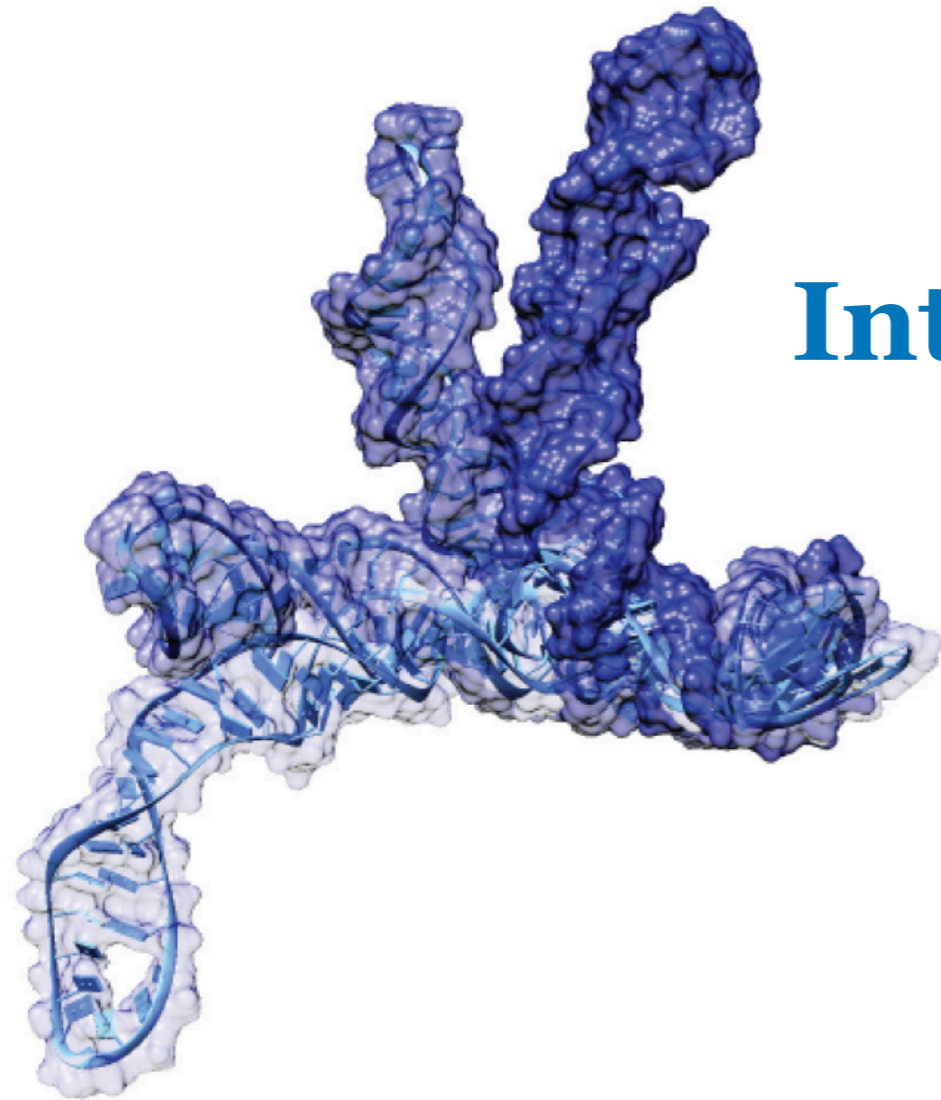# Université Paris Cité

Integrative modeling strategies and RNA polymorphism
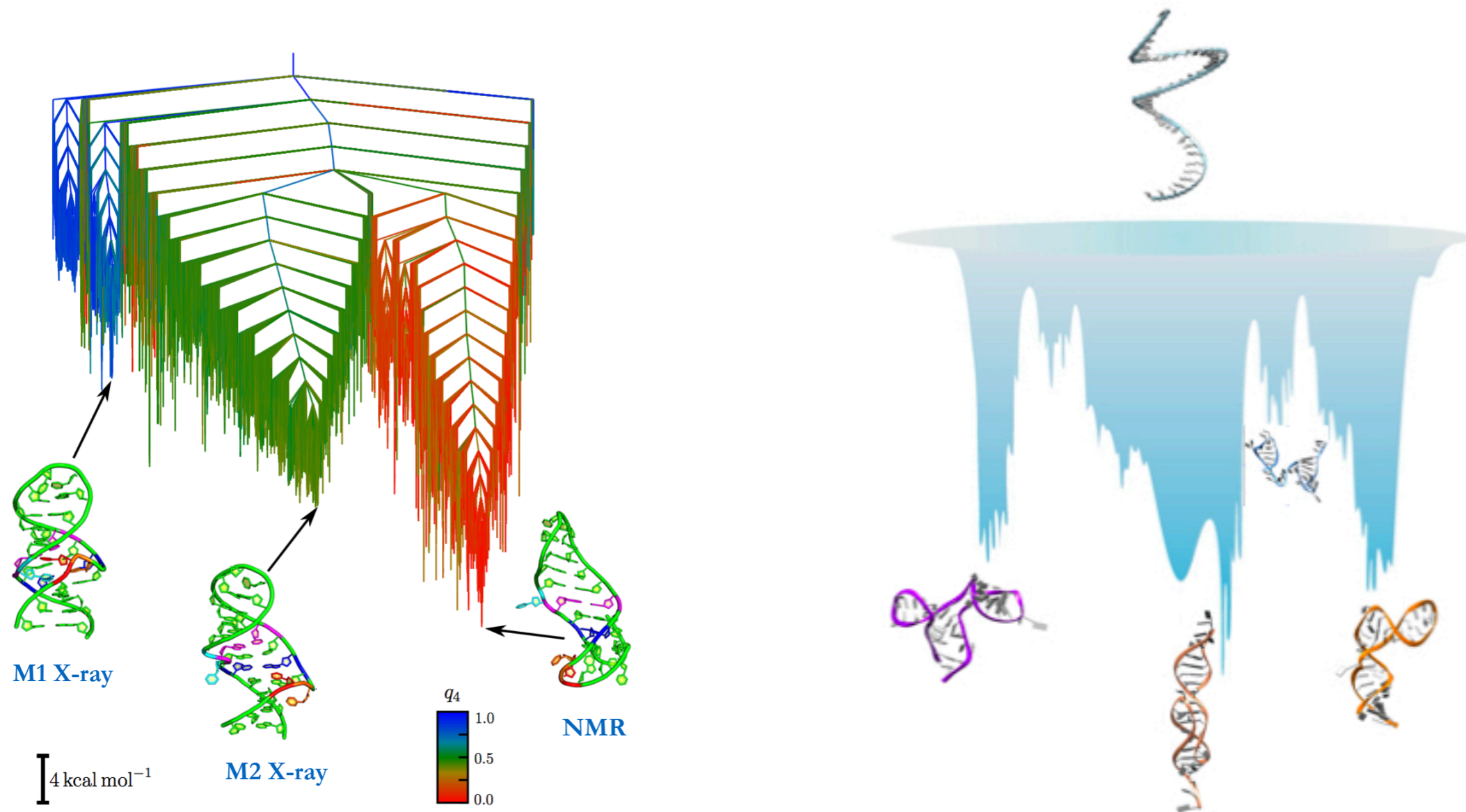
Samuela Pasquali

Laboratoire Biologie Fonctionnelle et Adaptative *et* Laboratoire CiTCoM

Polymorphism and energy landscapes

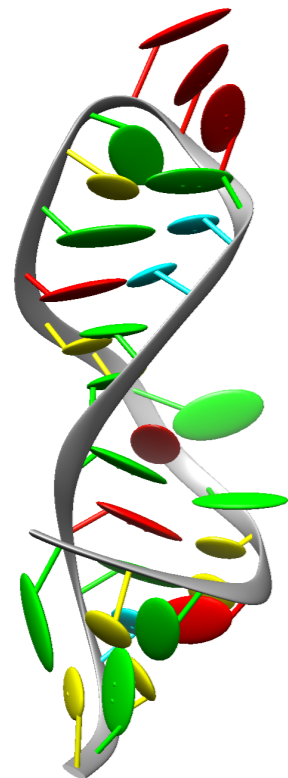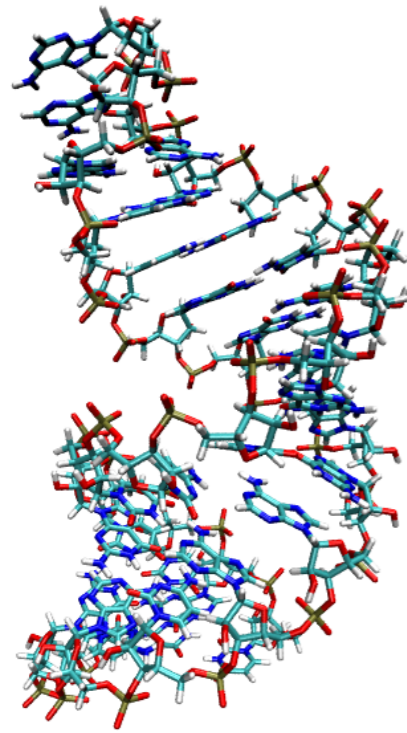## *A given sequence can adopt multiple structures of comparable energies*

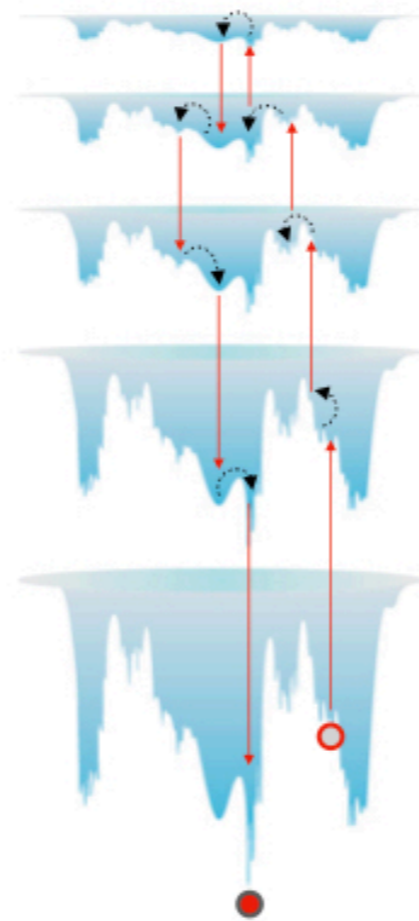(opposite to the fast folding paradigm for proteins)

M1 X-ray

4 kcal mol^{-1}

M2 X-ray

q_4

1.0

0.5

0.0

NMR

Observed structures depend on:

Temperature, Buffer, Ions (Mg^{++}), pH, Ligands, dynamics, ...

Modelling strategies
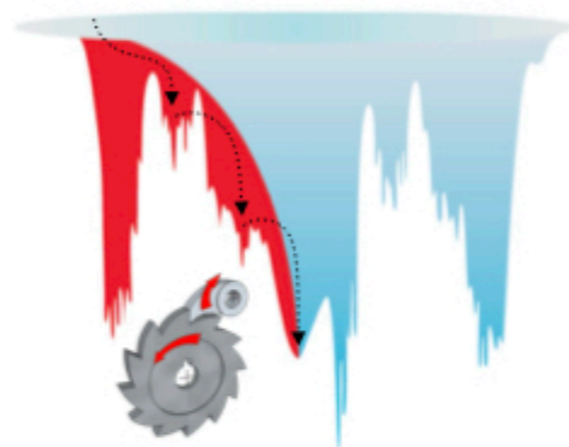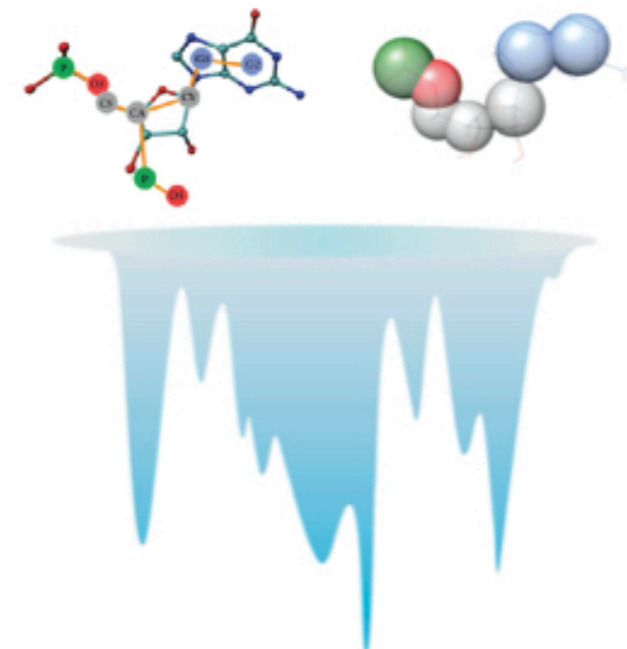
Why multiple methods ?
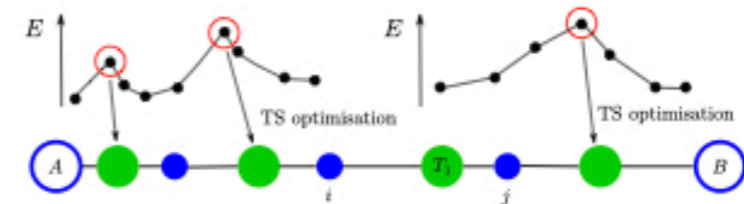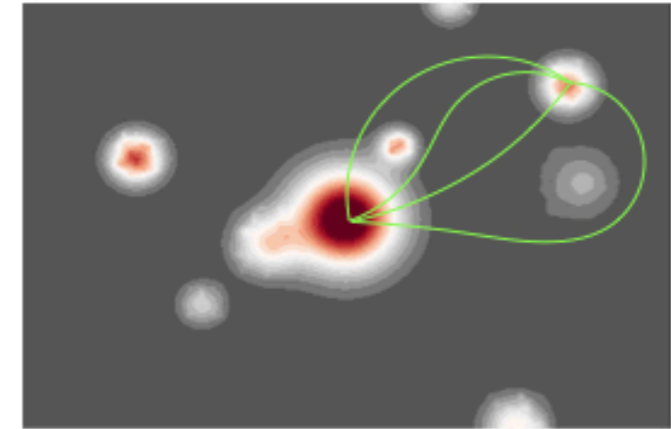
H-REX

✔ Full Solvation
Only solute is affected by the energy rescaling
Use standard force fields

✘ Need multiple copies of the system (~30)
Can deal only with small systems (~few dozens nt)

DPS

✔ Sample widely the conformational space
Allows to define "families" of structures (basins)
Allow to obtain kinetic data

✘ Implicit solvent
Can deal only with small systems (~few dozens nt)

rMD

✔ Full Solvation
Use standard force fields
Good statistics on folding trajectories

✘ Knowledge of native state
Cannot study alternative foldings in a single MD

CG models

✔ Allow to study larger systems for longer time scales
Allow to highlight essential elements of the system description

✘ Loss of atomistic details
Implicit solvent
No generally recognized model

Including experimental data and environment

- What does the experiment actually measure/what is the physical influence of the environment?

- Is it possible to compute the outcome from hypothetical structures?

- Can we bias simulations based on this account?
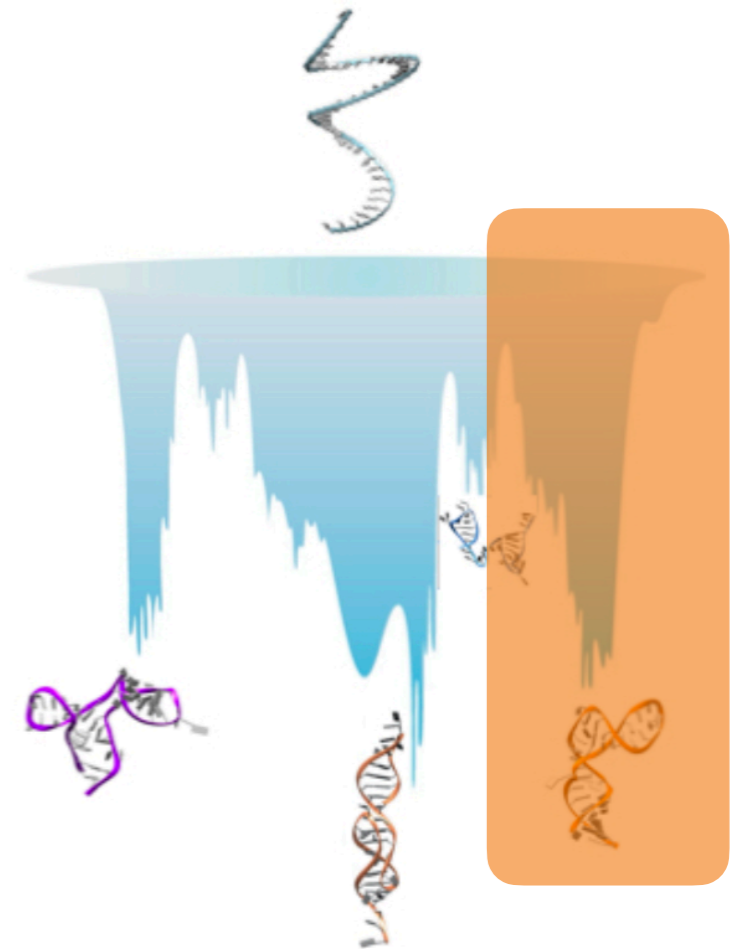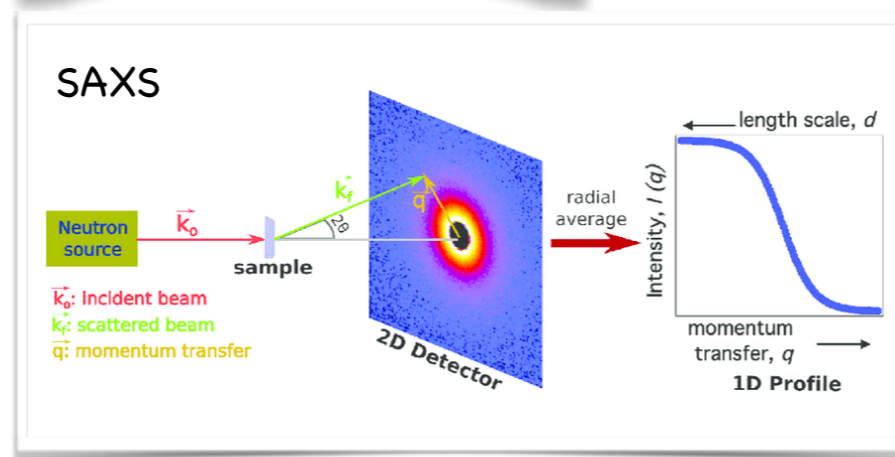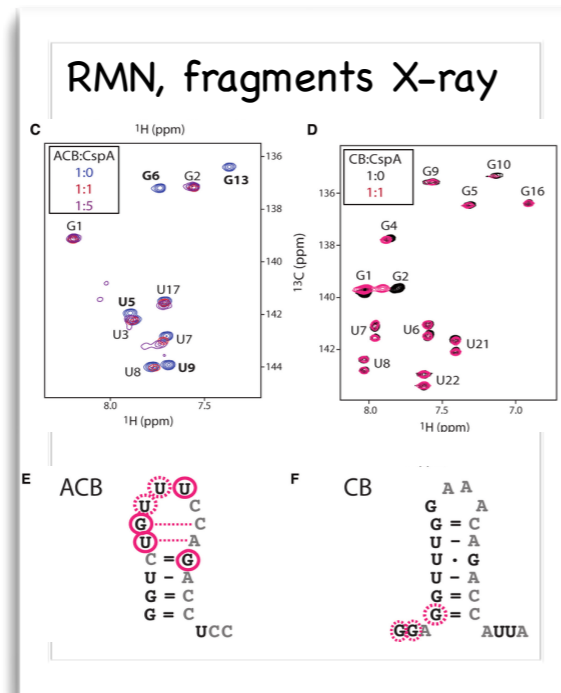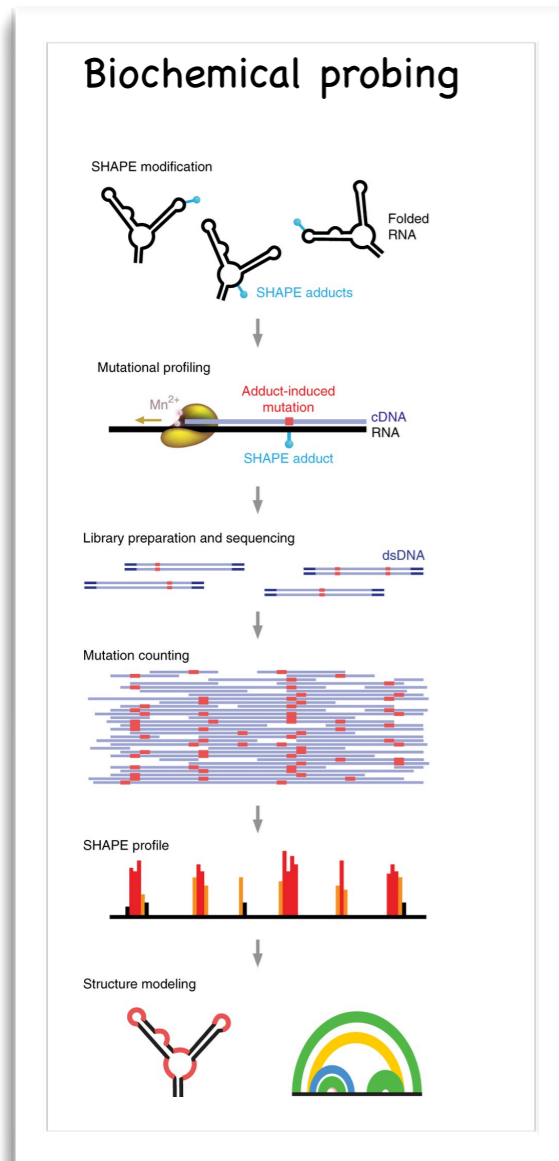
Experimental data restricts the conformational space to be explored

Use the CG model as a platform to integrate environment and experimental data to prototype possible approaches
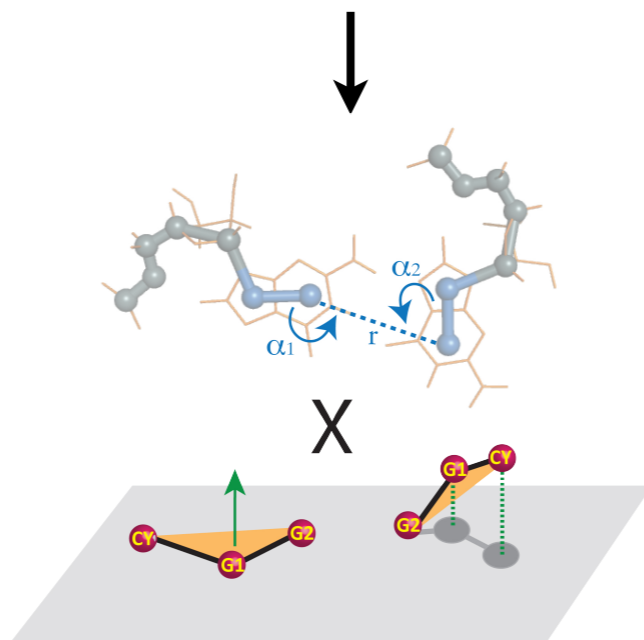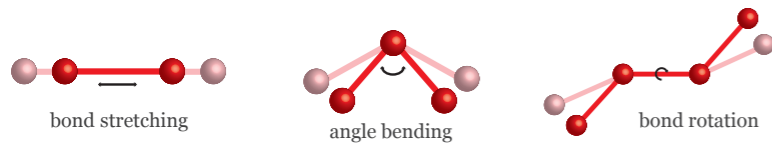
Coarse-grained RNA modeling: HiRE-RNA

simplified models to represent the meaningful degrees of freedom of the system and the process of interest
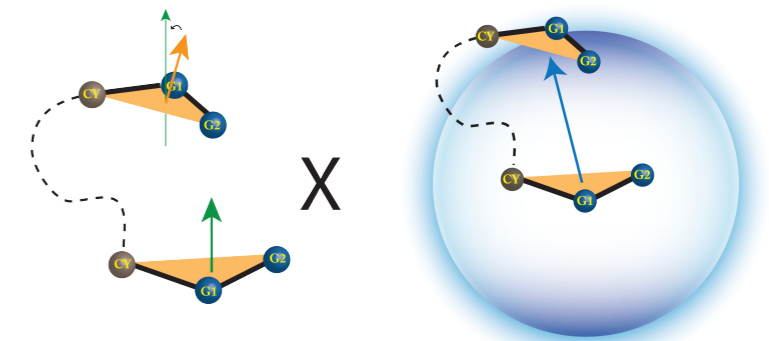

Adenine

Guanine

Uracil

Cytosine
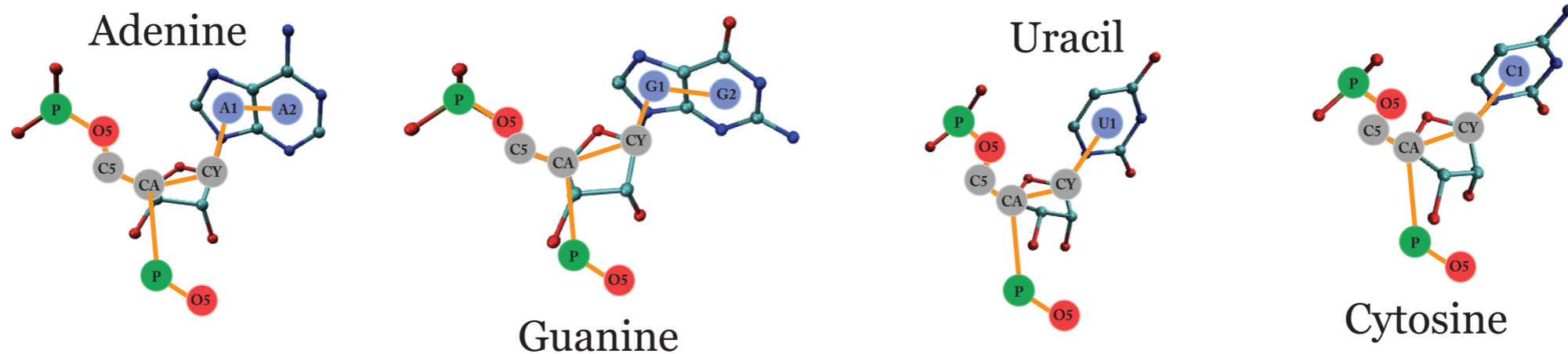
$$E = E_{\text{local}} + E_{\text{ex vol}} + E_{\text{BP}} + E_{\text{electrostatics}} + E_{\text{stacking}}$$
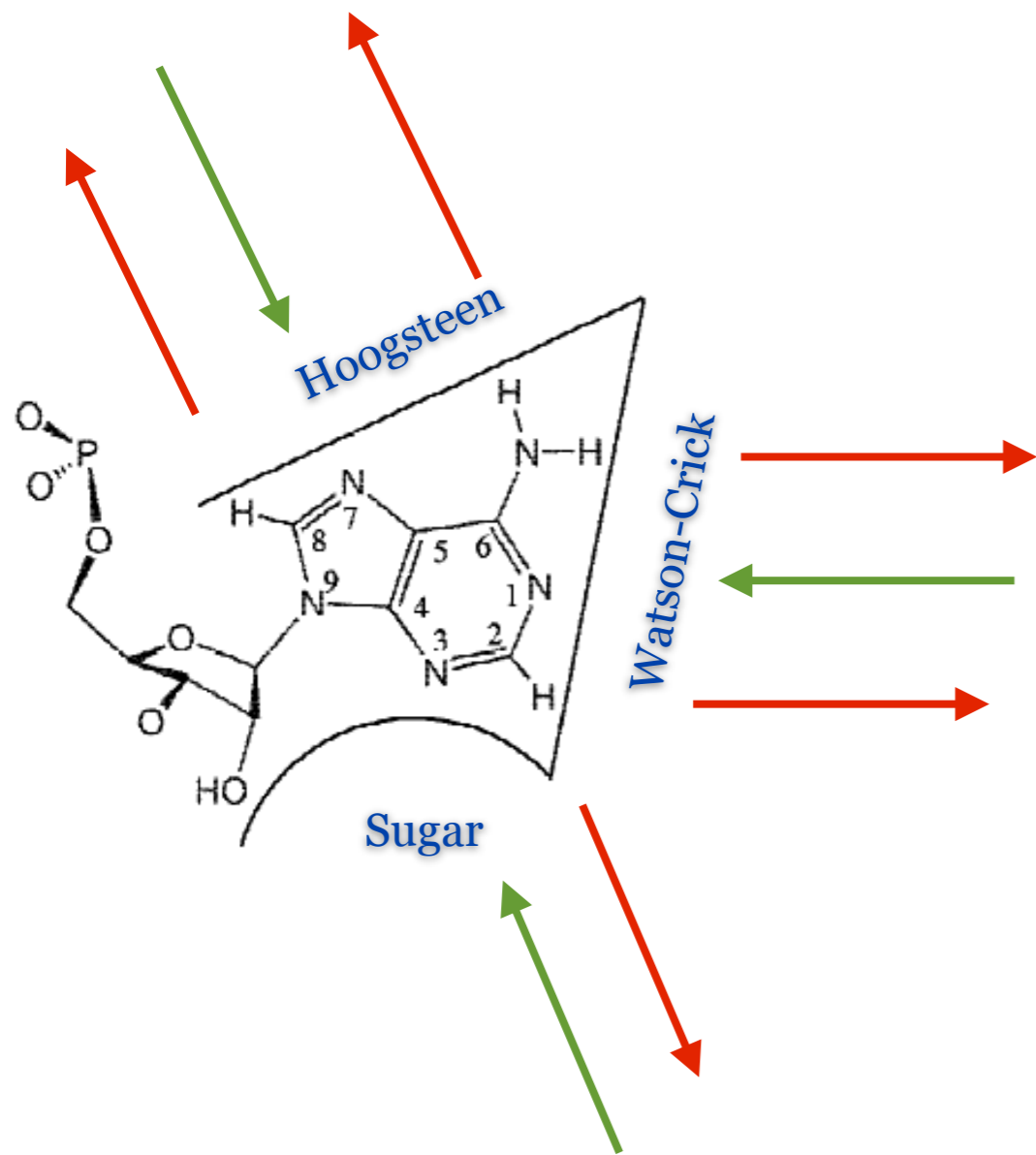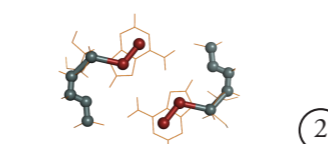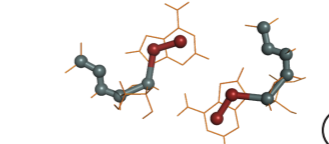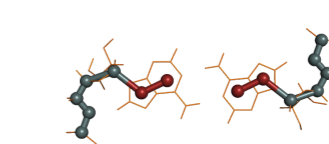

harmonic
statistical parameters

bond stretching   angle bending   bond rotation

X

X

Base orientation

X

$$E_{el} = \varepsilon_{el} \ \frac{q^2}{4\pi\epsilon_0\epsilon_r r} e^{-r/\ell}$$

Pasquali, Derreumaux, JPCB 2010

Cragnolini,, Derreumaux, Pasquali, JCTC 2015
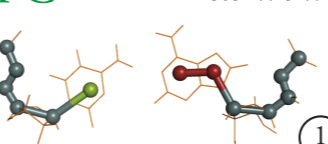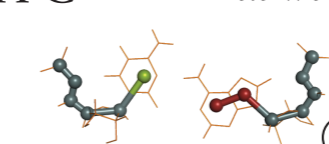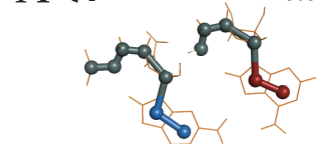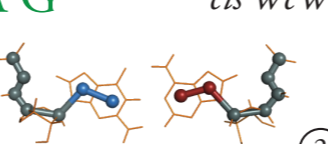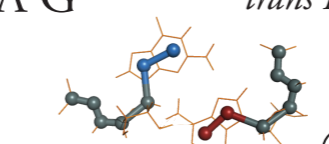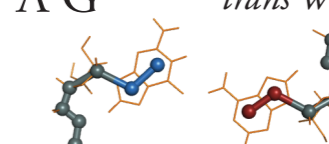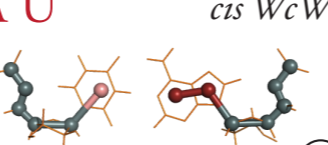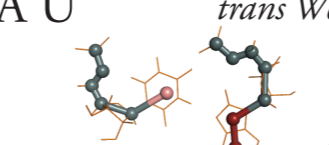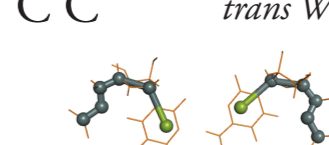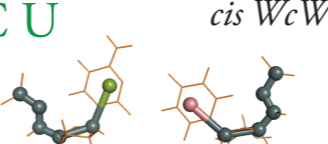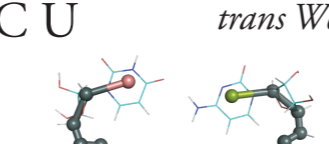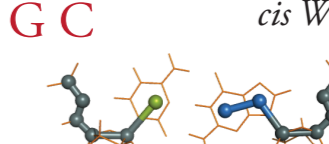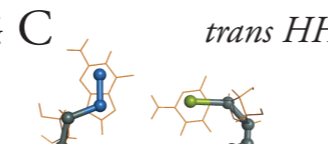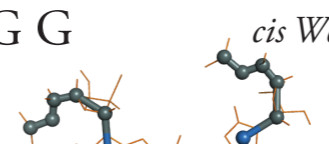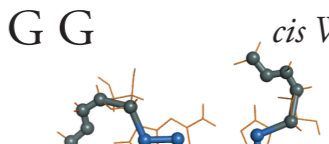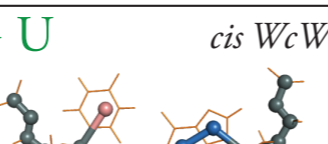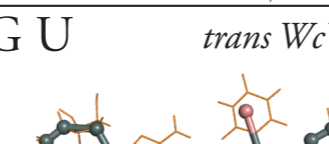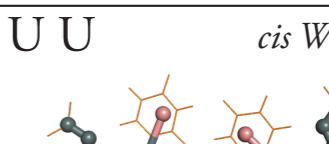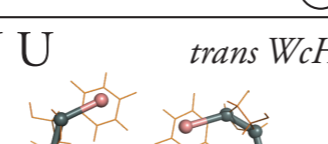
Canonical and non-canonical pairings

Hoogsteen

Watson-Crick

Sugar

A A	*trans HH*	A A	*trans HS*	A A	*trans WcWc*
A C	*cis WcWc*	A C	*cis WcWc*	A G	*cis WcS*
A G	*cis WcWc*	A G	*trans HS*	A G	*trans WcWc*
A U	*cis WcWc*	A U	*trans WcH*	C C	*trans WcWc*
C U	*cis WcWc*	C U	*trans WcH*	G C	*cis WcWc*
G C	*trans HH*	G G	*cis WcH*	G G	*cis WcH*
G U	*cis WcWc*	G U	*trans WcWc*	U U	*cis WcWc*
U U	*trans WcH*				

Wc: Watson-Crick
H: Hoogsteen
S: Sugar

Coarse-grained folding (non-canonical)

Triplets

Quartlet

Cragnolini, Laurin, Derreumaux, Pasquali, JCTC 2015

T Cragnolini, D Chakraborty, J Šponer, P Derreumaux, S Pasquali, D. Wales, J Chem Phys (2017)
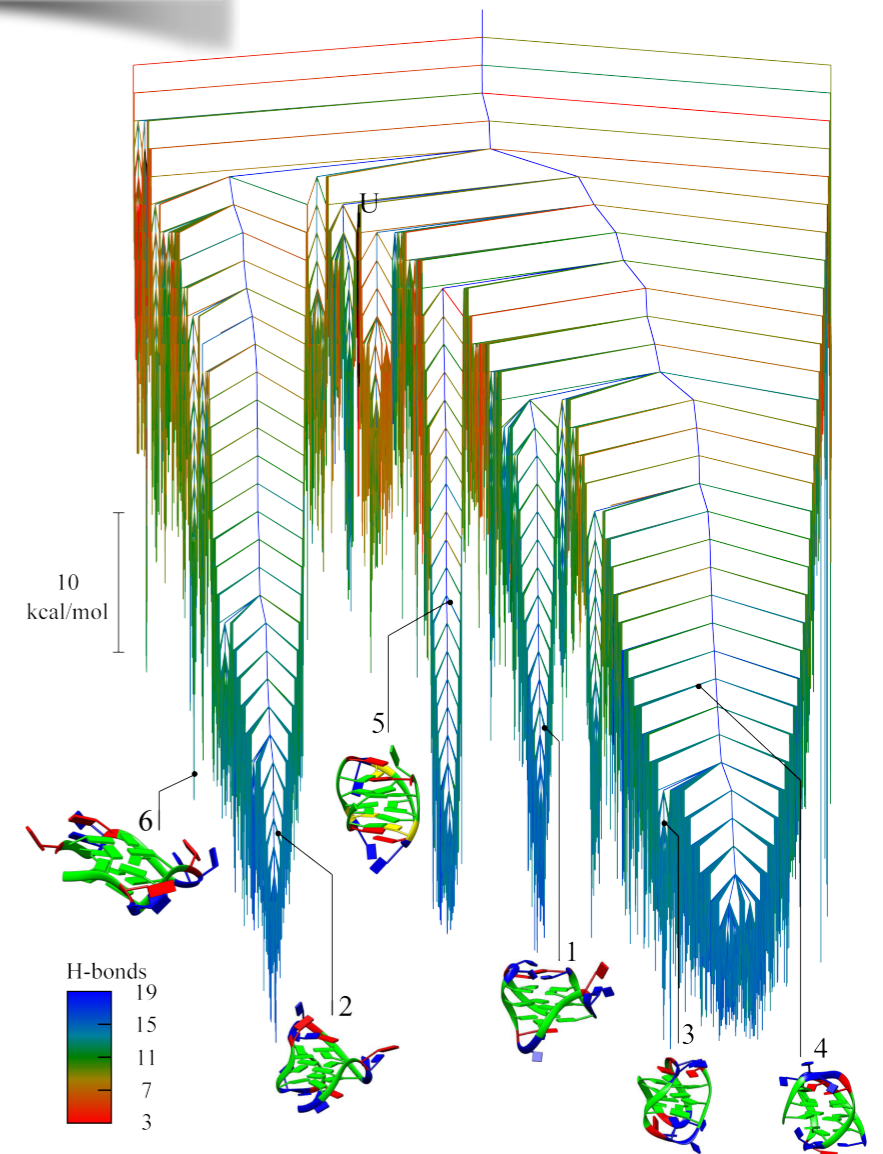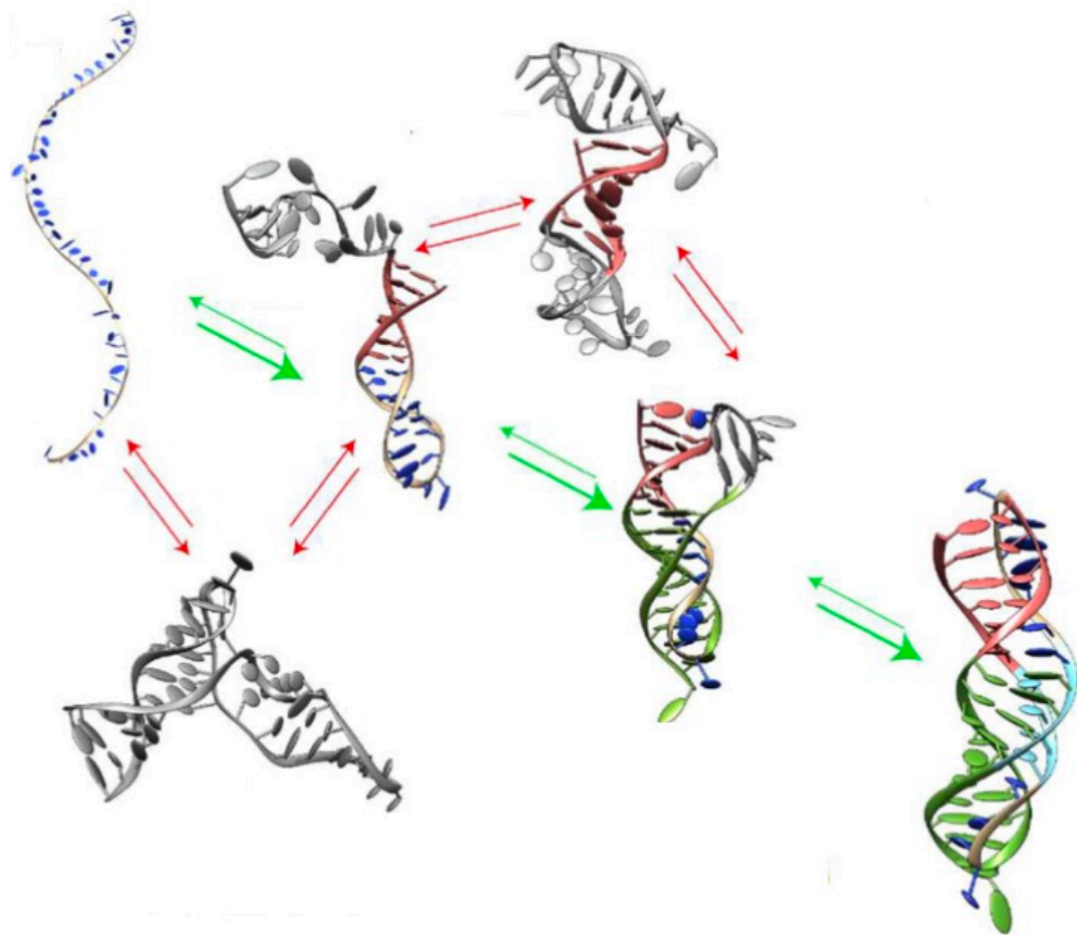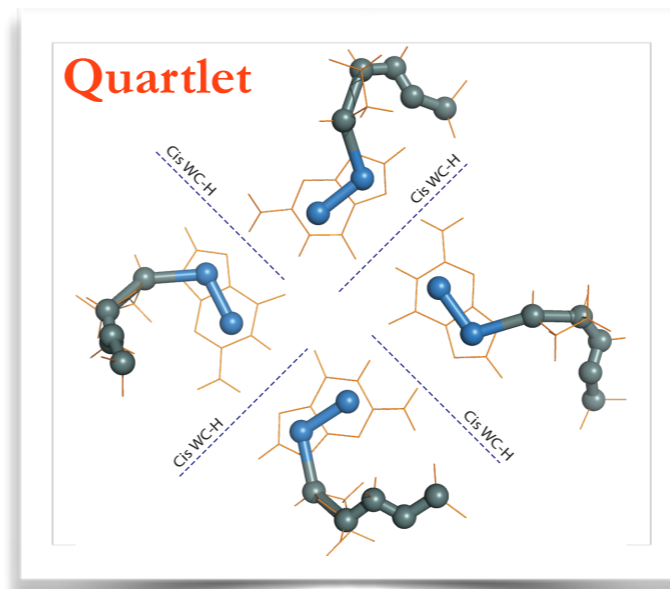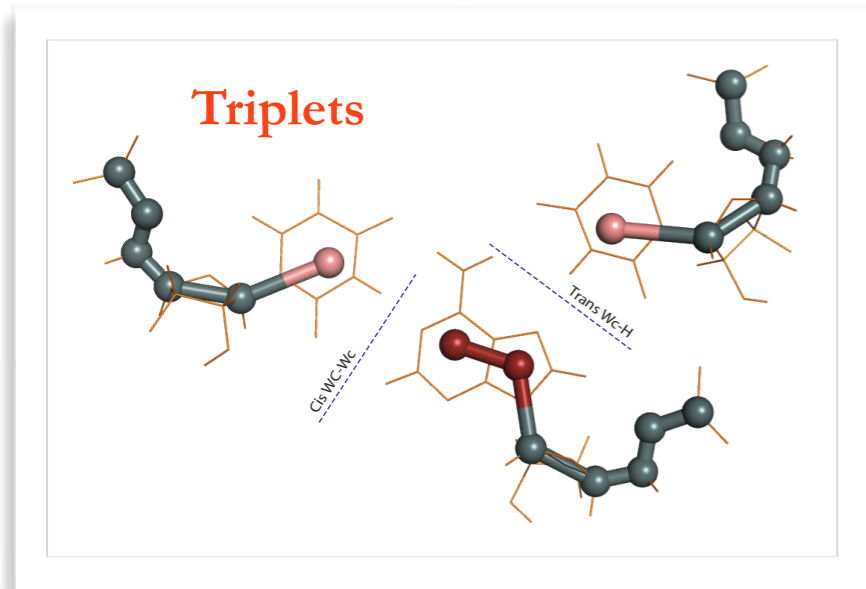
Low-resolution data modeling: SAXS
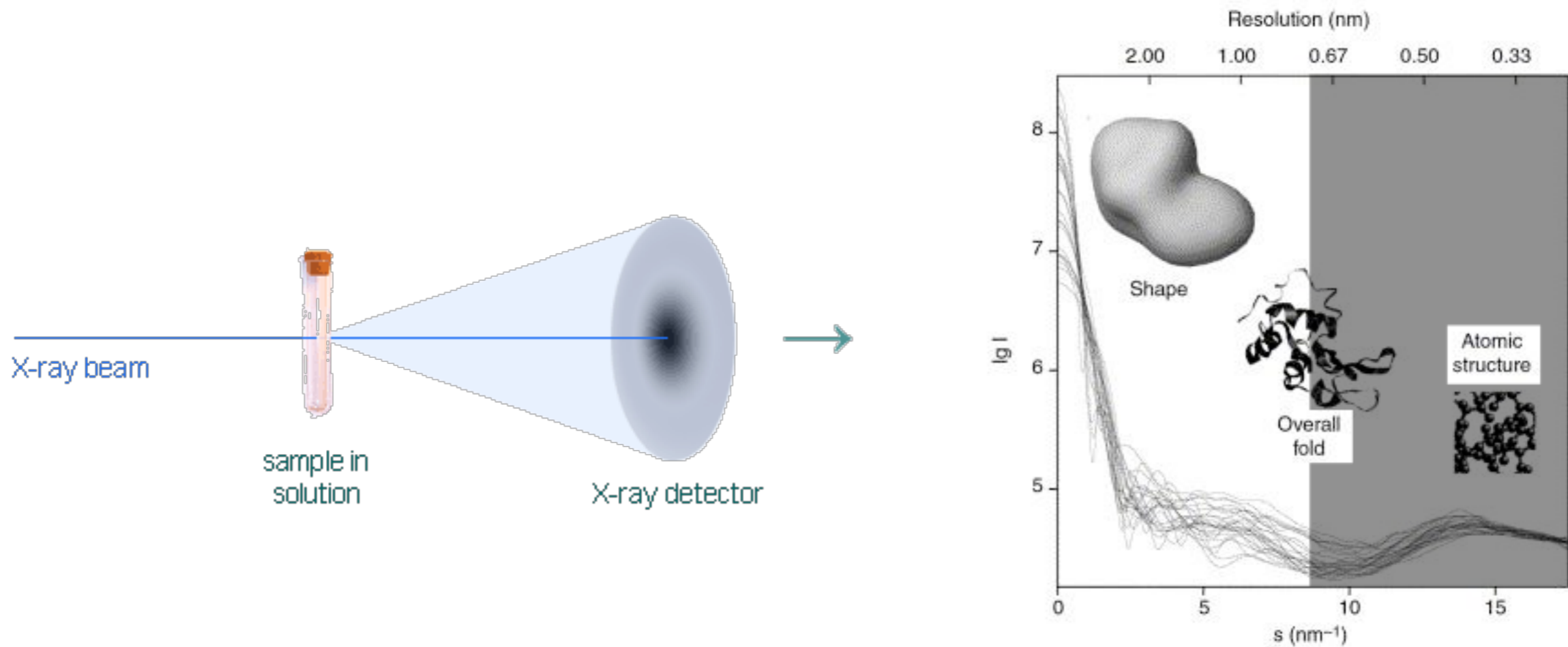
Analytical calculations of scattering curve (Debye)

$$I'_{CG}(q) = \left\langle |A_v(\vec{q}) - \rho_s A_s(\vec{q})|^2 \right\rangle_\Omega = \sum_{I,J}^{N} f_I(q) f_J(q) \, \frac{\sin(qR_{IJ})}{qR_{IJ}},$$

$$I'_{CG}(q) = \left\langle |A_v(\vec{q}) - \rho_s A_s(\vec{q})|^2 \right\rangle_\Omega = \sum_{I,J}^{N} F'_I(q) F'_J(q) \frac{\sin(qR_{IJ})}{qR_{IJ}},$$

$\downarrow$

coarse-grained form factors

Low-resolution data modeling: SAXS

Given a conformation for the RNA we can compute the "theoretical" scattering curve and compare it to the experimental

- Uniform background of constant dielectric properties
- Coarse-grained water molecules on hydration layer
- CG description of molecule

Static
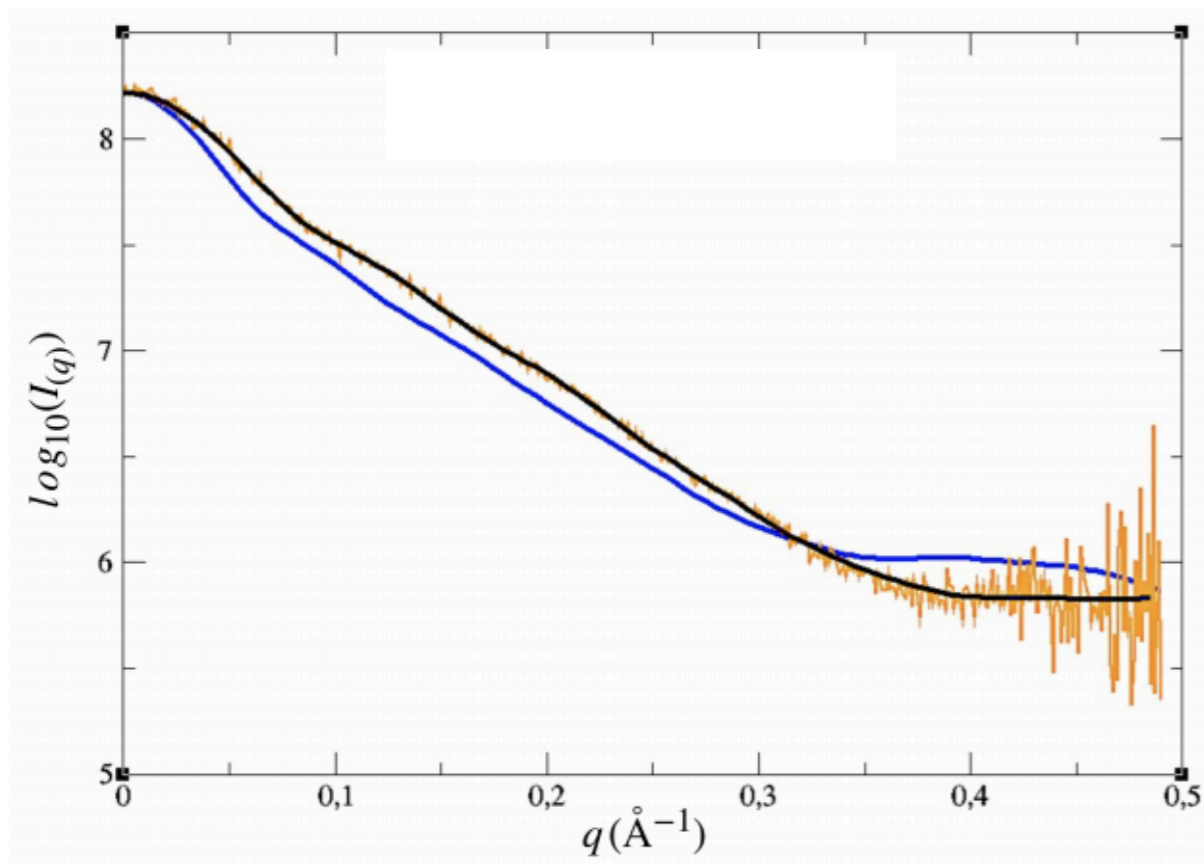
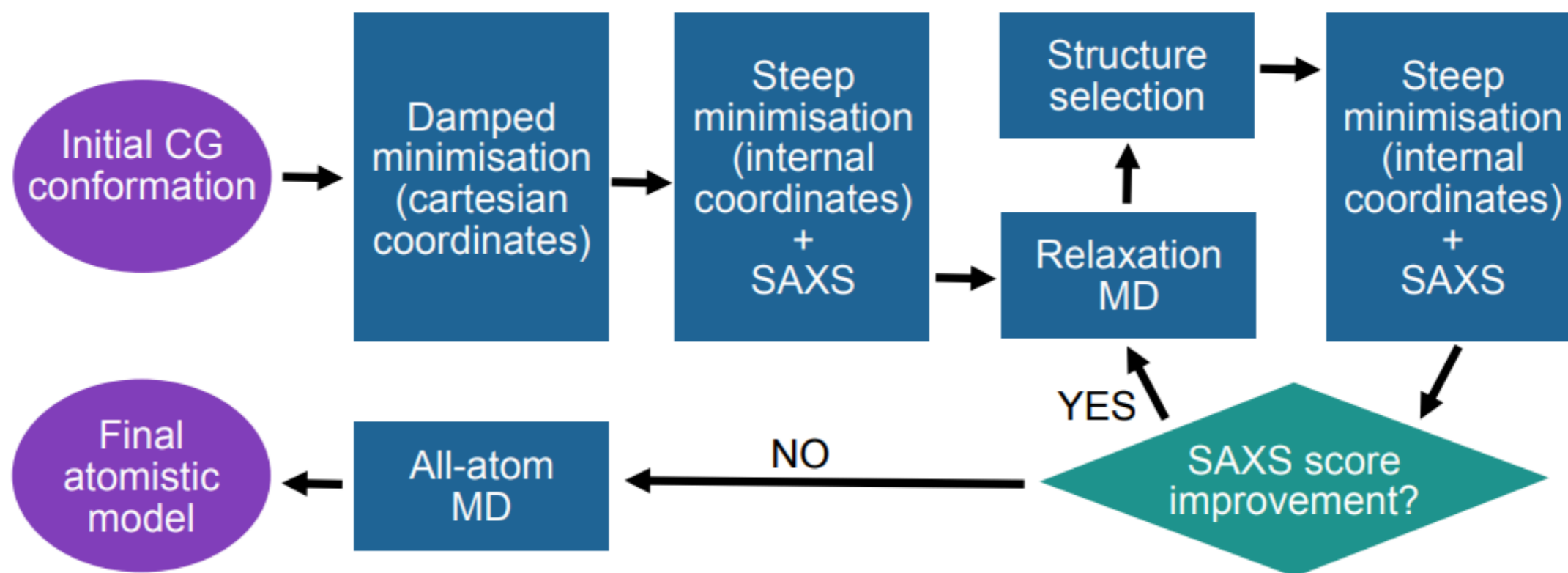

Measure the discrepancy between curves

$$\chi = \frac{\sum_q q^2 \left[\frac{I_0(0)}{I(0)} - I_0(q) \right]^2}{N_q I_0(0)^2}$$

SAXS "energy" and "force"
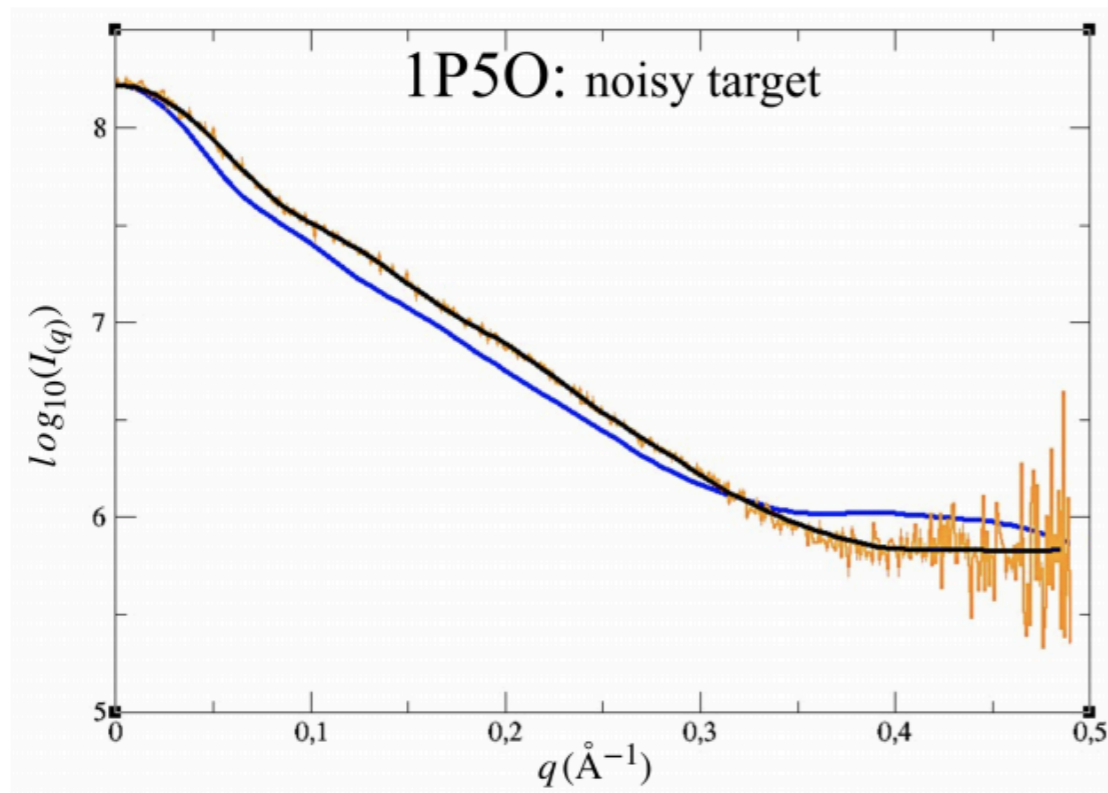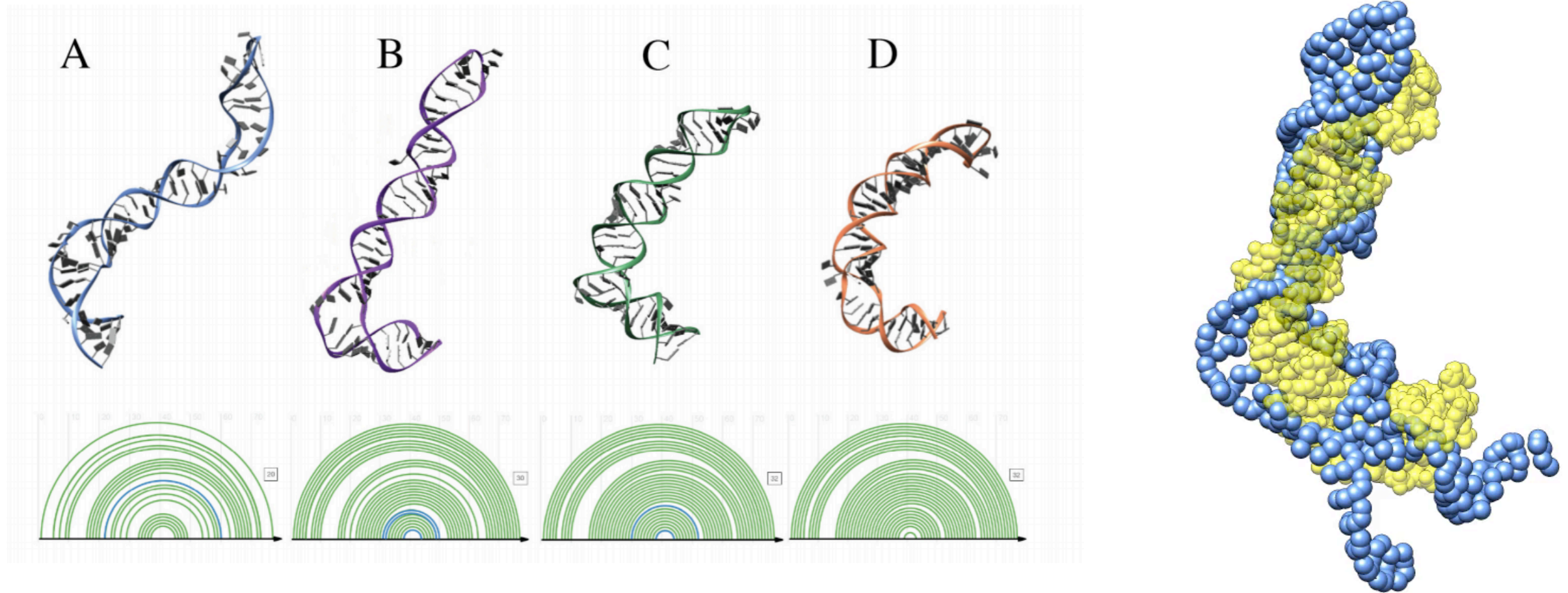
$$E_{SAXS} = \kappa_{SAXS} \, \chi$$

$$\vec{F}_{SAXS} = \sum_q \frac{dE_{SAXS}}{dI(q)} \frac{dI(q)}{d\vec{r}} \qquad F_{tot} = F_{ff} + F_{SAXS}$$

Mazzanti, Alferkh, Frezza, Pasquali, JCTC (2021)

HCV IRES Domain II (PDB ID: 1P5O)

Portion of tRNA (PDB ID: 1A60)
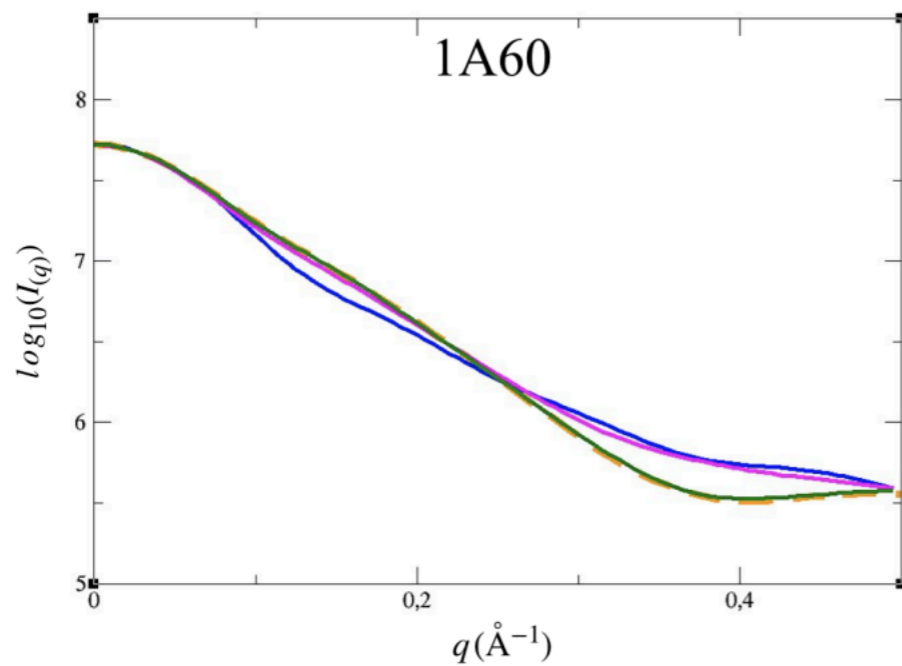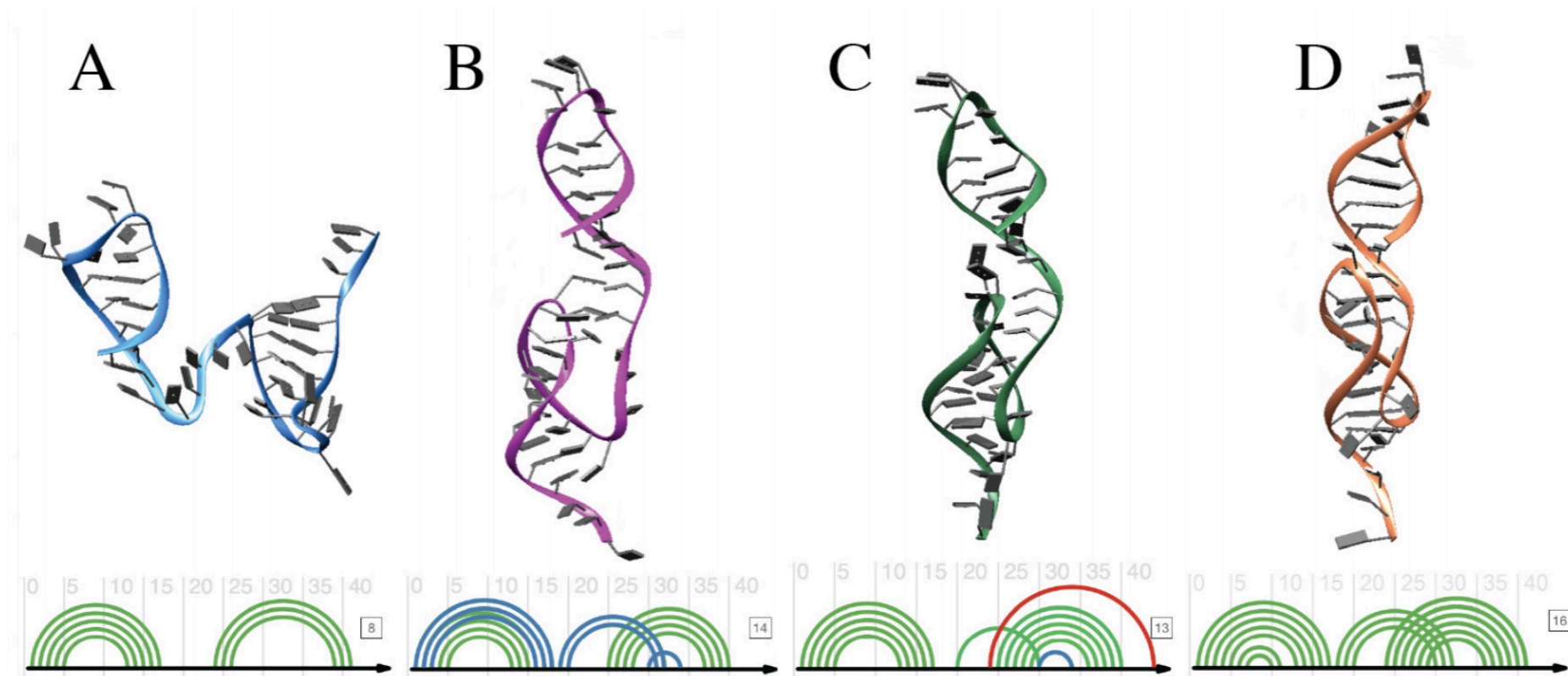
RMSD values comparison according to the native structure		
Starting structure	Control structure	Final structure
15.4Å	7.0Å	6.3Å

Ratio of base pairs matching in the native structure		
Starting structure	Control structure	Final structure
8/16	8/16	11/16

Information from chemical probing data?

DMS
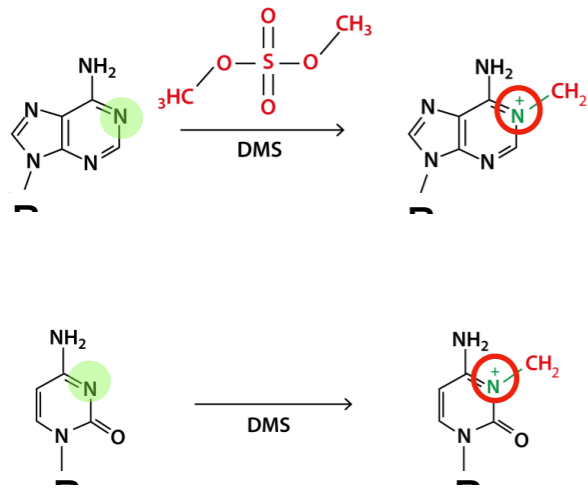
CMCT
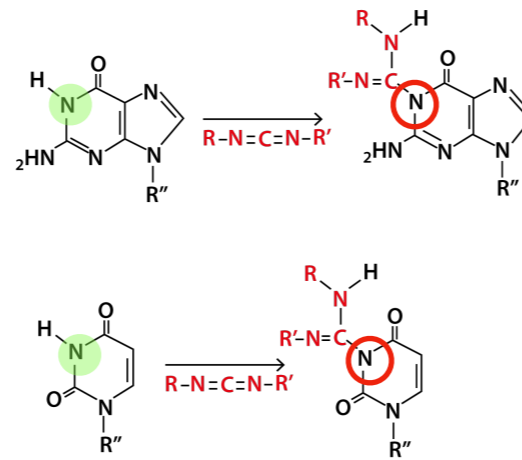
No canonical pairings

SHAPE

N-methylisatoic
anhydride (NMIA)

1-methyl-7-nitro-isatoic
anhydride

NMIA

SHAPE reagents

link to base-pairing?

Interpreting SHAPE data at the mechanical level
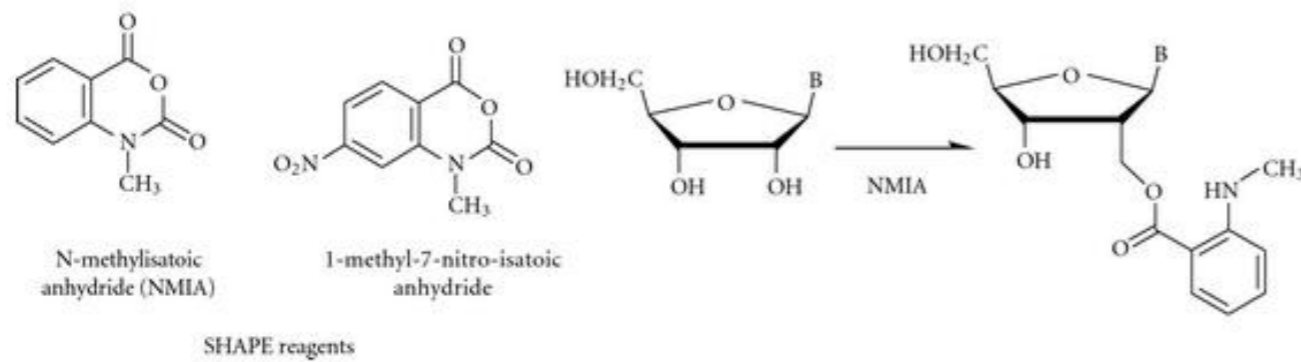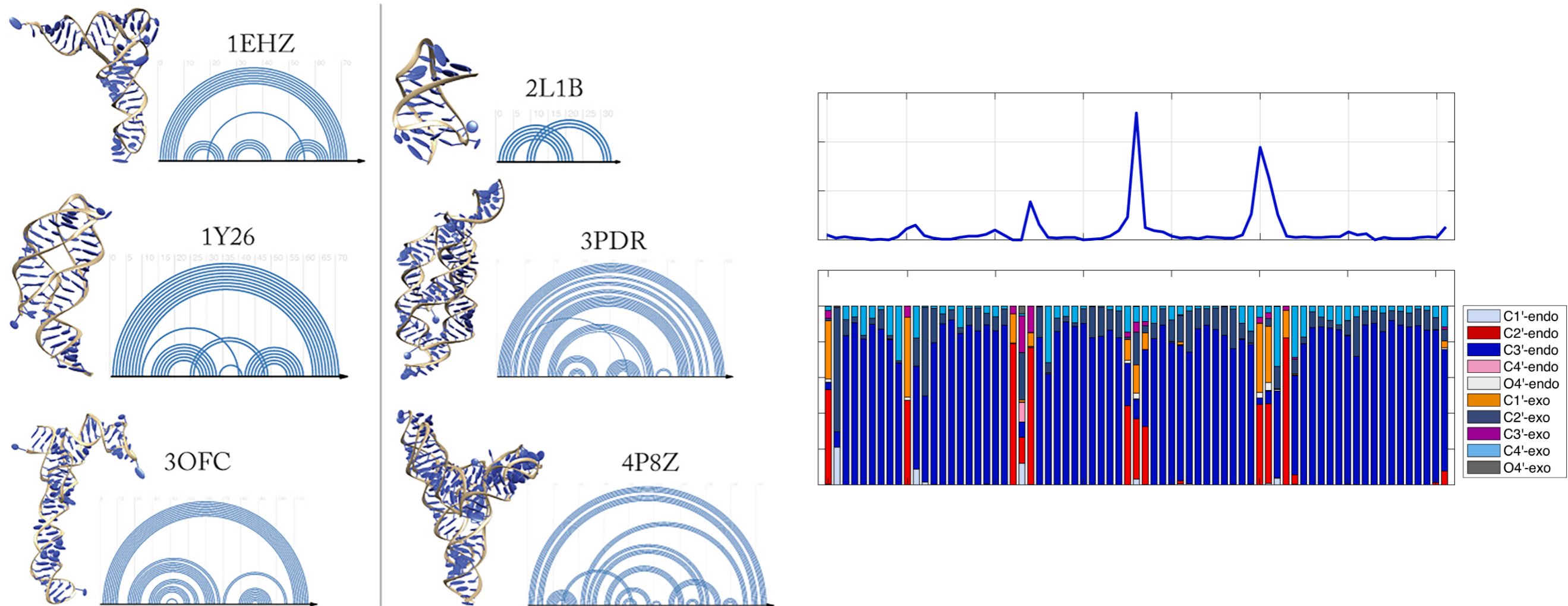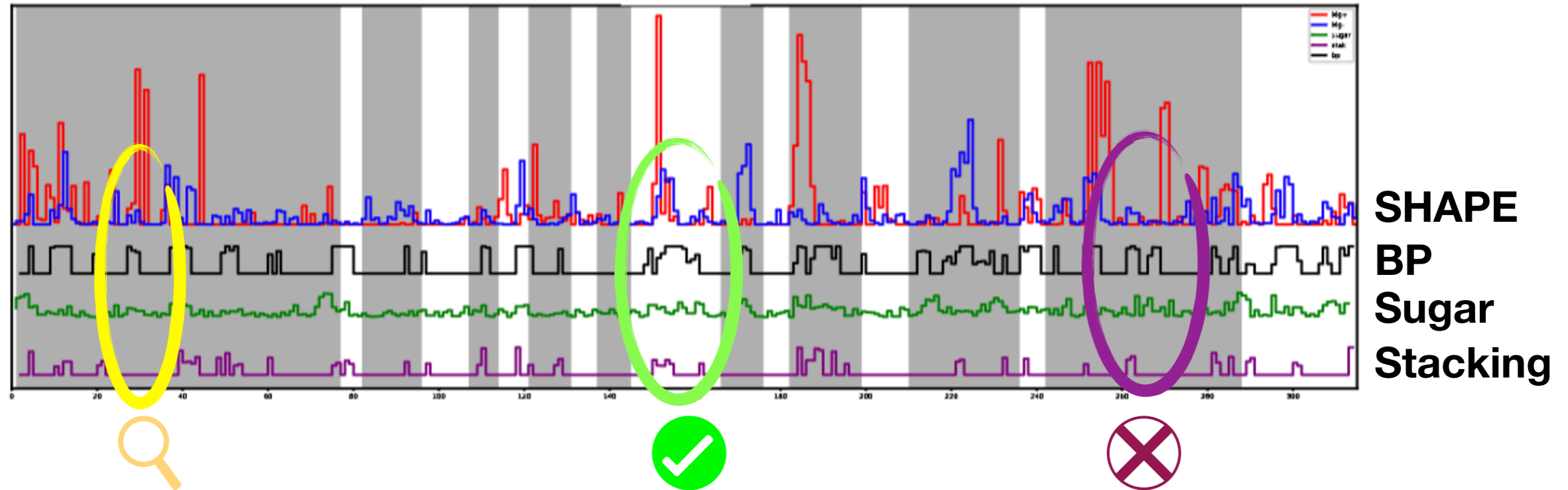
Look for correlations between reactivity and dynamical behavior (atomistic MD)

Sugar flexibility ➞ high reactivity (but not always)
Canonical base pairing ➞ low reactivity (but not always)
Stacking ➞ low reactivity (but not always)

Interpreting SHAPE data at the mechanical level

The best correlations occur when all 3 conditions are met
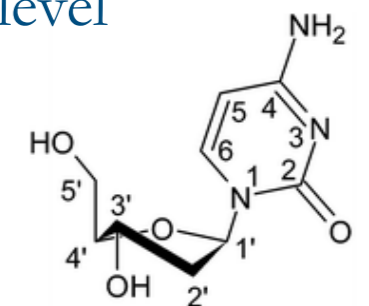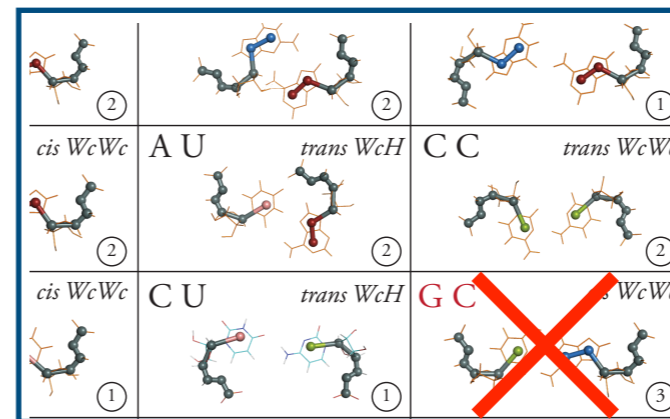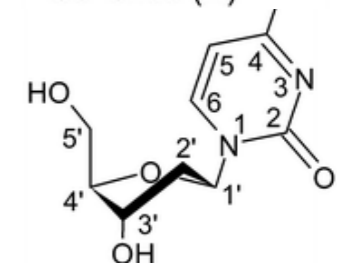**Qualitative information**

Focus on high-reactivity peaks only

Highlight reactive "regions" but harder to pinpoint at the single base level

Envision soft biases for CG simulations:
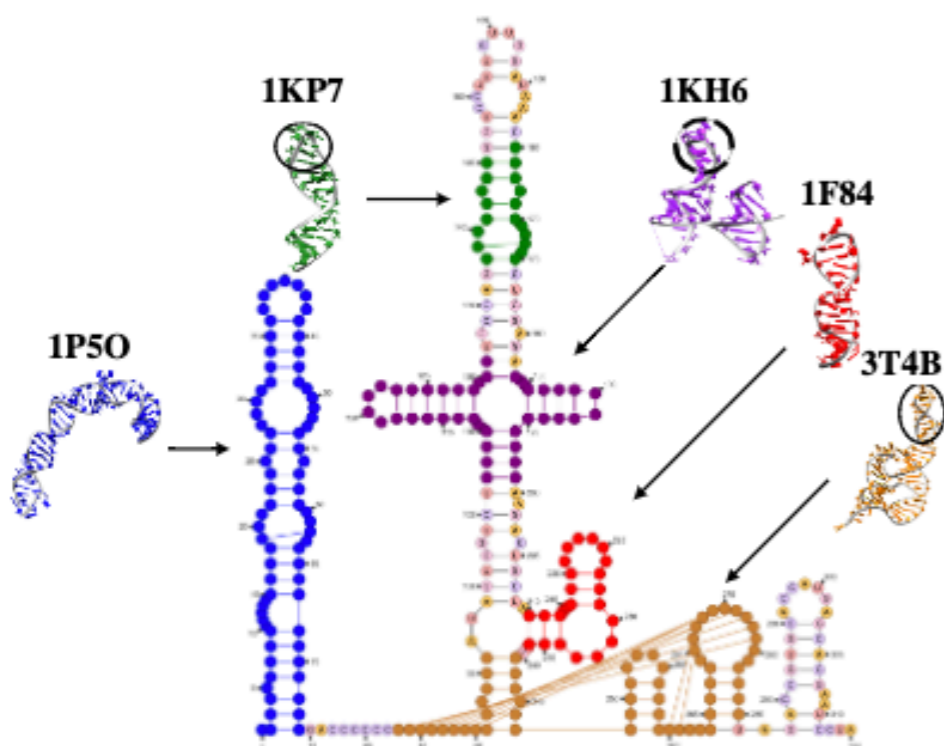sugar torsions
base-pairing energy
stacking energy

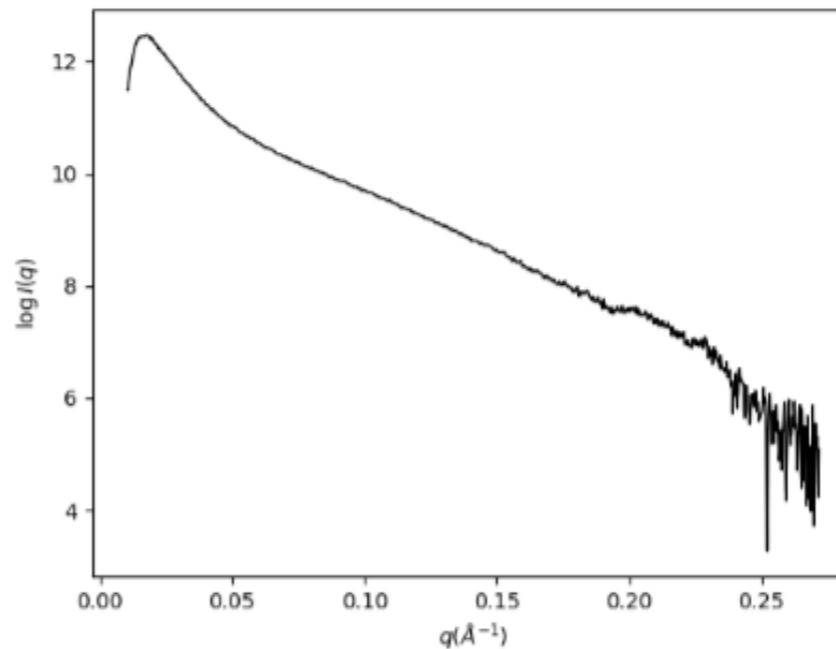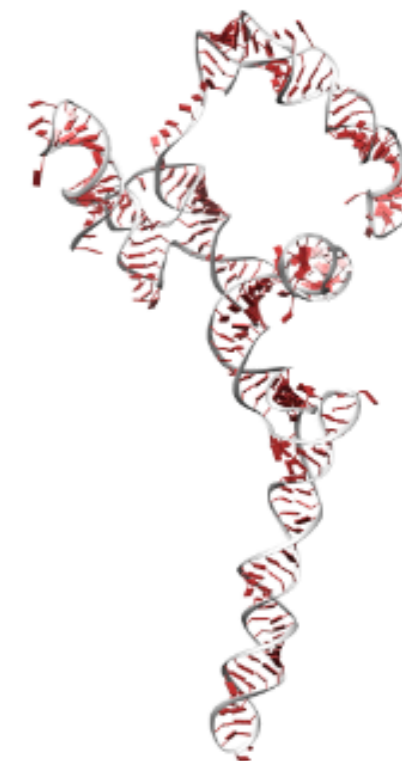# HCV Internal Ribosomal Entry Site (IRES)

MD

SAXS full system

MD clusters from atomistic simulation

Cluster	Structure number	*RMSD* of middle structure (Å)
1	27	0.324
2	28	0.356
3	64	0.388
4	47	0.360
5	48	0.374
6	82	0.472
7	118	0.500
8	151	0.585
9	64	0.457
10	239	0.635
11	124	0.551
12	131	0.438
13	180	0.557

CG minimization guided by SAXS data
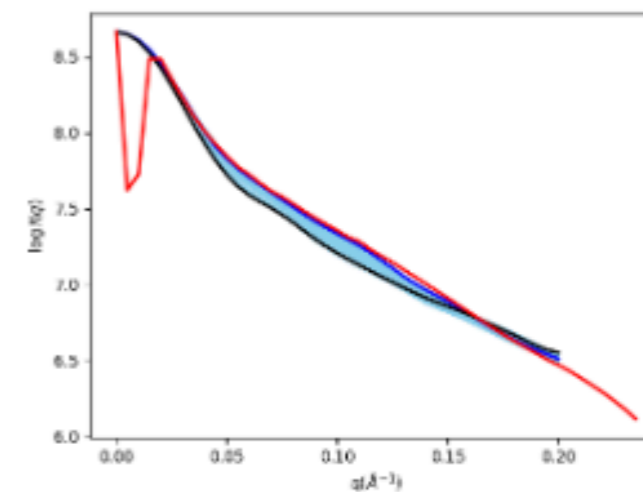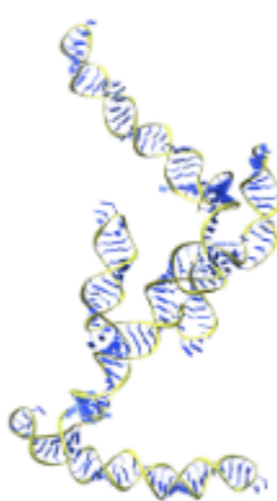
Model HCV IRES

Generate multiple conformers satisfying SAXS constraint (CG modeling)
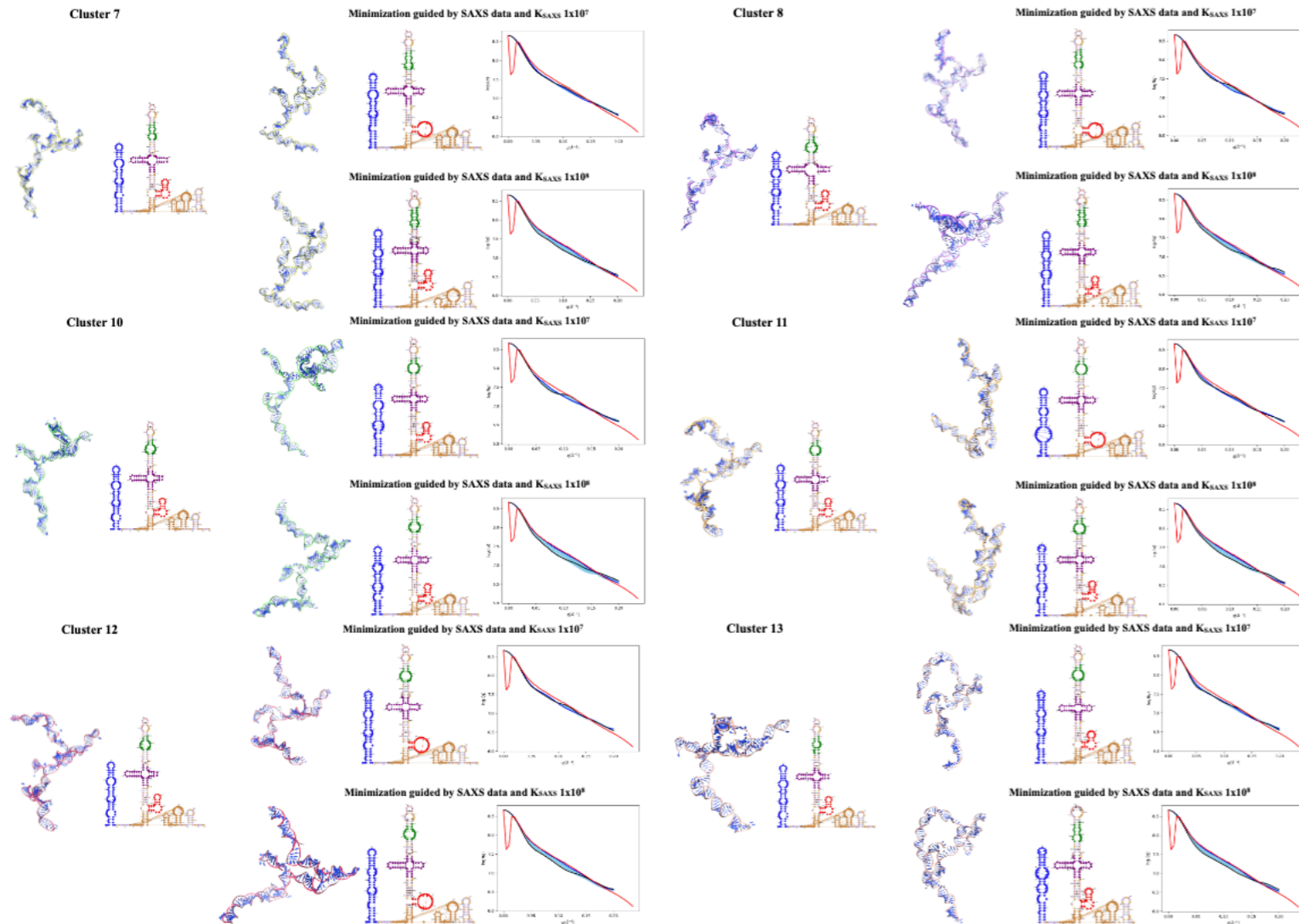
Select conformers matching SHAPE behavior in atomistic simulations
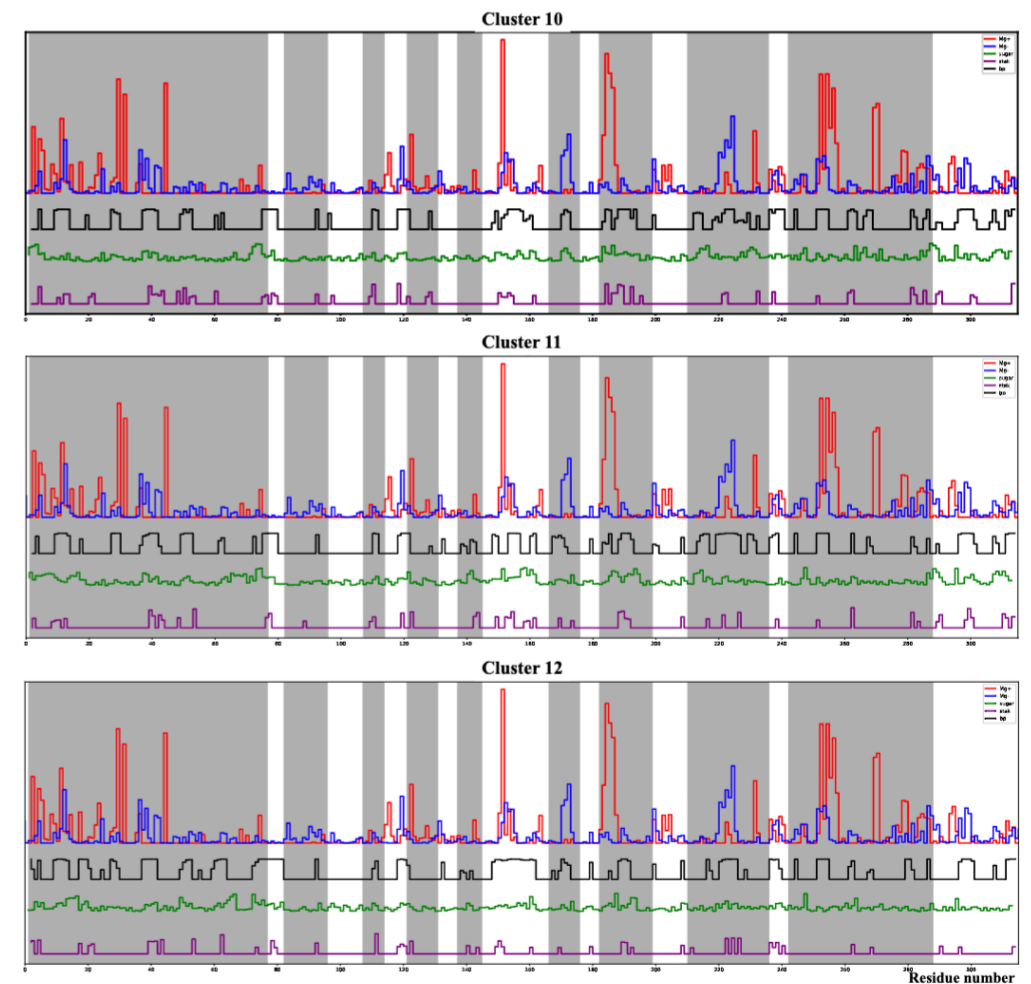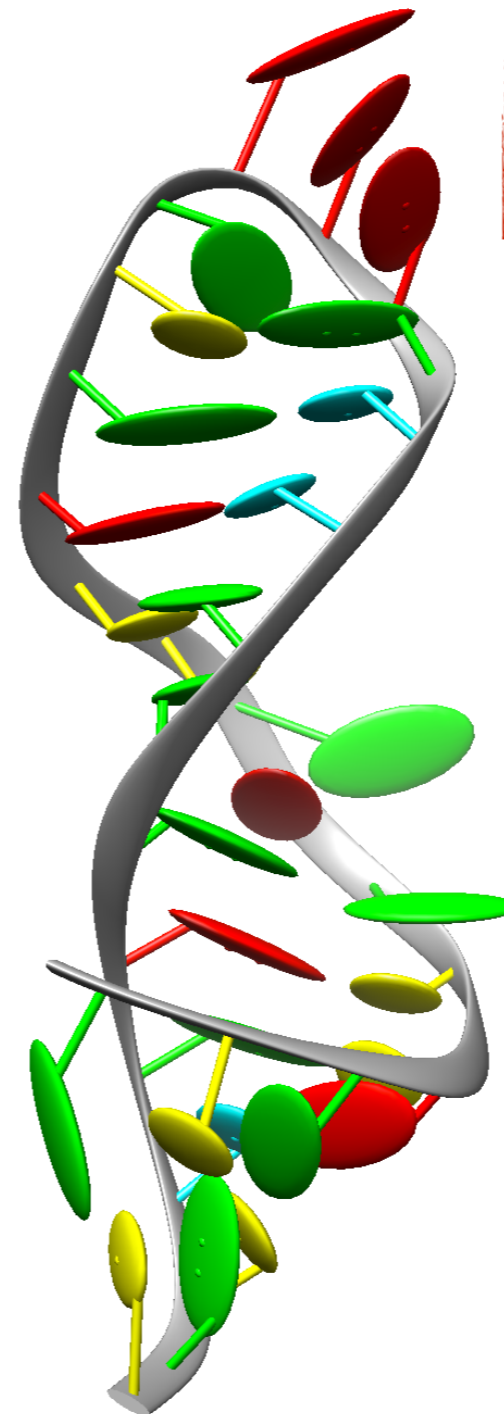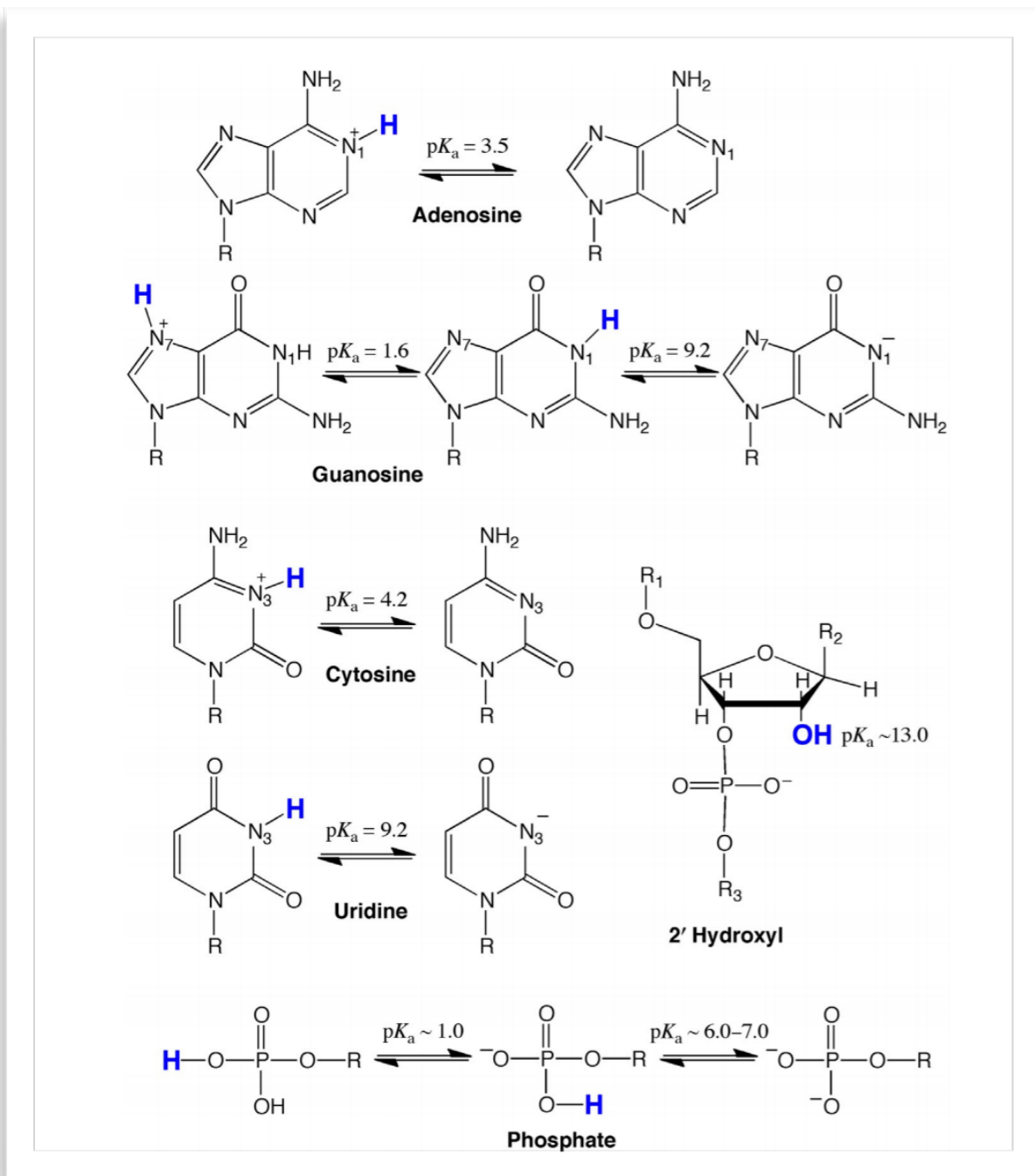

Select structure(s) most compatible with experimental data to attempt the conformational transformation from unbound to bound state (to the ribosome)

Accounting for pH
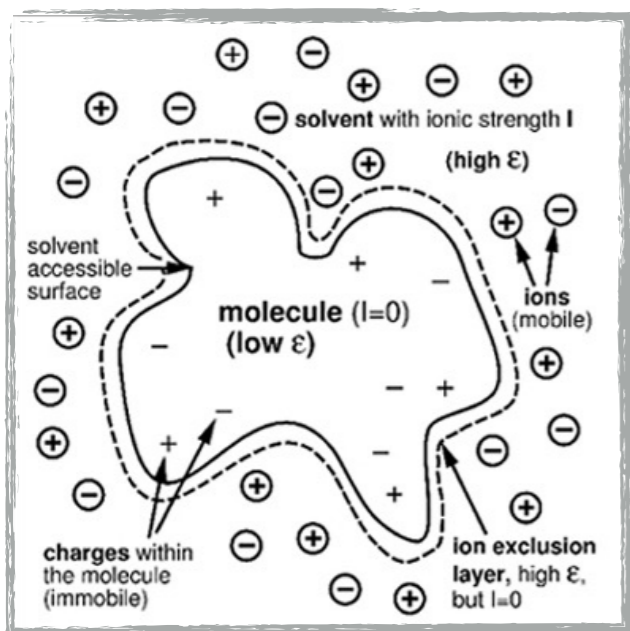

exposed ┈▶ protonable
pKa ~ isolated

neutral paired ┈▶ protected
lower pKa

N+ paired ┈▶ protected
higher pKa

RNA being strongly charged, charge regulation mechanisms are important for the structure adopted and for recognition mechanisms!

Fast MC titration scheme

J. Kirkwood, J Chem Phys 1934

C. Tanford, J. Kirkwood, JACS 1957

$$w_{TK} \approx \frac{e^2}{8\pi\epsilon_0\epsilon_r} \sum_{i>j}^{N_p} \left(\frac{z_i z_j}{r_{ij}} - \frac{Z_p^2 \kappa}{2(1+\kappa b)} \right) \pm (pH - pK_a)$$

protonation (+)
deprotonation (-)

Titratable sites are placed at their experimental position
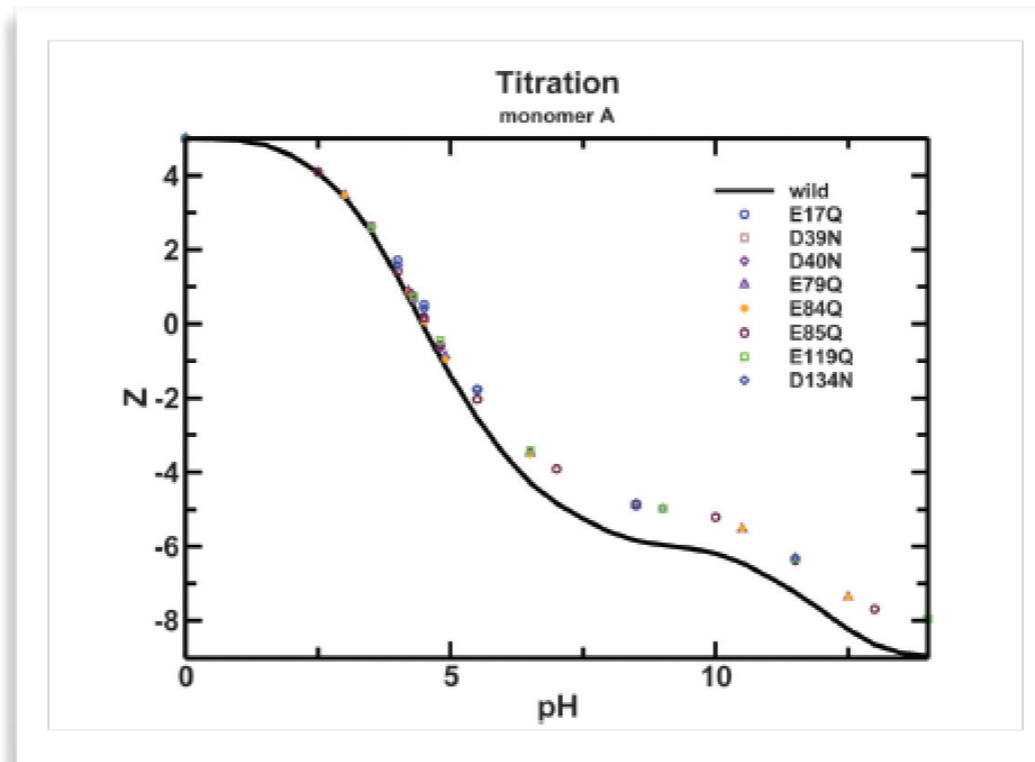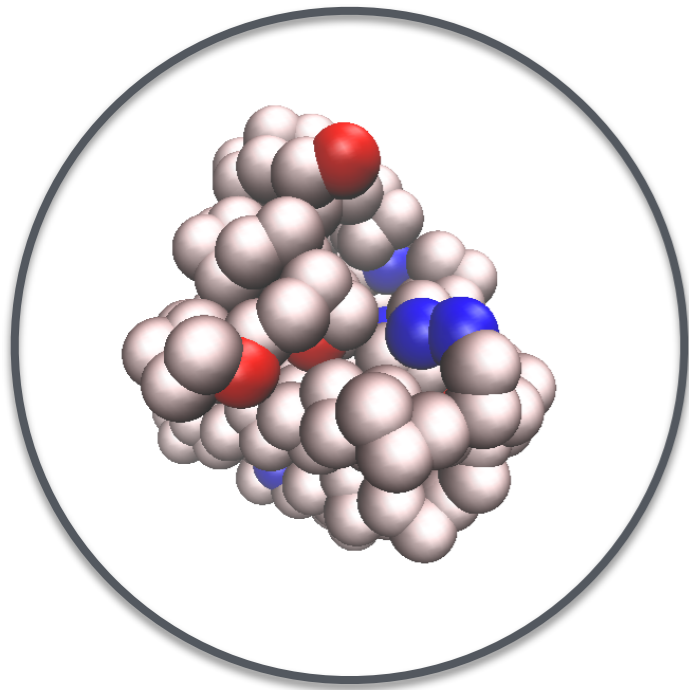from a PDB structure and kept fixed

Randomly pick one titratable site

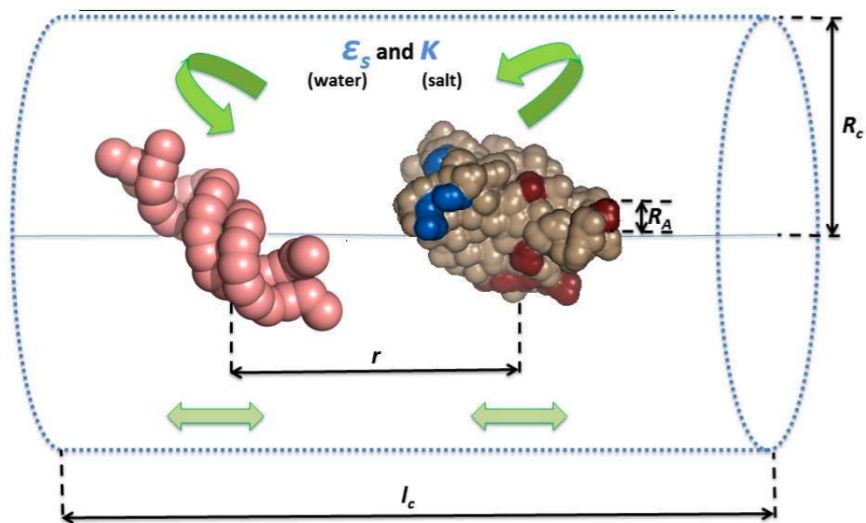- if deprotonated add proton
- if protonated remove proton

accept or reject based on the probability : $\min\left(1, e^{-\beta \Delta w_{TK}}\right)$

test for convergence of physical properties
(average residue charge, total dipole moment, capacitance)
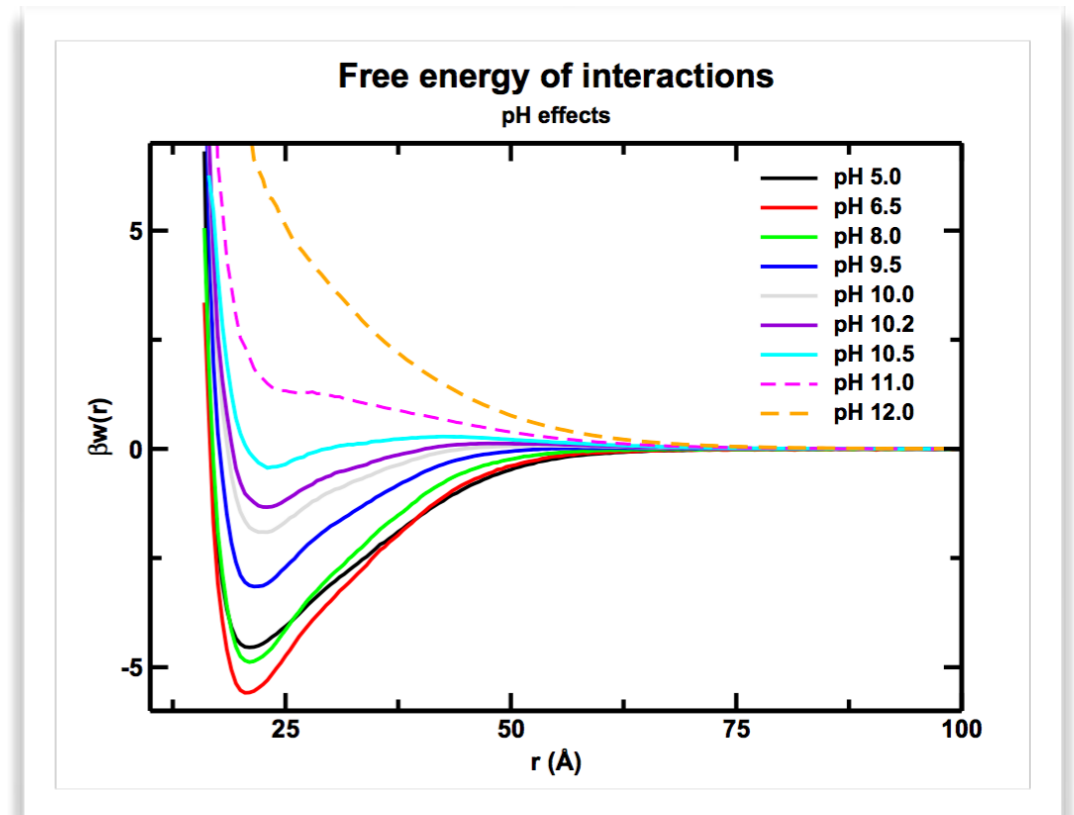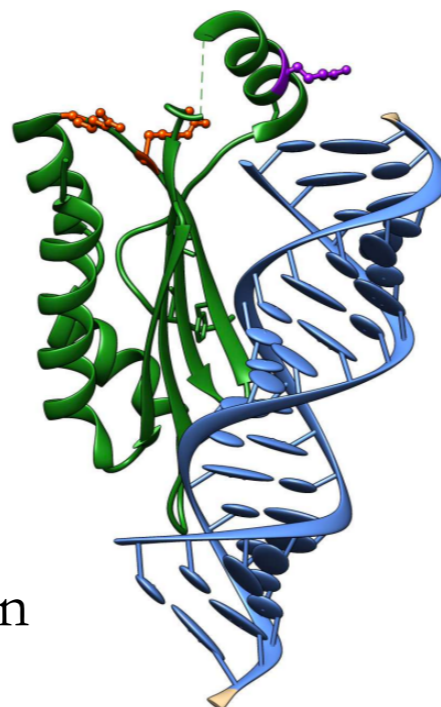
Texeira, Lund, Barroso da Silva, JCTC, 2010

Barroso da Silva, MacKernan, JCTC 2017

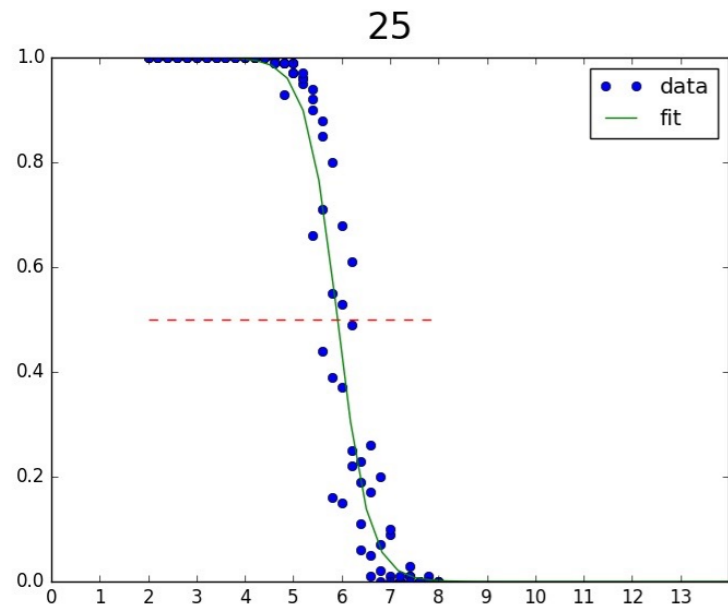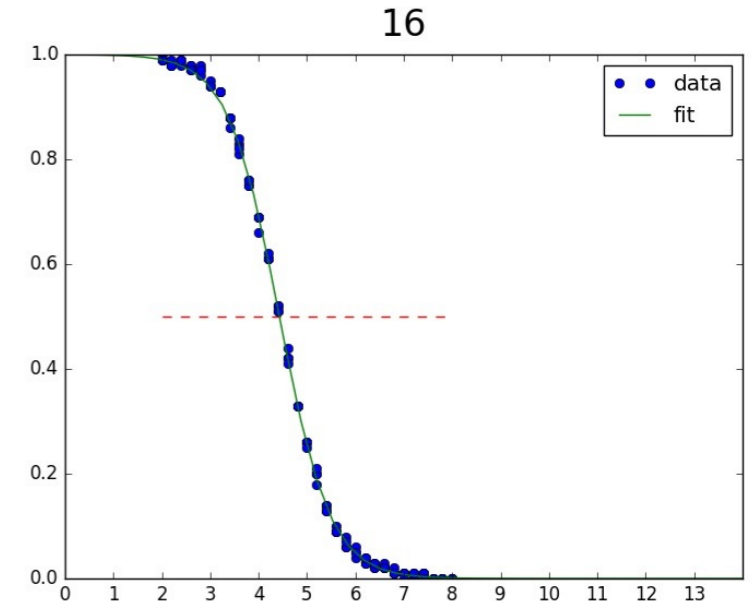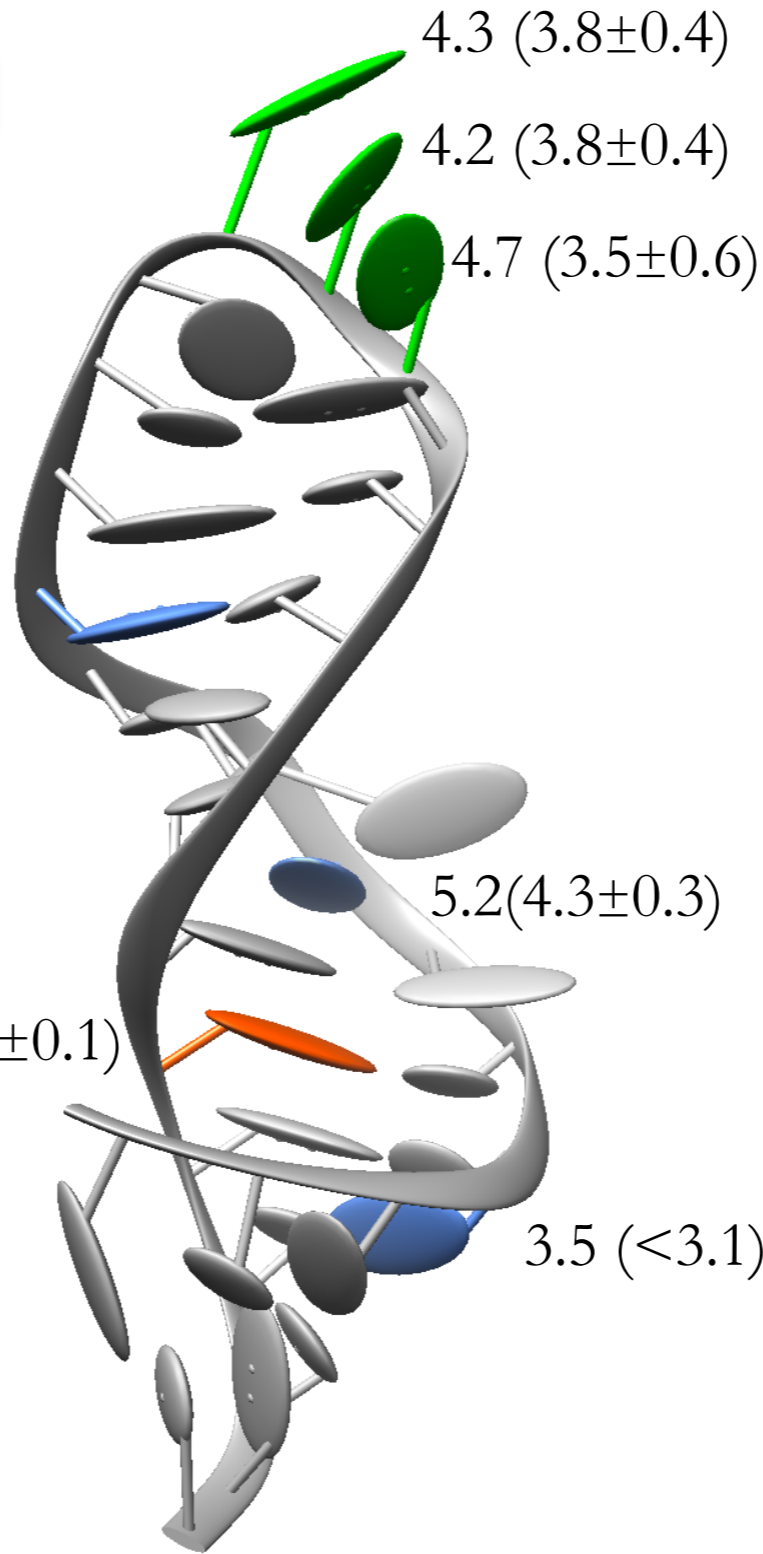# Fast MC titration scheme on fixed structures

Titration
monomer A

individual effective
pK$_a$ values

Argonaute protein
and si-RNA

Free energy of interactions
pH effects

Titration-HiRE-RNA pK$_a$

CG MD → titration → charges → new BP → CG MD

4.3 (3.8±0.4)
4.2 (3.8±0.4)
4.7 (3.5±0.6)

3.4 (<3.1)

5.2 (4.3±0.3)

6.0 (6.5±0.1)

3.5 (<3.1)
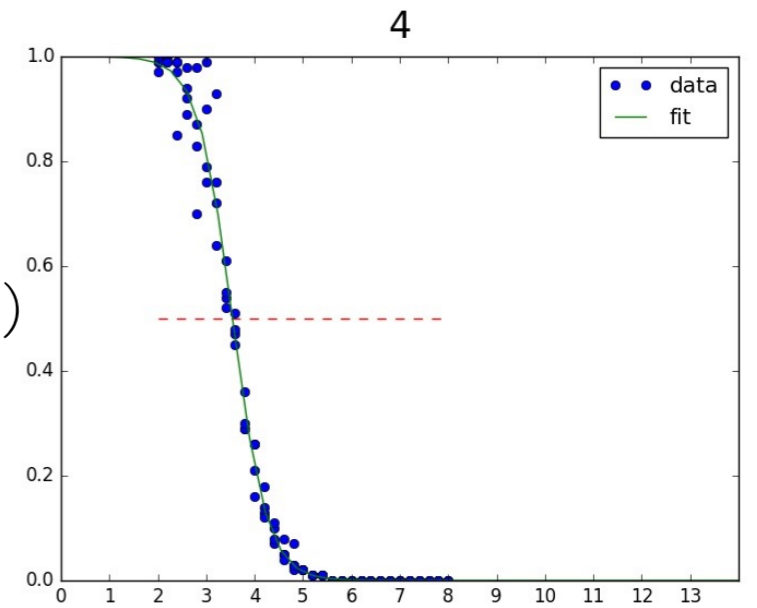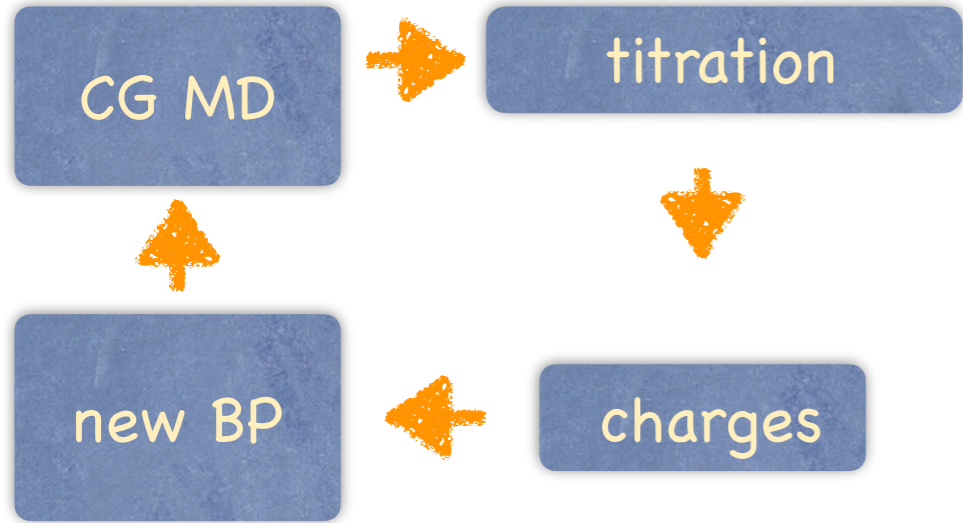
Observe different structures at different pH values

Pasquali, Frezza, Barroso da Silva, Interface Focus (2019)

FPTS-HiRE-RNA structures

pH 2

pH 8
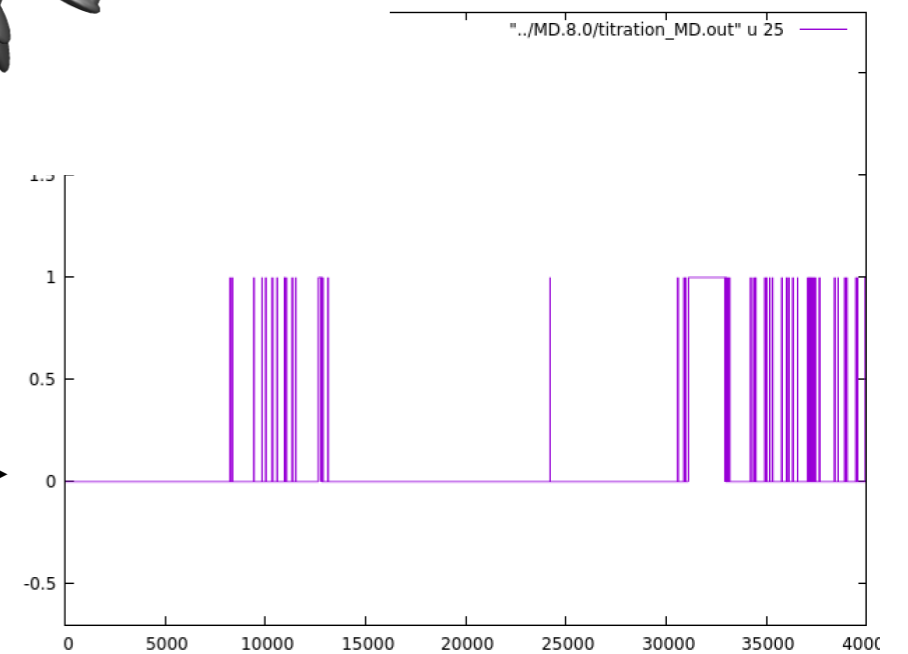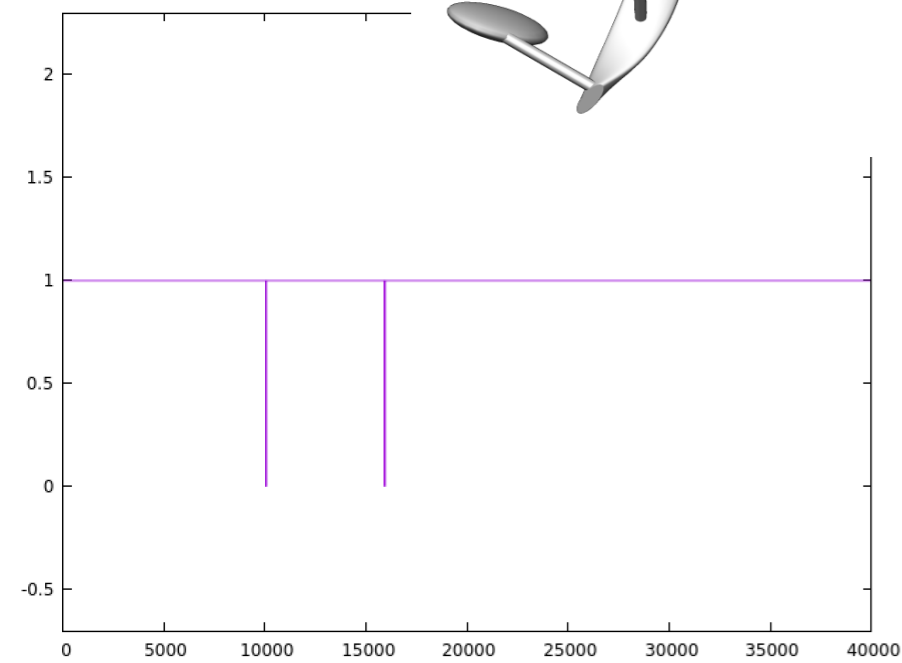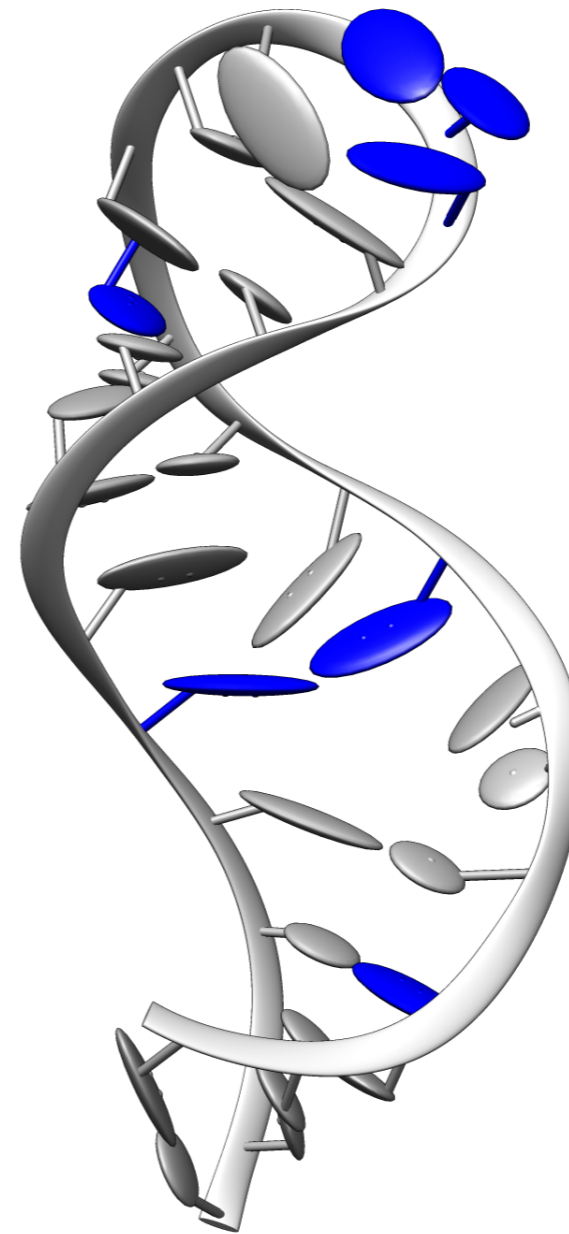
fluctuation of charge
of a given base

For the future

Study RNA landscapes in conjunction with drug design

- Optimize the coarse-grained model with Machine Learning

  → collaboration ENS Paris

  Looking for student for Master thesis and Ph.D. (funding already available)

- Develop multi scale path sampling (larger systems: few dozen to few hundreds nt)
  Study EL modification upon binding

  → collaboration University of Cambridge and King's College London

  Looking for postdoc (3 years, funding already available)

- Test EL modifications and binding with chemical probing (Bruno Sargueil)

Acknowledgments