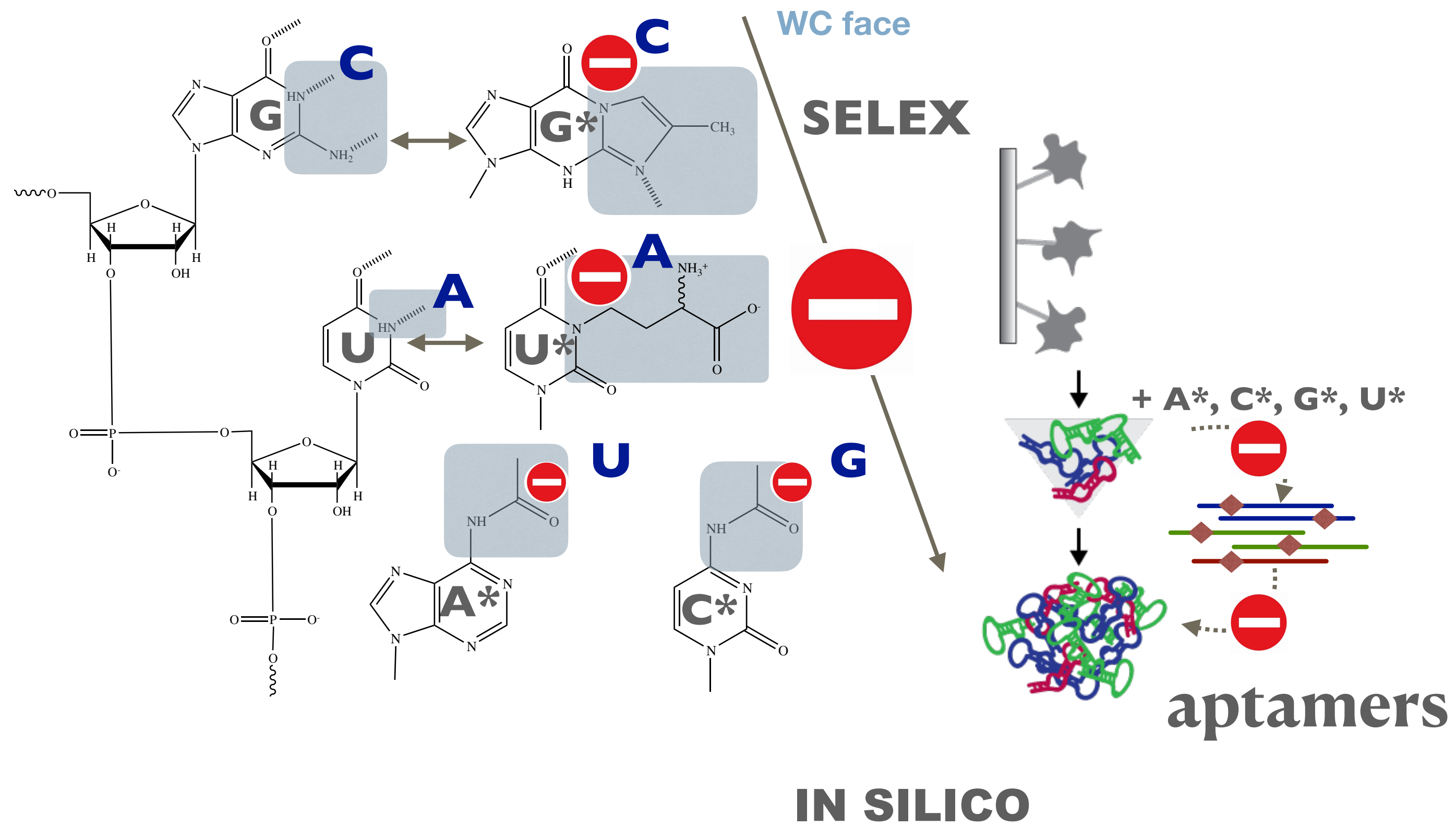
## strengths & weaknesses

- **fragment-based** approach applied to single nucleotide binding
- **explicit solvent** improve predictions of nucleotide binding
  - improved docking power (native poses)
  - improved screening power (“true” binders)
- **pitfalls:**
  - protein flexibility in the binding site
  - binding thermodynamics
  - binding kinetics



# Perspectives and opportunities

*in silico* approach to fragment-based design



# Acknowledgments

- Roy González-Alemán (Ph.D. student, 2019-)
- Nicolas Chevrollier (Ph.D., 2019)
- Coralie Rohmer (MSc. 2017)
- Manuel Simoes (Ph.D.)
- Martin Karplus (Prof. emeritus)
- Luis Montero-Cabrera (Prof.)

